

**A DEFENSE OF IMPRECISE CREDENCES IN INFERENCE  
AND DECISION MAKING<sup>1</sup>**

James M. Joyce  
The University of Michigan

Some Bayesians have suggested that beliefs based on ambiguous or incomplete evidence are best represented by families of probability functions. I spend the first half of this essay outlining one version of this *imprecise model of belief*, and spend the second half defending the model against recent objections, raised by Roger White and others, which concern the phenomenon of *probabilistic dilation*. Dilation occurs when learning some definite fact forces a person's beliefs about an event to shift from a sharp, point-valued subjective probability to an imprecise spread of probabilities. Some commentators find dilation disturbing, both from an epistemic and a decision-theoretic perspective, and place the blame on the use of sets of probabilities to model opinion. These reactions are based on an overly narrow conception of imprecise belief states which assumes that we know everything there is to know about a person's doxastic attitudes once we have identified the spreads of values for her imprecise credences. Once we recognize that the imprecise model has the resources to characterize a much richer family of doxastic attitudes than this, we will see that White's charges of epistemological and decision theoretic incoherence are unfounded.

**The Basic Ideas of Bayesian Epistemology**

Speaking broadly, the Bayesian approach to epistemology is based on four interconnected ideas:

1. Belief is not all-or-nothing. Opinions come in varying gradations of strength which can range from full certainty of truth, through equal confidence in truth and falsehood, to complete certainty of falsehood.
2. Gradational beliefs are governed by the laws of probability, which codify minimum standards of inductive consistency (or "coherence") to which rational opinions must conform.

3. Learning involves updating a “prior” state of gradational belief to obtain a “posterior” state that preserves as many of the prior beliefs as possible consistent with the evidence received.
4. Rational agents use their graded beliefs, in conjunction with their desires, to choose actions with maximum *expected* desirability. In particular, someone who is determinately more confident of  $X$  than of  $Y$  will place a higher expected value on a wager  $O_X$  that offers a prize when  $X$  and a penalty when  $\sim X$  than on a wager  $O_Y$  that offers the same prize when  $Y$  and the same penalty when  $\sim Y$ .

While 1–4 are common to all versions of Bayesianism, there is no consensus among Bayesians about how these broad principles should be implemented. One point of contention has to do with the degree of *precision* with which gradational beliefs are best represented. According to one view, the strengths of beliefs can typically be measured in sharp numerical degrees, so that, in most contexts, we can think of a believer’s level of confidence in a proposition  $X$  as being captured by a single real number of cardinal significance,  $c(X)$ , which falls between zero and one. This number, the believer’s *credence* for  $X$ , is zero when she is certain that  $X$  is false, one when she is certain that  $X$  is true, and is one-half when she is exactly as confident in  $X$ ’s truth as in its falsehood. Bayesians who adopt this *precise model* of graded belief construe 1–4 as follows:

- 1P. A believer’s overall *credal state* can be represented by a single *credence function*  $c$  that assigns a sharp numerical degree of belief  $c(X) \in [0, 1]$  to each proposition  $X$  in some Boolean algebra  $\Omega$ .<sup>2</sup>
- 2P. If the believer is rational then her credence function obeys the laws of probability, so that, for all  $X, Y \in \Omega$ , (a)  $c(X \vee \sim X) = 1$ , (b)  $c(X \& \sim X) = 0$ , (c)  $c(X \vee Y) + c(X \& Y) = c(X) + c(Y)$ , and (d)  $c(X \& Y) = c(X) \cdot c(Y | X)$  where  $c(Y | X)$  is  $Y$ ’s probability on the supposition of  $X$ ’s truth.<sup>3</sup>
- 3P. Learning is governed by *Bayes’ Theorem*. A believer who has a learning experience in which she becomes certain of a proposition  $D$  (and learns nothing else) should end up with a posterior credence function  $c^*$  such that  $c^*(X) = c(X) \cdot [c(D | X) / c(D)]$  for all  $X \in \Omega$ .<sup>4</sup>
- 4P. Rational decision making is a matter of choosing actions that maximize *expected utility* relative to one’s credences. More precisely, if an act  $A$  produces an outcome of utility  $u_A(X_n)$  for each  $X_n$  in a partition  $\{X_1, X_2, \dots, X_N\}$  of mutually exclusive, collective exhaustive events, then an agent with credence function  $c$  will assess  $A$ ’s choiceworthiness by its *expected utility*  $Exp_c(A) = \sum_n c(X_n) \cdot u_A(X_n)$ .

### The Need for Imprecise Credences

There are a variety of problems with the idea that credences must be precise. As many commentators have observed (see, e.g., Kyburg [1983], Levi

[1985], Kaplan [1996]), numerically sharp degrees of belief are psychologically unrealistic. It is rare, outside casinos, to find opinions that are anywhere near definite or univocal enough to admit of quantification. An agent with a precise credence for, say, the proposition that it will rain in Detroit next July 4<sup>th</sup> should be able to assign an exact “fair price” to a wager that pays \$100 if the proposition is true and costs \$50 if it is false. The best most people can do, however, is to specify some vague range.

While psychological implausibility is one worry, a more decisive problem is that precise degrees of belief are the wrong response to the sorts of evidence that we typically receive. As argued in Joyce [2005], since the data we receive is often incomplete, imprecise or equivocal, the epistemically *right* response is often to have opinions that are similarly incomplete, imprecise or equivocal. Here is a motivating example:

*Black/Grey Coins.* An urn contains a large number of coins which have been painted black on one side and grey on the other. The coins were made at a factory that can produce coins of any bias  $\beta:(1 - \beta)$  where  $\beta$ , the objective chance of the coin coming up black, might have any value in the interval  $0 < \beta < 1$ . You have *no information* about the proportions with which coins of various biases appear in the urn. If a coin is drawn at random from the urn, how confident should you be that it will come up black when tossed?

Proponents of precise models have two very different reactions to this example. They all agree that a rational believer must take a definite stand by having a sharp degree of belief for the event of black, hereafter  $B$ . Formally, the believer's opinions must be captured in a *probability density function* (pdf)  $f_c$  that maps  $\beta$ 's possible values into the non-negative reals subject to  $\int_0^1 f_c(x) dx = 1$ . This pdf fixes a sharp probability for  $B$  via  $c(B) = \int_0^1 x \cdot f_c(x) dx$ , so that the person's credence for  $B$  is her expectation of its objective bias. *Subjectivists* see  $f_c$  and  $c$  as matters of personal judgment based on the believer's inductive hunches, best guesses about similarities among cases, and so on. Your pdf in might be the *uniform* distribution  $c_U$ , which has  $f_U(x) = 1$  everywhere on  $[0, 1]$  and  $c_U(B) = 1/2$ , whereas I might feel that  $\beta$  is skewed toward lower values and have the pdf  $f_c(x) = 42 \cdot x \cdot (1 - x)^5$ , which fixes  $c(B)$  at  $1/4$ . According to the subjectivists, neither of us is irrational. We are each entitled to our own sharp opinions when the data runs out.

*Objectivists* find this absurd. We have no reason to think that any bias is any more or less likely than any other, they say, and epistemic rationality requires us to impartially assign symmetrically supported possibilities identical sharp probabilities. This means that, contra the subjectivists, the *only* rational credences to hold in B/G-Coins are those based on the uniform density  $f_U$ . This reasoning depends ultimately on an application of the famous (or infamous) *Principle of Indifference*.

**POI.** If disjoint events  $E_1$  and  $E_2$  are “equipossible” given all relevant evidence — so that any datum that tells for/against  $E_1$  is offset by a perfectly symmetrically datum that tells for/against  $E_2$  to the same degree — then any rational credence function based on the available evidence must treat  $E_1$  and  $E_2$  symmetrically by assigning them the same sharp probability.

In the extreme case where there is *no* evidence **POI** requires  $c(E_1) = c(E_2)$ . In B/G-Coins, this means that events in any partition  $\beta \in [j/N, j+1/N]$ , for  $j = 0, 1, \dots, N$ , must be assigned the same sharp credences of  $1/N$ , which is only possible if the credence function is generated by  $f_U$ .

Though **POI** has a checkered past [Joyce, 2009], its defenders continue to see it as the only hope for a Bayesianism that does not degenerate into a systematic logic of unsupported personal superstition. When asked to justify **POI**, its defenders typically invoke the *maximum entropy* (MaxEnt) principle [Jaynes, 2003]. MaxEnt shows that the uniform density uniquely minimizes the amount of *extra information* that one must add to get sharp degree of beliefs in cases like B/G-Coins. If we measure the amount of information encoded in density  $f_c$  by its *Shannon information*  $I(f_c) = \int_0^1 f_c(x) \cdot \log_2(f_c(x)) dx$ , then it can be shown that  $I(f_U) < I(f_c)$  for all non-uniform pdfs  $f_c$ . These considerations generalize in significant ways along a range of dimensions. In broadest terms, one starts with the largest family C of credence functions that are not logically excluded by the data, and MaxEnt then requires believers to adopt the unique credence function from C that minimizes Shannon information. Moreover, if someone with sharp prior  $c$  acquires further data that fixes a set of probability functions D, then she should update by shifting from  $c$  to the unique  $d$  from D whose pdf minimizes *Kullback-Leibler cross-information*:  $I(f_c, f_d) = \int_0^1 f_c(x) \cdot \log_2(f_c(x)/f_d(x)) dx$ .<sup>5</sup> Again, the thought is that rationality requires the adoption of sharp credences that encode the minimum amount of information beyond what is explicitly given in the evidence, so as to minimize the extent to which believers “jump to unsupported conclusions.”

Proponents of imprecise credences are unmoved. Even if one grants that the uniform density is the *least* informative sharp credence function consistent with your evidence in Black/Grey Coins, it is still *very* informative. Adopting it amounts to pretending that you have lots and lots of information that you simply don’t have. For example,  $f_U$  commits you to thinking that in a hundred independent tosses of the black/grey coin the chances of black coming up fewer than 17 times is *exactly*  $17/101$ , just a smidgen ( $= 1/606$ ) more probable than rolling an ace with fair die. Do you really think that your evidence justifies such a specific probability assignment? Do you really think, e.g., that you know enough about your situation to conclude that it would be an unequivocal mistake to let \$100 ride on a fair die coming up one rather than on seeing fewer than seventeen blacks in a hundred tosses? Or, to take another example, are you comfortable with the idea that upon seeing black on the first toss you should expect a black

on the second toss with a credence of exactly  $2/3$ , or, more generally, that seeing  $s$  blacks and  $N - s$  greys should lead you to expect a black on the next toss with a probability of precisely  $^{s+1}/_{N+2}$ ? This is Laplace's [1825] *rule of succession*. If you adopt the uniform pdf over  $\beta$ , as advocated by **POI**, you are stuck with it.<sup>6</sup> Again, the evidence you have about the coin's bias (viz., *nada!*) is insufficient to justify such a specific inductive policy. Of course, any sharp credence function will have similar problems. Precise credences, whether the result of purely subjective judgments or "objective" rules like **POI**, always commit a believer to extremely definite beliefs about repeated events and very specific inductive policies, even when the evidence comes nowhere close to warranting such beliefs and policies.

### The Imprecise Model

These sorts of difficulties have led many Bayesians to reject the precise model in favor of less exacting conceptions of graded beliefs. The first such approaches, explored extensively in Fine [1973] represented credal states using a *comparative confidence ranking*. This is a pair of binary relations ( $.>.$ ,  $.\geq.$ ) each defined on  $\Omega$ , where  $X .>. Y$  holds when the believer is determinately more confident in  $X$  than in  $Y$ , and  $X .\geq. Y$  holds when she is determinately no less confident in  $X$  than in  $Y$ . These relations are required to satisfy *the laws of comparative probability*, which can be distilled into a single general axiom:

If  $\langle X_1, \dots, X_N \rangle$  and  $\langle Y_1, \dots, Y_N \rangle$  are sequences of propositions from  $\Omega$  (that may contain repeats), and if the sequences contain the same number of truths as a matter of logic, then  $X_n .\geq. Y_n$  for all  $n = 1, 2, \dots, N - 1$  only if  $Y_N .\geq. X_N$  and not  $X_N .>. Y_N$ .

As shown in Kraft et al. [1959] and Scott [1964], this condition is necessary and sufficient for the existence of a (finitely additive) probability function  $c$  such that  $c(X) > c(Y)$  when  $X .>. Y$  and  $c(X) \geq c(Y)$  when  $X .\geq. Y$ . This function is unique when the ranking is *complete*, so that  $X .>. Y$  or  $Y .\geq. X$  for all  $X, Y \in \Omega$ , and *atomless*, so that  $X .>. F$  implies the existence of a  $Y \in \Omega$  with  $X \& Y .>. F$  and  $X \& \sim Y .>. F$ . The ordinal approach thus includes the precise model as a special case, but generalizes it by allowing for beliefs with less than mathematical precision and by recognizing that a believer might not even be able to make ordinal probability comparisons between some events.

Though more psychologically realistic, the ordinal approach can only represent a limited range of doxastic attitudes. Consider bare judgments of stochastic independence. People often believe that the probabilities of two events are entirely uncorrelated even when they know nothing about the events' chances of coming about. One might, for instance, have no information at all either about how likely it is to rain in Detroit next July 4<sup>th</sup> or about the chances of the Australians winning the Ashes crickett match in 2030, but one might still be

certain that these events have no probabilistic bearing on one another. The usual way of representing stochastic independence is by saying that the probability of one event  $X$  remains the same given the occurrence or non-occurrence of another  $Y$ , so that  $c(X) = c(X | Y) = c(X | \sim Y)$ , in which case learning  $Y$ 's probability tells one nothing about  $X$ 's probability (and conversely). While the ordinal framework can capture some aspects of this relationship (e.g.,  $X \& Y \geq X \& \sim Y$  when  $Y \geq \sim Y$ ), it lacks the resources to fully represent the richness of independence judgments.

One can remedy this deficit by supplementing the simple ordinal model with a third relation that holds between  $X$  and  $Y$  just when the believer sees  $X$  and  $Y$  as stochastically independent of one another. Unfortunately, even this addition still leaves many doxastic attitudes unrepresented. Consider what Couso et al. [2000] call an “unknown interaction” in which a believer is entirely ignorant about the manner in which two events might be correlated. Here is an illustration:

*Game Show.* You are a contestant on a game show. The host shows you two large urns that contain balls marked “\$1000”, “\$500”, and “\$0”. You know *nothing* about the relative proportions of ball types in either urn or about the ways in which their contents might be correlated. The host randomly pulls a ball from one urn, but does not reveal it. He explains that you can either have the amount on that ball plus \$500, or you can randomly draw a ball from the second urn and have the sum of the numbers on the two balls added to your fortune. Should you pick a second ball or stand pat?

You would like to know what you can expect to earn by drawing another ball, but since you are entirely ignorant about the composition of the second urn you have no evidential basis for assigning definite probabilities to “\$1000”, “\$500” or “\$0”. Moreover, since you know nothing about the correlation, you lack the data needed to determine the conditional probabilities  $c(\text{Win } \$(x + y) | \text{Draw } \$x)$  for any  $x, y \in \{1000, 500, 0\}$ . This prevents you from fixing any definite expected utility for your potential acts. Indeed, it is consistent with what little you know that your objective chances of getting \$2000, \$1500, \$1000, \$500 or \$0 after drawing can be *any* 5-tuple of non-negative numbers summing to one. Likewise, if you stand pat then your objective chances of getting \$1500, \$1000, \$500 can be *any* 3-tuple of non-negative numbers summing to one. So, you simply do not know enough about the composition of the urns to know whether taking a second ball is positively relevant, negatively relevant, or irrelevant to your prospects of increasing or decreasing your fortune.

Friends of **POI** will advocate a uniform pdf, so that  $c(\text{Draw } \$x) = c(\text{Win } \$(x + y) | \text{Draw } \$x) = 1/3$  for all  $x, y \in \{1000, 500, 0\}$ . Since you are ignorant about the relative proportions of ball types in the second urn, they say, you should assume them equal. Likewise, since you have no idea how the contents of the two urns might correlate you should assume that your draw and the host's draw are stochastically independent (because you apply

**POI** to the outcome of his draw conditional on the outcome of yours). If you follow this advice then drawing and standing pat will seem equally good: your expected utility for each is determinately \$1000.<sup>7</sup> A more subjectively inclined proponent of precise credences will not insist on uniform distributions, but will say that you should have some sharp values or other for both  $c(\text{Draw } \$x)$  and  $c(\text{Win } \$(x + y) \mid \text{Draw } \$x)$ , thereby committing yourself both to a definite view about the relative proportions of balls in your urn and to a definite view about the statistical relevance of the outcome of your draw to the host's draw.

Neither position is defensible. By hypothesis, you know *nothing* about the contents of either urn or about their correlation. In particular, you have no more or less reason to think that your draw and the host's draw are independent than you have to think they are perfectly correlated, or rigged so that you will end up with zero, or anything else. You simply have no evidence that bears on these issues at all. Postulating sharp values for  $c(\text{Win } \$(x + y) \mid \text{Draw } \$x)$  under such conditions amounts to pulling statistical correlations out of thin air.

Examples like Game Show make it clear that the intuitive notion of one event "providing no information about" another is ambiguous between stochastic independence, which requires a great deal of evidence about probabilistic relationships among events, and "unknown interaction," which involves having no such evidence. Judging  $X$  and  $Y$  to be independent involves believing that you can ignore  $Y$  entirely when making probabilistic inferences about  $X$ . In this context "provides no information about" means "is evidentially irrelevant to". In an unknown interaction it is consistent with what you believe that  $X$ 's truth-value is perfectly correlated with  $Y$ 's truth-value, that the two are perfectly anti-correlated, or that there is any intermediate level of correlation. You lack the information that you need to determine whether you can ignore  $Y$  when making inferences about  $X$ . In this context "provides no information" means "is maximally evidentially ambiguous."

To formulate a theory of graded belief that properly distinguishes these notions we need an *imprecise probability model*. The basics run as follows:

- 1I. A believer's overall credal state can be represented by a *family*  $C$  of credence functions defined on some Boolean algebra  $\Omega$ . Facts about the person's opinions correspond to properties common to all the credence functions in her credal state.
- 2I. If the believer is rational then every credence function in  $C$  is a probability.
- 3I. If a person in credal state  $C$  learns that some event  $D$  obtains (and nothing else), then her post-learning state will be  $C_D = \{c(\bullet \mid D) = c(X) \cdot [c(D \mid X) / c(D)] : c \in C\}$ .
- 4I. A rational decision maker with credal state  $C$  is obliged to prefer one action  $A$  to another  $A^*$  when  $A$ 's expected utility exceeds that of  $A^*$  relative to every credence function in  $C$ .

Each point calls for commentary.

*Commentary on II:* Elements of  $C$  are, intuitively, probability functions that the believer takes to be compatible with her total evidence. Taken collectively they represent her imprecise opinions in light of that evidence. We can illuminate the idea using a picturesque analogy.<sup>8</sup> Think of  $C$  as a huge committee in which each member (= credence function) has a definite degree of confidence for each proposition in  $\Omega$ . The person's credal state is a kind of amalgam of the opinions of her committee members, where the amalgamation process is constrained by Pareto considerations: if all members agree about some matter this reflects a determinate fact about what the person believes. If  $c(Y) > c(X)$  for all  $c \in C$  then she regards  $Y$  as more likely than  $X$ . If  $c(X | Y) > c(X)$  for all  $c \in C$  then she sees learning  $Y$  as increasing her evidence for  $X$ . Issues that divide the credence committee are issues about which the person cannot be said to have any settled view. If  $c(Y) > c(X)$  for some  $c \in C$  and  $c(Y) = c(X)$  for the rest, then the person determinately regards  $Y$  to be at least as likely as  $X$ , but it is not determinate whether she regards  $Y$  as more likely than  $X$  or sees them as equally probable. So, it's not majority rule, and it's not a system in which committee members' views are differentially weighted to force agreement:<sup>9</sup> it's unanimity or ambiguity.

It is sometimes suggested that this model is even more psychologically implausible than its precise cousin since believers must keep track of a (typically infinite) family of credence functions, rather than just one. This is the wrong way to think. Rather than being a model of a believer's psychology, the credal state is a highly formalized representation of her doxastic situation. Though the person's opinions are modeled by the shared properties of her "committee members", she herself will not think in these terms. Instead, she will make qualitative or comparative assessments of probability and utility — that  $X$  is more likely than  $Y$ , that  $X$  and  $Y$  are independent, that  $X$  is the evidence for  $Y$ , that  $A$  is a better act than  $A^*$ , *et cetera* — and these concrete judgments are modeled abstractly by requiring that all  $c \in C$  satisfy certain conditions, e.g.,  $c(X) > c(Y)$ ,  $c(X \& Y) = c(X) \cdot c(Y)$ ,  $c(Y | X) > c(Y)$ ,  $Exp_c(A) > Exp_c(A^*)$ , *et cetera*. The believer only keeps track of her explicitly held qualitative and comparative beliefs: the formal representation takes care of itself.

As with the precise model, there is room here for a lively subjectivist/objectivist debate. Subjectivists will say that rational individuals can have different imprecise belief states even when they face the same objective data. A man in Game Show might have a hunch that the urns are highly correlated, and so have  $c(\text{Win } \$x + x | \text{Draw } \$x) > 3/4$  across his  $C$ , whereas a woman in the same situation may think the urns are anti-correlated and reverse these inequalities. Subjectivists deem either view legitimate. Objectivists will deny this, arguing that there is always a single correct imprecise credal state that a person should hold in light of a given body of data. Presumably, this state would be the one that introduces the least amount of additional information beyond what is in the data. It is not clear how such an "imprecise minimum information requirement" might be formulated, but it seems clear that  $C_1$  encodes more information than  $C_2$  whenever  $C_1 \subset C_2$  or when  $C_2$  arises from  $C_1$  by conditioning.



A few aspects of this debate deserve further comment. The first concerns the relationship between credences and evidence about objective chances. Often data imposes direct constraints on the range of possible chance hypotheses for some event. In B/G-Coins, for example, we are told that coin's bias falls within  $(0, 1)$ , but we might imagine variants in which  $\beta$  is confined to some narrower interval  $(a, b)$ . Precise credence functions then correspond to probability densities defined over the chance hypotheses left open by the evidence. So, in B/G-Coins with  $\beta \in (a, b)$ , each  $c \in C$  has a pdf with  $f_c(x) = 0$  outside  $(a, b)$ .

One vexed question concerns whether considerations beyond knowledge of chances should influence credal states. One position is what Roger White calls the *Chance Grounding Thesis*.

**CGT.** "Only on the basis of known chances can one legitimately have sharp credences. Otherwise one's spread of credence should cover the range of chance hypotheses left open by your evidence." [2010, 174]

In B/G-Coins this says that you should have a sharp credence of  $c(B) = b$  only if you are certain that the chance of black coming up is  $b$ , so that  $c(\beta = b) = 1$ . Otherwise, your credal state should have committee members whose opinions instantiate every pdf over  $(0, 1)$ . While White portrays **CGT** as essential to the imprecise approach, it merely the most extreme of a range of possible positions. Indeed, it is too extreme in one respect since sharp credences are clearly called for in some situations where chances are unknown. Suppose you are told that the urn in B/G-Coins contains a coin of bias  $1 - \beta$  for every coin of bias  $\beta$ . This requires all the pdfs in your credal state to be symmetric, so that  $f_c(x) = f_c(1 - x)$  for all  $x$ . The range of chance hypotheses left open by your evidence is still  $(0, 1)$ , but your committee members are, and should be, unanimous that  $c(B) = 1/2$ .

This brings us to a second question: how should symmetry conditions factor into the imprecise model? Even without direct evidence that forces each function in the credal state to be symmetric, some may insist that asymmetric densities over  $(a, b)$  should be excluded whenever data indicates *only* that the true chance lies in that interval. This is a half-step from POI.<sup>10</sup> Instead of trying to ensure evenhanded beliefs by constituting a credence committee consisting of the single most even-handed member, this proposal aims to achieve the same effect by ensuring that each member is "fair and balanced" the extent of not preferring one side of the interval over the other. As with the uniform distribution, this yields a sharp credence of  $a+b/2$ . An even weaker symmetry requirement, which does not generate sharp credences, is this:

*Symmetry.* If  $\{E_1, \dots, E_N\}$  is a partition of events that are "equipossible" in light of all the available evidence, then a rational believer must treat the  $E_n$  symmetrically in her credal state by including committee members who take all

possible contrary views on their relative probabilities, so that for any  $c \in C$  and any permutation  $\sigma$  of  $\{1, 2, \dots, N\}$  there is always a  $c_\sigma \in C$  with  $c_\sigma(E_n) = c(E_{\sigma(n)})$ .

Like **POI**, this requires believers to afford equal treatment to possibilities not discriminated by the data. However, “equal treatment” does not mean “equal sharp probability” or even “symmetrical imprecise probability”, but something more like “equal strength of support across the committee.” Picturesquely, instead of trying to ensure an impartial belief by putting the matter in the hands of the single most even-handed committee member, or in the hands of “fair and balanced” members who are equally inclined toward either side of the issue, Symmetry ensures impartiality by seeing to it that each “unbalanced” committee member is opposed by another who is exactly as radical, relative to the committee average, but in a contrary direction. In **B/W-Coins** with  $\beta \in (a, b)$  this ensures that every “left winger” whose  $f_c$  generates a credence  $c(B)$  between  $a$  and the midpoint  $^{a+b}/_2$  is precisely offset by a “right winger” with  $f_{c^*}(x) = f_c(a + b - x)$  whose credence  $c^*(B)$  falls between the midpoint and  $b$ . Unlike the other two cases, Symmetry does not entail sharp credences: every value in  $(a, b)$  coincides with the credence of some committee member.

A final question along these lines concerns the appropriate level of “sharpness” for a credal state. Many versions of the imprecise model (e.g. Joyce [2005]) assume with **CGT** that if the data indicates *only* that the chance of an event is in  $(a, b)$  then the credal state should contain credences for the event that cover  $(a, b)$ . The most natural way to ensure this (and satisfy Symmetry) is by having a credal state that reflects every pdf defined over the interval, including those that place almost all their weight close to the ends and those that place it all on one specific value. This is a plausible view (and *I* personally think it is the only view that makes epistemological sense), but it has costs that some proponents of the imprecise model might be unwilling to bear. For example, it precludes inductive learning in situations of extreme ignorance. Suppose you (independently) toss the black/grey coin a thousand times and see 500 heads. This seems like overwhelming evidence for thinking that the coin is very, very nearly fair, and so seems like a spectacular justification for confining your credence in  $B$  to some very, very narrow interval around  $^{1}/_2$ . But, this sort of learning is impossible if your credal state reflects every pdf on  $(0, 1)$ . Nearly all your committee members will have an initial pdf  $f_c(\bullet)$  that updates to a posterior pdf  $f_c(\bullet \mid 500 \text{ black})$  that produces a credence for *black* that is closer to  $^{1}/_2$ . That sounds like learning! Unfortunately, two countervailing considerations entirely nullify this effect. First, “pigheaded” pdfs that concentrate all their weight in some proper subinterval  $(a, b)$  of  $(0, 1)$  will never assign  $B$  a credence outside that interval: for them,  $c(B \mid 500 \text{ black}) \in (a, b)$ . If  $(a, b)$  is bounded away from  $^{1}/_2$ , then there is no way for  $c(B \mid 500 \text{ black})$  to approach  $^{1}/_2$ . Second, and in some ways worse, for every non-pigheaded committee member  $f_c$  who moves her credence closer to  $^{1}/_2$  there will always be a more extreme member whose

pdf  $f_{c^*}$  is concentrated more heavily toward one end or the other of  $(0, 1)$  and is such that  $f_{c^*}(\bullet \mid 500 \text{ black}) = f_c(\bullet)$ . As each “extremist” finds her views tempered by the data, an even more radical extremist slides in from the wings to take her place. So, while each non-pigheaded committee member becomes more convinced that the coin is fair, your credal state as a whole remains exactly where it was!

There are two ways for proponents of the imprecise model to respond to this result. Purists will say that if you really know *nothing* about the black/white coin’s bias, then you also really know *nothing* about how your opinions about  $B$  should change in light of frequency data. For each  $0 < \delta < 1/2$  you have a committee member who feels that your credence for  $B$  should move exactly  $\delta$  probability units toward  $1/2$  given the data 500 *heads*. So, your views about the evidential relevance of this data are maximally imprecise, which means that your credence for  $B$  should remain imprecise as well even after taking the data into account. You cannot learn anything in cases of pronounced ignorance simply because a prerequisite for learning is to have prior views about how potential data should alter your beliefs, but you have no determinate views on these matters at all.

The alternative response involves taking a small step in the direction of precision. One might think that the most pigheaded and extreme committee members — those that place all their credal density on a proper subinterval of  $(0, 1)$  or those that focus almost entirely on the ends of the interval — represent positions of such fanatical overconfidence that they should be discounted. Perhaps the right way to secure inductive learning is to *sharpen* your credal state by (a) throwing out all the pigheaded committee members and insisting that each  $c$  be based on a pdf that assigns some positive credence to every event of the form  $a < \beta < b$ , and (b) silencing “extremist” elements by insisting that each committee member assign a credence to  $B$  that falls within some sharpened interval  $(c_-, c^+)$  with  $c_- > 0$  and  $c^+ < 1$ . (Symmetry would dictate  $c_- = 1 - c^+$ .) Inductive learning is then possible: the width of the interval for  $c(B \mid 500 \text{ black})$  will be narrower than  $(c_-, c^+)$ , and the narrower the prior interval is the faster the posterior interval shrinks. (In a sense, the amount of sharpening one is willing to tolerate can be seen as an indicator of one’s natural inductive boldness.) The basic strategy generalizes to cases where the objective chance of some event  $X$  lies in an arbitrary interval  $(a, b)$ . Here, sharpening involves starting with the set of all densities over  $(a, b)$ , eliminating the pigheaded ones, and then requiring that every pdf in the credal state to generate a credence for  $X$  that falls within a subinterval of  $(a, b)$ , so that more extremist members, relative to the group norm, are eliminated. In keeping with Symmetry, one might preserve balance by ensuring that each “left winger” with  $c(X) = 1/2 \cdot (a + b) - \varepsilon$  who is ushered off is accompanied by a “right winger” with  $c^*(X) = 1/2 \cdot (a + b) + \varepsilon$ , so that the interval’s midpoint is preserved.

While sharpening has some attractive features, purist defenders of imprecision will have difficulty detecting a coherent rationale for the strategy. Why

banish extremists? Why not ditch moderates? There seems to be no reason *based in the evidence* for doing one rather than the other. Likewise, what principled reason is there to confine credences to one proper subinterval of  $(a, b)$  as opposed to another? Whichever way you go, you end up acting as if you have evidence that you do not actually possess. I will not try to adjudicate these issues here, except to say (i) that my sympathies lie with the purists, but (ii) the sharpening strategy (as we shall see) may play a role in practical decision making even if it is defective as epistemology.

*Commentary on 2I:* Requiring each  $c \in C$  to be a probability forces graded beliefs to have some nice normative properties. First, it entails that  $C$ 's associated confidence ranking, defined by  $X \succ Y$  iff  $c(X) > c(Y)$  for all  $c \in C$  (and likewise for  $\succ$ ), is a comparative probability. This means, for example, that a person who is more confident in  $X$  than in  $Y$  is thereby committed, whether she knows it or not, to being more confident in  $X \vee Z$  than in  $Y \vee Z$  when  $Z$  is contrary to  $X$  and  $Y$ . Second, 2I guarantees that each  $X$  in  $\Omega$  has a coherent *lower* and *upper credence*. These are the lower and upper bounds of the values that elements of  $C$  assign to  $X$ :  $C_-(X) = \inf\{c(X) : c \in C\}$  and  $C^+(X) = \sup\{c(X) : c \in C\}$ . It follows from 2I that  $C_-(X) \leq 1 - C_-(\sim X) = C^+(X)$ , and that  $C_-(X \vee Y) \geq C_-(X) + C_-(Y)$  when  $X$  and  $Y$  are contraries. One can think of  $C_-(X)$  and  $C^+(X)$  as the lowest/highest probability for  $X$  that the evidence permits, according to a conception of evidence on which data tells for/against  $X$  to exactly the same degree it tells against/for  $\sim X$ . Upper and lower probabilities are useful, in part, because they reflect limitations on a believer's epistemic and practical commitments. To see how imagine, as Bayesians often do, that a person with a sharp credence for  $X$  will always set a "fair price" of  $c(X)$  utiles on a wager [ $u(X) = 1, u(\sim X) = 0$ ] that pays one utile if  $X$  and nothing if  $\sim X$ . It is then possible to show that a person with imprecise credences that satisfy 2I can avoid becoming the victim of a "Dutch book" by having the policy of never paying more than  $C_-(X)$  utiles to buy the wager and never selling it for less than  $C^+(X)$  utiles. In other words, she can avoid accepting a series of wagers that are certain to leave worse off in the aggregate merely by paying prices that all her "committee members" regard as fair or advantageous.

*Commentary on 3I.* According to 3I, updating in light of data involves having each "committee member" update on the data individually, and taking the posterior to be the new, better informed committee. There is an alternative model of updating, I have heard suggested, which simply eliminates those probabilities in  $C$  that are incompatible with the data received. On this proposal, if a person with credal state  $C$  learns  $D$  (and nothing else), then her post-learning credal state is  $C \cap D = \{c(\bullet) : c \in C \text{ and } c(D) = 1\}$ . So, rather than reporting the new evidence to the committee and letting members modify their opinions accordingly, members who did not anticipate the data are shown the door and the rest, whose views do not change at all, constitute the new

committee. This model seems to be based on a confusion between learning that some event has occurred and learning that its prior objective chance of occurring was 1.

To see this point, suppose you are facing two urns each of which contains five black or grey balls. You know that there is one more black ball in urn<sub>1</sub> than there is in urn<sub>2</sub>, but otherwise you know nothing else about either urn's contents. A coin will be tossed, and a ball will be drawn randomly from urn<sub>1</sub> if heads and from urn<sub>2</sub> if tails. You know that the coin is biased toward heads, but for all you know the chances of a head coming up might be any number in  $(\frac{1}{2}, 1]$ . Without knowing how the coin fell, you learn that a black ball was drawn. How confident should you be that the coin landed heads? Suppose your credal state  $C$  contains every probability function defined over the events  $\langle n \rangle$  &  $\pm Heads$  &  $\pm Black$  such that  $c(Black \mid \langle n \rangle \ \& \ Heads) = \frac{n}{5}$  and  $c(Black \mid \langle n \rangle \ \& \ \sim Heads) = \frac{n-1}{5}$ , where  $\langle n \rangle$  means that  $Urn_1$  contains  $n \in \{1, 2, \dots, 5\}$  black balls. Learning *Black* is moderately informative about *Heads* on 3I's model: Upon conditioning, each committee member raises her credence for *Heads* a little bit, and  $C_{Black}$  ends up containing credence functions whose values for *Heads* cover the smaller interval  $(\frac{5}{9}, 1]$ . For "show 'em the door" updating, however, learning *Black* is *maximally* informative about the coin toss. Since the *only* credence in  $C$  for which  $c(Black) = 1$  also has  $c(Heads) = 1$  and  $c(\langle 5 \rangle) = 1$ , learning that ball was black licenses the conclusion that the coin surely came up heads, clearly the wrong answer. Of course, this would be the right answer if you had learned that the *initial* (pre-draw) chance of *Black* was one, but this is not what you learned.

*Commentary on 4I.* 4I lays down a *sufficient* condition for imprecise preferences. In the same way that each event in  $\Omega$  has an upper and lower probability, each prospect  $A$  that produces an outcomes of utility  $u_A(X_n)$  across some partition  $\{X_1, X_2, \dots, X_N\}$  will have an upper and a lower expected utility, with  $Exp^+(A) = -Exp_-(\sim A)$ . 4I says that a rational agent will definitely prefer  $A$  to a payment of  $u < Exp_-(A)$  utiles and definitely prefer a payment  $u > Exp^+(A)$  utiles to  $A$ . Questions about payments in the range between  $Exp_-(A)$  and  $Exp^+(A)$  are left unresolved. More generally, if the credence committee members unanimously agree that one prospect is better or worse than another then 4I requires the agent to have that preference, but when there is disagreement it is ambiguous what the agent should do. There is no consensus among proponents of the imprecise model about what choices are permissible or impermissible in the ambiguous region. We shall be discussing these issues more fully below.

### **What Does the Imprecise Model Represent?**

Before addressing objections to the imprecise model we should consider another issue that divides its proponents. As with any attempt to represent

some phenomenon mathematically, it is critical to figure out which aspects of a representation reflect the reality being modeled and which are artifacts of the formalism. In particular, we need to know which features of a believer's credal state actually convey facts about her beliefs. One view has it that the useful information in  $\mathbf{C}$  is exhausted once we know the upper and lower probabilities that  $\mathbf{C}$  associates with the propositions in  $\Omega$ . More formally, the view is this:

**Lower Probability (LP).** If credal states  $\mathbf{C}$  and  $\mathbf{C}^*$  assign the same lower (hence upper) probabilities to all events in  $\Omega$ , then  $\mathbf{C}$  and  $\mathbf{C}^*$  encode the same beliefs about  $\Omega$ .

Some, e.g., Henry Kyburg of [1983], have gone farther and advocated jettisoning  $\mathbf{C}$  altogether and thinking of the believer as having a single credence function that assigns *intervals*, rather than sharp numbers, to propositions. Others have suggested that following the slightly more general view:

**PSET.** If  $\mathbf{C}$  and  $\mathbf{C}^*$  assign the same range of values to all events in  $\Omega$ , so that the sets  $\mathbf{C}(X) = \{c(X) : c \in \mathbf{C}\}$  and  $\mathbf{C}^*(X) = \{c^*(X) : c^* \in \mathbf{C}^*\}$  are identical for all  $X \in \Omega$ , then  $\mathbf{C}$  and  $\mathbf{C}^*$  encode the same beliefs about  $\Omega$ .

White [2010, p. 173] seems to accept **PSET**, writing, “often we are just interested in the spread of values in our [credal state] for a particular proposition... I will speak of one's credence in a proposition as being possibly a set of values.” He also uses the notation  $\mathbf{C}(X) = \{c(X) : c \in \mathbf{C}\}$  and speaks of *the* credence of an event when he clearly means the range of its probabilities in  $\mathbf{C}$ .

Unfortunately for **LP** and **PSET**, there are cases in which  $\mathbf{C}$  and  $\mathbf{C}^*$  clearly capture distinct beliefs despite generating identical sets of probabilities for all  $X \in \Omega$ . Here is an example:

*Three-sided Die.* Suppose  $\mathbf{C}$  and  $\mathbf{C}^*$  are defined on a partition  $\{X, Y, Z\}$  corresponding to the result of a roll of a three sided-die. Let  $\mathbf{C}$  contain all credence functions defined on  $\{X, Y, Z\}$  such that  $c(Z) \geq 1/2$ , and let  $\mathbf{C}^*$  be the subset of  $\mathbf{C}$  whose members also satisfy  $c(X) = c(Y)$ .

It is easy to show that  $\mathbf{C}$  and  $\mathbf{C}^*$  generate the same range of probabilities for all Boolean combinations of  $\{X, Y, Z\}$ , and so **LP** and **PSET** deem them equivalent. But, they are surely different: the  $\mathbf{C}^*$ -person believes everything the  $\mathbf{C}$ -person believes, but she also regards  $X$  and  $Y$  as equiprobable.

One can avoid this particular problem by focusing more broadly on the ranges of *expected values* that credal states generate (as in Walley [1991]). These expectations are associated with random variables of the form  $g(X_n) = a_n \in \mathfrak{R}$  with  $\{X_1, X_2, \dots, X_N\}$  a partition from  $\Omega$ . As before, one can either focus

on the upper and lower values of such expectations or on their entire ranges. Taking the later approach yields:

**ESET.**<sup>11</sup> If **C** and **C\*** assign the same range of expected values to all random variables defined on  $\Omega$ , so that the sets  $\mathbf{C}(g) = \{Exp_c(g) : c \in \mathbf{C}\}$  and  $\mathbf{C}^*(g) = \{Exp_{c^*}(g) : c^* \in \mathbf{C}^*\}$  are always identical, then **C** and **C\*** encode the same beliefs about  $\Omega$ .

This avoids the Three-sided Die problem since **C** and **C\*** assign different ranges of expected values to the function  $(g(X) = 1, g(Y) = -1, g(Z) = 0)$ , viz.,  $\mathbf{C}(g) = [-1/2, 1/2]$  and  $\mathbf{C}^*(g) = 0$ .

Despite this success, there are still other cases in which **C** and **C\*** capture distinct beliefs despite generating identical expectations. Consider the following example:

*Complementarity.* A black/grey coin will be tossed followed by a head/tail coin. You and I know nothing about the either coin's bias except that there is some chance of all four outcomes. So, our credal states contain probability functions that assign *H* and *B* every value in  $(0, 1)$ . Compatible with our shared ignorance, we have divergent opinions about the stochastic connection between the tosses. You regard them as independent, i.e., you think the probability of seeing a head on the second toss is the same whether black or grey comes up. So, every probability in your credal state satisfies  $c(H | B) = c(H | \sim B)$ . In contrast, I treat heads as *complementary* to black (my term), which means that I think the probability of seeing a head after seeing black is the reverse of the probability of seeing a head after seeing grey. So, every probability in my credal state satisfies  $c(H | B) = c(\sim H | \sim B)$ . Here is a picture of the situation, where the area of each region corresponds to its probability:

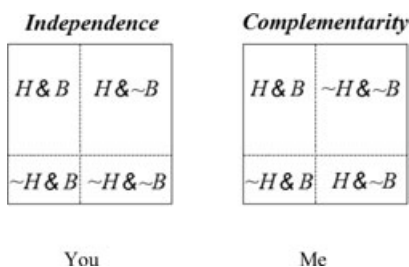


Figure 1

The broken vertical and horizontal lines in each diagram vary independently, and one can place them anywhere within the *interior* of the boxes and remain within the relevant credal state. By moving lines up and down and left and right it becomes easy to see that the range of probabilities for Boolean combinations of the four events are the same in either picture: indeed, the range is always  $(0, 1)$

(except for T and  $\perp$ ). Likewise, if we assign a numerical value to each quadrant and take expectations we can obtain any value in the proper span of the four values using either picture. According to **ESET**, then, you and I have the same beliefs. This manifestly wrong! You regard the *product* of the probabilities of  $H \& B$  and  $\sim H \& \sim B$  as equal to the *product* of the probabilities of  $\sim H \& B$  and  $H \& \sim B$ , while I see the *ratio* of the probability of  $H \& B$  to that of  $\sim H \& \sim B$  as equal to the *ratio* of the probability of  $\sim H \& B$  to that of  $H \& \sim B$  (even though neither of us assigns any of these four propositions any specific probability).

The moral here is that there is more to our doxastic attitudes than can be represented by imprecise models that accept **ESET**, **PSET** or **LP**. An adequate model of imprecise opinion should be able to capture the difference between independence and complementarity,<sup>12</sup> and the only way to do this is by recognizing that certain sorts of beliefs will only be revealed by functional relationships that hold among the credence functions in one's credal state, relationships that are obscured if we focus only on the sets of probabilities assigned to propositions or the sets of values assigned to random variables. Here are some of the functional relationships I have in mind:

- $X_1, X_2, \dots, X_N$  are probabilistic *contraries*:  $\sum_n c(X_n) = \text{constant} < N$  for all  $c \in \mathbf{C}$ .
- $X_1, X_2, \dots, X_N$  are probabilistic *subcontraries*:  $\sum_n c(X_n) = \text{constant} > 0$  for all  $c \in \mathbf{C}$ .
- $X$  is *determinately more probable than*  $Y$ :  $c(X) > c(Y)$  for all  $c \in \mathbf{C}$ .
- $X$  and  $Y$  are *stochastically independent*:  $c(X | Y) = c(X | \sim Y)$  for all  $c \in \mathbf{C}$ .
- $X$  is *complementary* to  $Y$ :  $c(X | Y) = c(\sim X | \sim Y)$  for all  $c \in \mathbf{C}$ .

These are all legitimate epistemic attitudes, and while some can be captured within a framework that accepts **ESET** the only way to represent them all is by allowing  $\mathbf{C}$  to encode facts about beliefs that go well beyond that is found in sets of probability values and expectations.

### Epistemological Objections to the Imprecise Model

White [2010] seeks to undermine the idea of imprecise credences using a series of examples in which the imprecise model seems to fail from an epistemological perspective. Here is perhaps the most interesting (slightly retold):

**Coin Game.** I have a *fair* coin with heads on one side and tails on the other, and a coin of *entirely unknown bias* that is black on one side and grey on the other. Since you know the head/tail coin is fair, each  $c$  in your credal state will satisfy



$c(H) = 1/2$ , where  $H$  says that the coin comes up heads on its next toss. Since you know nothing about the black/grey coin the imprecise interpretation says that there will be a  $c$  in your credal state with  $c(B) = x$  for every  $0 < x < 1$ . I will toss the coins, independently, and observe the results without showing them to you. I will then tell you something about how the outcomes are correlated by reporting either " $H \equiv B$ " (if I see head and black or tail and grey), or " $H \equiv \sim B$ " (if I see head and grey or tail and black). Since the head/tail coin is fair and tosses are independent, you anticipate these reports with equal probability:  $c(H \equiv B) = c(H \equiv \sim B) = 1/2$  for all  $c \in C$ . The case in which you learn  $H \equiv B$  can be pictured as follows:

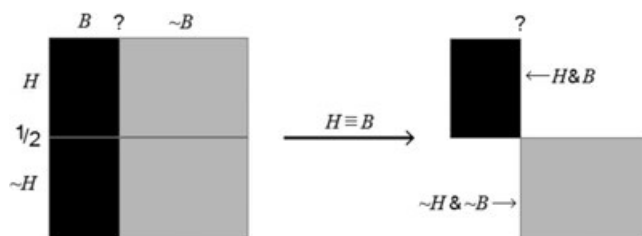


Figure 2

The horizontal line separating heads and tails in the left-hand diagram is accurately placed: it reflects your initial knowledge that the head/tail coin is fair. We will discuss the status of the vertical line separating black and grey shortly. White's question is this: How confident should you be in  $H$  after you hear me say " $H \equiv B$ "?

Some things are obvious. First, learning  $H \equiv B$  shifts you from  $C$  to  $C_{H \equiv B} = \{c(\bullet | H \equiv B) : c \in C\}$ , which means that your posterior beliefs about  $H$  and  $B$  will be identical since the laws of probability require  $c(H | H \equiv B) = c(B | H \equiv B)$  for all  $c \in C$ . Second, since the tosses are independent and  $c(H) = 1/2$  it follows that  $c(B | H \equiv B) = c(H) \cdot c(B) / [c(H) \cdot c(B) + c(\sim H) \cdot c(\sim B)] = c(B)$ .<sup>13</sup> Thus, we have

♣ For every  $c \in C$ ,

- the credence that  $c$  assigns to  $B$  upon learning  $H \equiv B$  is identical to the *prior* credence that  $c$  assigns to  $B$ ,  $c(B | H \equiv B) = c(B)$ ;
- the credence that  $c$  assigns to  $H$  upon learning  $H \equiv B$  is identical to the credence  $c$  assigns to  $B$  upon learning  $H \equiv B$ ,  $c(H | H \equiv B) = c(B | H \equiv B)$ .

This is surprising! Learning that  $H$  and  $B$  are perfectly correlated provides you with *no relevant evidence* about  $B$ , but it forces you to conform your beliefs

about the outcome of the head/tail toss (about which you know a lot) to your prior beliefs about  $B$  (about which you know nothing).

People have three sorts of reactions to Coin Game, each reflecting a different interpretation of the vertical line in FIGURE 2. A precise Bayesian of the objectivist school will say that the example only seems puzzling because the figure is so misleadingly drawn. The vertical line should be in the center to indicate that you should invoke **POI** and settle on a determinate probability of  $1/2$  for  $B$ . Then, the identities in ♣ are exactly what one would expect, since your credences about  $H$  and  $B$  (and their logical combinations) are exactly as they would be if you took the black/grey coin to be fair:  $c(H \& B) = c(H \& \sim B) = c(\sim H \& B) = c(\sim H \& \sim B) = 1/4$ .

Precise subjectivist Bayesians also find nothing puzzling in White's example. They see the vertical line in FIGURE 2 as a personal choice to be made on the basis of inductive hunches, gut feelings, or whatever. While rationality requires you to draw a single vertical line at some point along the horizontal axis no particular point is mandated. Once  $c(B)$  has a definite value, however, the identity  $c(B \mid H \equiv B) = c(B)$  has a compelling rationale. Bayes' Theorem entails that learning  $H \equiv B$  can only alter the evidence for or against  $B$  if there is a difference between  $H \equiv B$ 's probability given  $B$  and its probability given  $\sim B$ . But, since you take heads and tails to be equiprobable and since you see the tosses as independent, there is no difference. If I tell you black came up, your credence for  $H \equiv B$  is just your credence for heads. If I tell you grey came up, your credence for  $H \equiv B$  is just your credence for tails. Either way, the credence is  $1/2$ , which means that the biconditional's truth or falsity is irrelevant to  $B$ 's probability. It should also be clear why your credence for heads should shift so as to coincide with your sharp credence for black. It helps to think sequentially: suppose that you first learn  $H \equiv B$ , and so become convinced that  $H$  and  $B$  are equiprobable. If you then discover that  $B$ 's probability (in light of the correlation) is precisely  $p$ , you thereby come to know that  $H$ 's probability is  $p$  as well. If, conditional on  $H \equiv B$ , you have a precise credence for  $B$  then you should have the same precise credence for  $H$ .

While Coin Game poses no problems when credences are precise, it has disconcerting consequences for imprecise models. Suppose your credal state  $C$  is a set of probabilities whose values for  $B$  range over all of  $(0, 1)$ . This amounts to treating the vertical line in FIGURE 2 as an arbitrary exemplar of some infinite family of lines which, in the aggregate, represents your imprecise state of confidence. If the line can be drawn anywhere, then the ♣-identities entail that your posterior  $C_{H \equiv B}$  will contain functions whose values for both  $B$  and  $H$  range over all of  $(0, 1)$ . For  $B$  this is no surprise. Since you started with imprecise beliefs about  $B$ , and since (as explained above) learning  $H \equiv B$  when  $c(H) = 1/2$  provides no relevant evidence about  $B$ , it stands to reason that you should wind up imprecise about  $B$ . What is surprising is that your credence for  $H$  goes from being sharply  $1/2$  in  $C$  to being spread over the whole

of  $(0, 1)$  in  $C_{H \equiv B}$ ! This *probabilistic dilation* has been studied by Seidenfeld and Wasserman [1993]. It is entirely unavoidable on any view that allows imprecise probabilities.

White sees this as a serious problem for the imprecise model. In addition to having deep conceptual problems, he argues, dilation is the wrong response to Coin Game. The right response, he feels, lies in recognizing that

- Neither  $H \equiv B$  nor  $\sim H \equiv B$  is the least bit are evidentially relevant to  $H$  when you are entirely ignorant about  $B$ .<sup>14</sup>
- Given this, you should retain a determinate credence of  $1/2$  for heads upon learning  $H \equiv B$ , thereby keeping your credence for  $H$  aligned with its know objective chance.

When coupled with the ♣-identities this requires your *prior* credence for  $B$  to be  $1/2$ , a result that White and others who endorse **POI** are happy to accept.

I shall argue, to the contrary, that dilation is the only proper response the imprecise evidence that you receive in Coin Game. My main claims will be:

- Both  $H \equiv B$  nor  $\sim H \equiv B$  are highly evidentially relevant to  $H$  even when you are entirely ignorant about  $B$ . It only seems otherwise because their evidential impact on  $H$ 's probability is indeterminate, both in magnitude and direction.
- Given this, it would be wrong to retain a credence of  $1/2$  for  $H$  upon leaning  $H \equiv B$  or  $\sim H \equiv B$ . Instead, you should have an indeterminate credence for  $H$  that reflects the indeterminate evidential import of your new evidence.

Let me start by explaining where White's arguments go wrong, since this will teach us some surprising things about the resources of the imprecise model.

White first challenges the idea that  $H \equiv B$  can provide evidence about  $H$ . He recognizes that *if you have evidence about B* then you may have to revise your credence in  $H$  upon learning  $H \equiv B$ . For instance, if you know the black/grey coin has bias  $\beta$  toward black then your credence for  $H$  must change from  $1/2$  to  $\beta$  upon learning  $H \equiv B$ . In the terminology of Lewis (1980), knowledge of  $\beta$  makes the information  $H \equiv B$  *inadmissible* with respect to head/tail coin toss: it is direct evidence about the toss's *outcome* which does not alter your views about the coin's fairness. Possession of inadmissible information undermines the chance-to-credence link enshrined in the *Principal Principle*.

- (**PP**) Your current credence for  $X$  should coincide with your current expectation of  $X$ 's time- $t$  objective chances, so that  $c(X \mid \text{chance}_t(X) = p) = p$  whenever  $c(\text{chance}_t(X) = p) > 0$ .

As Lewis recognized, **PP** only holds subject to the proviso that *you do not possess any information that is inadmissible with respect to  $X$ 's time- $t$  chances*.<sup>15</sup> Such information, being directly about  $X$ 's outcome (or later chances) makes  $c(X) = p$  insufficient, all on its own, to justify a credence assignment of  $c(X) = p$ . As everyone agrees, this would be your situation if you initially knew that the black/grey coin had a specific bias  $\beta$  and later came to learn  $H \equiv B$ .

In Coin Game, however, you know nothing about the black/grey coin's bias. This entirely changes the situation, according to White. He argues as follows (p. 177–178):

- You started out knowing that the head/tail coin was fair.
- “*You haven't a clue whether as to whether  $B$* ”.
- If “*you haven't a clue whether as to whether  $B$* ,” then learning  $H \equiv B$  provides “nothing to suggest that the coin landed heads or that it landed tails.”
- Hence, you should “just ignore this useless bit of information and keep your credence in heads at  $1/2$ ,” aligned with the known initial objective chances.

The key move is the third step, which implies that  $H \equiv B$  is admissible when you know nothing about  $B$ 's initial ( $t = 0$ ) chances. Though White is not explicit about why he thinks that  $H \equiv B$  “suggests nothing” about  $H$  in this situation, his view does have a lot of intuitive force. If there is any inadmissible evidence in  $H \equiv B$  it is sure hard to spot since, in the absence of evidence about  $B$  or  $\sim B$ ,  $H \equiv B$  seems to be entirely uninformative about  $H$ . How can you learn anything about an event by discovering how it correlates with another event that you know nothing about?

This plausible sounding reasoning hinges on a subtle equivocation. As we have seen, there are two ways in which one event can be “uninformative about” another: the two might be stochastically independent or it might be in an “unknown interaction” situation. Regarding  $H \equiv B$  and  $H$  as independent in Coin Game means having a credal state in which  $c(H | H \equiv B) = c(H) = c(B) = 1/2$  holds everywhere. While proponents of **POI** will find this congenial, friends of imprecise probabilities will rightly protest that there is no justification for treating the events as independent. You have no evidence to suggest *anything* about the relevance of  $H \equiv B$  to  $H$ . For all you know, the black/grey coin might be strongly biased toward black, in which case  $H \equiv B$  tells heavily in favor of  $H$ . It might be strongly biased toward grey, in which case  $H \equiv B$  tells heavily against  $H$ . Or, it might have an intermediate bias, and so be of intermediate evidential significance for  $H$ . Behaving as if these two events are independent amounts to pulling a statistical correlation out of thin air! The fact is that your evidence leaves the relationship between  $c(H | H \equiv B)$  and  $c(H)$  entirely open.

This helps to explain why the inadmissible information is so difficult to spot. When  $\mathbf{C}$ 's values for  $c(H | H \equiv B)$  span all of  $(0, 1)$ , the "committee members" all *agree* that  $H \equiv B$  is inadmissible with respect to  $chance_0(H)$ , but *disagree* about *how much and in which direction* this data should alter  $H$ 's probability. To see why, it helps to remind oneself that  $H \equiv B$  is direct evidence about the outcome of the head/tail toss. (Remember, I see how the coins fall before making my report.) This is reflected in your credal state by the following identities, which hold all across  $\mathbf{C}$  when  $x = 1/2$ :

$$(\$) c(chance_0(H) = x | H \equiv B) = c(chance_0(H) = x) = 1$$

$$c(H | chance_0(H) = x \ \& \ H \equiv B) = c(H | H \equiv B).$$

In fact, something stronger is true. Even if we alter the example, so that you start out uncertain about the head/tail coin's bias, (\$) will still hold whenever  $c(chance_0(H) = x) > 0$ . On this matter your credal committee speaks with one voice:  $H \equiv B$  is inadmissible with respect to *any* initial chance assignment for  $H$ , and the initial chance of heads ceases to be of evidential consequence once  $H \equiv B$  becomes known. This unanimity tends to be obscured, however, by the members's disparate views about the *impact* of  $H \equiv B$  on  $H$  (owing to their divergent views about  $B$ ). Since  $c(H | H \equiv B) = c(B)$  holds across  $\mathbf{C}$ , those with unconditional probabilities for  $B$  above  $1/2$  will say that learning  $H \equiv B$  raises  $H$ 's probability; those with unconditional probabilities for  $B$  below  $1/2$  will say that  $H \equiv B$  lowers  $H$ 's probability; those with  $c(B) = 1/2$  will say that  $H$ 's probability remains fixed.<sup>16</sup> Just as it is easy to confuse unknown interaction with stochastic independence, it is easy to confuse ambiguity in evidential impact with lack of evidential impact. This is exactly what White seems to have done, misclassifying inadmissible information of (maximally) imprecise evidential significance as admissible information of no evidential significance at all. Once one appreciates that  $H \equiv B$  is inadmissible, however, it becomes more plausible that  $H$ 's credence should dilate in response to  $H \equiv B$ 's ambiguous message. This consideration is not decisive (see next paragraph), but what is clear is that the initial fairness of the head/tail coin cannot justify a posterior credence of  $1/2$  all of its own, as White's argument suggests.

Even if  $H \equiv B$  does trump your knowledge of  $H$ 's initial chance, it does not automatically follow that it is evidentially relevant to  $H$ . Perhaps the proper response to data of maximally ambiguous evidential import is to treat it as having no impact at all. This is what one would think if, consonant with **POI**, one felt that each  $c \in \mathbf{C}$  is "cancelled out" by a complementary  $c^*$  such that  $c^*(H | H \equiv B) = 1 - c(H | H \equiv B)$ , so that silence results when committee members scream equally loudly in favor of contrary conclusions. A plausible seeming argument for this view appeals to the *likelihood principle*, which, for current purposes, says that  $H \equiv B$  can only be evidence for  $H$  if there is some difference between  $H \equiv B$ 's probability conditional on  $H$  and its

probability conditional on  $\sim H$ . For a sharp credence function  $c$ , this means that, by  $c$ 's lights,  $H \equiv B$  is only evidence for or against  $H$  if  $c(H \equiv B | H) \neq c(H \equiv B | \sim H)$ . The imprecise analogue is the following:

*Likelihood.* According to an imprecise credal state  $C$ ,  $H \equiv B$  is evidentially relevant to  $H$  if and only if  $C_H$  and  $C_{\sim H}$  encode different beliefs about  $H \equiv B$ .

If we accept this it looks like the game is up.  $C_H$  and  $C_{\sim H}$  both contain credence functions that assign  $H \equiv B$  every probability in  $(0, 1)$  and so, one might argue, they encode the same beliefs about  $H \equiv B$ . It then follows from Likelihood (which we should accept) that  $H \equiv B$  is not evidence for or against  $H$ .

The flaw in this reasoning lies in its assumption that credal states encode the same beliefs about an event when they assign the same range of values to that event, in particular that  $C_H$  and  $C_{\sim H}$  encode the same beliefs about  $H \equiv B$  because  $C_H(H \equiv B) = C_{\sim H}(H \equiv B) = (0, 1)$ . This is just **ESET**, a view we have already rejected. In fact, from  $C$ 's perspective,  $C_H$  and  $C_{\sim H}$  encode *complementary* beliefs about  $H \equiv B$ , even though each is maximally imprecise about its probability.  $C$ 's committee members agree that  $c(H \equiv B | H) = 1 - c(H \equiv B | \sim H)$ , and so are unanimous in thinking that learning  $H$  will increase/decrease  $H \equiv B$ 's probability to the same degree that learning  $\sim H$  will decrease/increase its probability (while disagreeing entirely about the magnitude and direction of these changes). Your credal attitudes toward  $H \equiv B$  upon learning  $H$  and upon learning  $\sim H$  are thus mirror images of one another, equally imprecise but flipped right to left. While this difference in orientation does not show up in the range of credence values for  $H \equiv B$ , it appears in other ways. For example, all committee members will agree that the following wagers have the same expected utility for any values of  $x_1, x_2, y_1, y_2$ :

Table 1

	$(H \equiv B) \& H$	$(H \equiv \sim B) \& H$	$(H \equiv B) \& \sim H$	$(H \equiv \sim B) \& \sim H$
<i>Bet-1</i>	$\$x_1$	$\$y_1$	$\$y_2$	$\$x_2$
<i>Bet-2</i>	$\$x_2$	$\$y_2$	$\$y_1$	$\$x_1$

This entails, among other things, that you are determinately indifferent between (i) a conditional bet that pays a prize for  $H \equiv B$  and a penalty for  $H \equiv \sim B$  when  $H$  holds and is called off when  $\sim H$ , and (ii) a conditional bet that pays the same prize for  $H \equiv \sim B$  and the same penalty for  $H \equiv B$  when  $\sim H$  holds and is called off in when  $H$ .

Given the complementary nature of your beliefs, Likelihood says that, from  $C$ 's perspective,  $H \equiv B$  is evidence about  $H$ . It only seems otherwise because it is so easy to confuse complementarity with independence. If  $H \equiv B$  and  $H$

where independent, then  $c(B) = 1/2$  would hold all across  $C$ , and  $C_H$  and  $C_{\sim H}$  would encode the same beliefs about  $H \equiv B$ . Here, however, we do not have independence of precise credences, but complementarity of imprecise credences.  $H \equiv B$  is determinately relevant to  $H$  in virtue of the complementarity, but the magnitude and direction of its effect on  $H$ 's probability is indeterminate because of the imprecision.

One can put the point differently by starting not from Likelihood, but from the (equivalent) principle:

*Conditionality.* According to  $C$ ,  $H \equiv B$  is evidentially relevant to  $H$  if and only if  $C_{H \equiv B}$  and  $C_{H \equiv \sim B}$  encode different beliefs about  $H$ .

$C_{H \equiv B}(H)$  and  $C_{H \equiv \sim B}(H)$  span the whole of  $(0, 1)$ , which can make it seem as if they encode the same beliefs about  $H$ . Again, however, the beliefs are complementary: every committee member agrees that  $c(H | H \equiv B) = 1 - c(H | H \equiv \sim B)$ . So, far from being the same, from your prior perspective, your beliefs about  $H$  upon learning  $H \equiv \sim B$  are *antithetical* to your beliefs about  $H$  upon learning  $H \equiv B$ . Learning  $H \equiv B$  puts you in a position in which your uncertainty about  $H$  versus  $\sim H$  is uncertainty about  $H \& B$  versus  $\sim H \& \sim B$ . Learning  $H \equiv \sim B$  puts you in a position in which your uncertainty about  $H$  versus  $\sim H$  is uncertainty about  $H \& \sim B$  versus  $\sim H \& B$ . These are determinately *different* states because  $H \& B$  and  $H \& \sim B$  are contraries, and  $\sim H \& B$  and  $\sim H \& \sim B$  are contraries. So, according to  $C$ , your opinions about  $H$  will be determinately different depending on whether you learn  $H \equiv B$  or  $H \equiv \sim B$ , but this difference will *not* show up in the range of probabilities you assign to  $H$  in the two situations. Again, this means that  $H \equiv B$  is determinately relevant to  $H$ , but that the magnitude and direction of its effect on  $H$ 's probability is indeterminate.

This observation allows us to dispatch a number of White's other objections to the imprecise model. Consider the objection he calls "Reflection". White writes that, "according to the [imprecise] proposal you know in advance that once you [hear my report] you will rationally dilate your credence in *heads*. So, you ought to just dilate it now before the coin[s] even land. . . which is absurd." (p. 178) Here White is claiming that combining dilation with

*Reflection* (van Fraassen [1984]). If you know you will undergo a veridical learning experience in which you discover  $H \equiv B$ 's truth-value (and nothing else), and if you can now identify a single belief state for  $H$  that you will inhabit *whatever* you learn about  $H \equiv B$ , then you should be inhabiting that belief state already.

entails the manifestly incorrect conclusion that your *initial* credence for  $H$  must be spread over  $(0, 1)$ . As in White's first argument, the crucial unstated premise is that you end up in *the same* maximally imprecise belief state relative to  $H$

whether you learn  $H \equiv B$  or  $H \equiv \sim B$ . But, we have already seen that this is not the case. From the perspective of your prior, the beliefs about  $H$  you will come to have upon learning  $H \equiv B$  are *complementary* to the beliefs you will have upon learning  $H \equiv \sim B$ . There is, in fact, no single belief state for  $H$  that you will inhabit whatever you learn about  $H \equiv B$ . You will inhabit an imprecise state either way, but these states will differ depending on what you learn. Reflection does *not* tell you to have imprecise beliefs now if you know that you will have imprecise beliefs in the future. Rather, it says that your current credence for  $H$  should coincide with your current *expectation* of your future credence for  $H$ . But, crucially, an expectation of future imprecise states need not itself be imprecise. That's exactly what happens here: all C's committee members agree that  $H$ 's expected future credence is precisely  $1/2$ ! So, contra White, the combination of Reflection and imprecise model yields exactly the right answer.

Another of White's objections to dilation concerns what happens in a variant of Coin Game:

*Many Coins.* Coin Game is repeated a large number of times with the same fair head/tail coin and different black/grey coins of entirely unknown biases. After each trial you are told how the coins came up and are allowed to update your expectations for future trials in light of this information. Suppose you play the game  $N - 1$  times for some large  $N$  (say, around ten thousand), and that, as one would expect from a fair coin, the frequency of heads has settled down some tiny distance from  $1/2$ . How should you modify your credence for  $H_N$ , the event of a head coming up on the  $N^{\text{th}}$  trial, when you learn  $H_N \equiv B_N$ ?

The imprecise model requires your credence for  $H_N$  to go from precisely  $1/2$  to entirely imprecise, so that your credal state after the first  $N-1$  trials has values for  $c(H_N | H_N \equiv B_N)$  that span  $(0, 1)$ . But, White writes, "it will be hard not to notice after a while that about half the time when you are in the [imprecise] state concerning  $H_N$  the coin does land heads. And it will be hard not to infer *inductively* that it will continue to be the case that about half the occasions in this scenario in which you have seen the coin land and your credence is  $C(H_N) = (0, 1)$ ,  $H_N$  is true." (p. 180, with minor changes) So, in White's picture, observation and good inductive reasoning should, eventually, lead you to a credal state in which your probability for heads remains at  $1/2$ , or thereabouts, even after you learn of how the head/tail and black/grey outcomes are correlated.

Again, the argument goes awry by assuming that credal states which assign the same range of credences for an event reflect the same opinions about that event. To make the error explicit, let's pause to consider an aspect of Many Coins that might go unnoticed. It is essential to the example that, upon learning  $H_N \equiv B_N$  or  $H_N \equiv \sim B_N$ , your credence for  $H_N$  goes from being precisely  $1/2$  to being spread across  $(0, 1)$ . This requires your initial probability to be such



that, given any data stream  $D = (\pm H_1 \ \& \ \pm B_1) \ \& \dots \ \& \ (\pm H_{N-1} \ \& \ \pm B_{N-1})$ , one has

(\*) For all  $c \in \mathbf{C}$ ,  $c(H_N \mid D) = 1/2$  and  $c(H_N \mid D \ \& \ (H_N \equiv \pm B_N)) = c(B_N \mid D)$ .

(\*\*) For all  $b \in (0, 1)$  there is a  $c \in \mathbf{C}$  with  $c(B_N \mid D) = b$ .

(\*\*) means that your initial opinions must prevent you from learning anything inductively about the black/grey coin's bias on trial  $N$  on the basis of information about earlier trials. This is a strange a situation. We are accustomed to thinking of frequency data as yielding useful information about underlying chance processes, but the first  $N - 1$  trials are entirely irrelevant to how the black/grey coin will fall on the  $N^{\text{th}}$  trial.

This strangeness helps to explain why we are so easily misled into thinking, as White does, that inductive inference should be possible here. White wants to argue this way:

- On each trial  $k < N$  you (a) knew the head/tail coin was objectively fair, and (b) received an item of maximally imprecise information about the relationship between  $H_k$  and  $B_k$ .
- In the first  $N - 1$  trials you saw heads come up about  $N^{-1}/2$  times.
- Your situation in trial  $N$  is similar to your situations in the earlier trials since you (a') know that the head/tail coin is objectively fair, and (b') receive an item of maximally imprecise information about the relationship between  $H_N$  and  $B_N$ .
- So, since similar cases should receive similar inductive treatment, your credence for  $H_N$  should be about  $1/2$ .

There are two flaws in this reasoning. First, (b) and (b') are misleading since, as we now know, you do *not* see  $H_k \equiv B_k$  and  $H_k \equiv \sim B_k$  as providing the same information about the outcome of the  $k^{\text{th}}$  toss. Indeed, it is consistent with your evidence that the probability of  $H_k$  is high given  $H_k \equiv B_k$  and low given  $H_k \equiv \sim B_k$ . So, the argument will only succeed if its second premise is replaced by something stronger. You need to know not merely that heads came up half the time, but also that there is no pattern in the data that relates the appearance of heads with reports of  $H_k \equiv B_k$  or  $H_k \equiv \sim B_k$ . If, say, heads came up seventy percent of the time when  $H_k \equiv B_k$  was announced and thirty percent of the time when  $H_k \equiv \sim B_k$  was announced, then heads would still appear about half the time overall, but it would be crazy to use  $1/2$  as your credence for  $H_N$ :  $7/10$  or  $3/10$  would be better choices, depending what you learn. To fix the argument we must assume that, relative to the data  $D$  you receive, your beliefs about  $H_N$  will be (nearly) the same whether you condition on  $D \ \& \ (H_N \equiv B_N)$  or  $D \ \& \ (H_N \equiv \sim B_N)$ . But, since  $H_N \equiv B_N$  and  $H_N \equiv$

$\sim B_N$  are complementary when there is any spread in  $B_N$ 's credence values, this requires  $B_N$  to have a (nearly) sharp credence given the data, so that  $c(B_n | D) = c^*(B_n | D)$  holds (approximately) for all  $c, c^* \in \mathbf{C}$ . But, (\*\*) precludes this! The example presupposes both that you start out with an entirely imprecise credence for  $B_N$  and that the data you receive on the first  $N - 1$  trials does not affect your beliefs about  $B_N$ , in particular, that it does not alter the fact that  $\mathbf{C}$ 's values for  $c(B_N | D)$  range across all of  $(0, 1)$ .

This brings the argument's other flaw into the open. *Ex hypothesi*, you reject any connection between what happens on the first  $N - 1$  tosses of the black/grey coins and what happens on the  $N^{\text{th}}$  toss. The similarity claim in the third premise is then undermined. Your evidential situation upon learning  $H_N \equiv B_N$  or  $H_N \equiv \sim B_N$  is unrelated, in any inductively significant way, to earlier situations in which you learned of a correlation between the outcomes of the two tosses. You see every trial as a whole new experiment, unrelated to the previous, and each is just the original Coin Game. It only seems otherwise because it is so easy to slip into thinking that the  $H_k \equiv B_k$  and  $H_k \equiv \sim B_k$  all leave you in the same evidential state. In fact, from your perspective, the items of information you receive in different trials have nothing to do with one another, and within each trial  $H_k \equiv B_k$  and  $H_k \equiv \sim B_k$  form a complementary pair. In such a situation dilation is the right response to the evidence you receive!

White offers a variant of Many Coins, which some people find more convincing.

*Many Coins-II.* Coin Game is repeated one million times with the same fair head/tail coin and different black/grey coins. You are not shown the outcomes, and the true member of each  $H_k \equiv B_k$  or  $H_k \equiv \sim B_k$  pair is written down on a piece of paper which is temporarily kept hidden from you. You are told, however, that each black/grey coin has a bias toward black between 0.4 and 0.6, and that the tosses are independent. Consider the proposition "*half-heads*" which says that between 450,000 and 550,000 heads came up. Before you are shown the paper your credence for *half-heads* is astronomically close to one. What should it be afterward?

As White points out, the imprecise model entails that your credence for *half-heads* will dilate from being as close to 1 as anyone can imagine to being spread out over almost the whole of  $(0, 1)$ . This happens because, upon learning a truth  $\mathbf{T} = (H_1 \equiv \pm B_1) \& \dots \& (H_{1000000} \equiv \pm B_{1000000})$ , members of your credal committee who placed most of their credal density for  $B$  between 0.40 and 0.41 or between 0.59 and 0.60 will find *half-heads* wildly improbable, whereas members whose densities are focused around the middle of the interval will assign it a probability very close to one.

This counterintuitive result seems to reflect badly on the imprecise model. However, it is intuition, not the model, that has it wrong. The idea that half-

heads should be highly probable derives from our natural tendency to think of coin tossing as a Bernoulli process in which all tosses are independent and have the same chance of producing success. These two features of the situation ensure that the frequencies of various possible outcomes are likely to match their objective chances in the long run. This is exactly how you should think of the heads/tails trials *before* you learn **T**. But, learning **T** destroys the basis for the Bernoulli-process-hence-frequency-converges-to-chance reasoning by destroying a crucial feature of your epistemic situation: when you learn **T** you can no longer legitimately treat the head/tail tosses as independent trials! **T** is a tremendously informative statement that establishes an individual correlation between each of the million head/tail tosses and a million independent events, albeit events whose probabilities are entirely unknown. After learning of all these correlations, you can no longer think of the different head/tail trials as independent any more than you could if learned that, for each  $k$ , a head came up on the  $k^{\text{th}}$  toss exactly if the  $k^{\text{th}}$  name in the Chicago telephone directory ends in a “g” (which would make *half-heads* pretty unlikely). The only difference between the cases is that, owing to your maximally imprecise knowledge of the chances of the various  $B_k$ , you are unable to replace the independence assumption with any even moderately definite alternative. Learning **T**, in other words, leaves you in a state of “unknown interaction” — you know *nothing* about the statistical relationship between the various head/tail outcomes. Though, as we have seen, it easy to mistake unknown interaction with stochastic independence, they are entirely different animals. In general, the right response to unknown interaction is dilation. In this case, that means that your credence for half-heads should be spread out over all of  $(0, 1)$ . To think otherwise is to pretend that you have evidence of independence that you do not really possess.

### **Dilation Everywhere?**

Another line of objection found in White [2010, p. 183], can be illustrated by the following example (slightly retold):<sup>17</sup>

*Bonbons.* You have just received a box of bonbons from *Tic-Tac-Toe Chocolatiers* (a company you know nothing about). It says on the top that the box contains nine chocolate-covered bonbons, eight of which have cherry fillings and one with caramel filling. You draw a chocolate at random. Since you like cherry a lot and mildly dislike caramel, 8:1 odds seems like a good bet, indeed anything better than 2:7 seems worth it. Just as you are about to pop the candy into your mouth you notice that it marked with an X. “I get it,” you think, “*Tic-Tac-Toe* — they mark their candies with Xs and Os!” But, it then occurs to you that you know nothing about how the company marks its candies. It might put Xs on cherry and Os on caramel, Os on cherry and Xs on caramel, or it might mix Xs and Os

up in any random way. It might even be that the whole box is Xs. Should you eat the bonbon?

White thinks it obvious that you should eat the bonbon. Merely reflecting on the filling/marking relationship should not alter your probabilities a bit, he thinks. Since the only way to achieve this is by investing equal credence in *cherry* conditional on X and *cherry* conditional on O, White maintains that you have no option but to apply **POI** subject to this constraint.

In contrast, proponents of imprecision are forced into the seemingly absurd position of saying that, once you see the X, your credence for *cherry* dilates from  $\frac{8}{9}$  to range over  $[0, 1]$ . Since you know nothing about the way Xs and Os are assigned, your credal state will contain committee members with values of  $\lambda = c(\text{cherry} \equiv X)$  that cover  $[0, 1]$ , and this leaves you with values of  $c(\text{cherry} | X) = \frac{8\lambda}{1+7\lambda}$  that cover  $[0, 1]$  as well. Upon seeing X you cannot assign *cherry* any determinate probability at all! But, White argues (pp. 183–184),

this is just not the way we think about such matters. And it's not just about chocolates. We are constantly faced with the possibility of correlations between attributes about which we know nothing. If we were not committed to certain credences in accordance with **POI** then our credences would be mushy all over the place in implausible ways.

So, the worry is that countenancing dilation in Bonbons will force us to accept it everywhere. Indeed, *no matter how the bonbon looks*, if we can imagine it looking another way then we can wonder how looks and flavor are correlated. When we are entirely without information on the topic, dilation ensues. Information about the objective chance of an event thus seems irrelevant as soon as we learn any trivial thing about it!

Proponents of imprecise probabilities have to bite the bonbon here, but it doesn't taste so bad! The first thing to notice is that it is rare for us to face a problem like this without having some inkling about how the properties in question are related. Our prior knowledge of the causal structure of the world usually provides us a pretty good sense of when we need to pay attention to some potential correlation and when we don't. ("The bonbon has a little nick on the side: that's an accident. It has a red dot on top: that may mean something.") So, the prospects for a catastrophic metastasis of dilation may not be as dire as White makes them seem. We are not "faced with the possibility of correlations between attributes about which we know nothing" all that frequently.

Still, we cannot rely on background knowledge to eradicate the problem entirely. I suspect that most of us, knowing what we know about candy makers, will think the X-to-cherry relationship is likely to be either perfect correlation ( $\text{cherry} \equiv X$ ), perfect anti-correlation ( $\text{cherry} \equiv O$ ), or stochastic independence. I also suspect, however, that few of us have any clue about which of the three it is. This uncertainty suffices to produce the maximal dilation.

One way out (a council of desperation?) is to embrace the sharpening strategy discussed earlier. Many people are averse to uncertainty, and become uncomfortable when they cannot associate definite probabilities with events, especially when there are choices to be made. To avoid the uncertainty, and resolve decisions, such a person might adopt policies that lead to fairly sharp, stable beliefs even when her evidence is thin or absent. I do not think there is any question that, as a psychological matter, people often do exactly this. To avoid the discomfort associated with “being at sea,” they banish uncertainty and adopt more precise opinions than are strictly required by their evidence. They reason, for example, that events are likely to be independent when no evidence suggests a correlation (thereby ignoring the fact that no evidence suggests independence either). Indeed, much of the intuitive appeal of belief-forming strategies like **POI** is that they lead to sharp, settled states of belief (and seem to have some underlying epistemological motivation).

One need not go as far as **POI**, however, to reap the benefits of precision. In Bonbon, if you are mainly concerned with the question of whether or not to eat the chocolate, you might merely banish the more extreme extremists by confining your credence for  $cherry \equiv X$  to the interval (0.05, 0.95), thereby eliminating committee members who are too distant from independence. The initial 8:1 odds in favor of cherry then make themselves felt by ensuring that only those committee members who regard  $cherry \equiv X$  as highly unlikely ( $< 1/29$ ) will recommend against eating. So, by adopting this rather minimal sharpening, you get a definite answer: Eat! Notice also that if, impressed by the symmetry of your ignorance, you sharpen only to credal states with  $c(cherry \equiv X)$  values in some symmetric interval around  $1/2$ , then eating is never definitively prohibited and is often mandated. There are sharpenings that leave it indeterminate whether you should eat or not eat, sharpenings that definitely advocate eating, but no symmetric sharpening determinately recommends against eating. Impressed by this, a hungry chocolate-cherry lover might find it appealing to do a little sharpening.

Despite its psychological advantages, from a purely epistemic point of view, sharpening seems *ad hoc*. If you *really* know *nothing* about how Xs and Os correlate with flavors, then you really cannot say whether the X you see is a positive, neutral, or negative indicator of cherry filling, and so you really don't know whether it should lead you to alter your opinions, or in which direction, or by how much. The trivial nature of the case tends to obscure the true level of your ignorance, but increasing the stakes makes dilation more plausible. Consider, for example, the following:

*Crime & Punishment.* You have been (unjustly) convicted of a crime in a foreign land. The standard prison sentence is ten years, but the judge offers you an option. She presents you with nine identical sealed envelopes, lying face-down on a table. She tells you that eight contain pardons but one contains a twenty-five year sentence. She has you designate an envelope at random, and tells you

that you can open it and accept the sentence it contains or leave it unopened and take the ten years. On the basis of the  $\frac{8}{9}$  probability of a pardon you are inclined to open the envelope, but as you pick it up you notice that it is marked with a big X on the other side. Would this give you pause?

Here the X does seem salient: is it an accident? does it mean the long sentence? freedom? might all the envelopes have Xs? These questions matter, and I suspect your inability to answer them will leave you at sea about what to think and what to do. While it may not be obvious how to proceed, the one course that seems definitely wrong is to blithely assume that the X is irrelevant, as White and POI advocate. If you do open the letter, your reason surely will not be that the X doesn't signal anything about your fate. *You know no such thing!* At best, you will be hoping, without evidence, either that the X is irrelevant to the envelope's contents or that it's a positive sign. Your initial confidence, based on knowledge of the chances, that the envelope is likely to hold a short sentence is undermined by the appearance of X. However, since you do not know how to interpret the X's evidential relevance you do not know how your probability should change. The proper result is dilation.

This answer disturbs people for two related reasons. First, in Bonbon and Crime & Punishment, it seems like the initial chances should matter *a lot* even after X is observed. Second, in both problems it seems clear what one should do. In light of the 8:1 odds, what could possibly be said in favor of passing up the chocolate or leaving the letter unopened? Both options seem wrong (though the latter seems *less* wrong), even after you see X. The remedy for both worries lies in recognizing that your knowledge of the initial 8:1 odds does remain, albeit somewhat hidden, in your credal state, and that its presence can have a significant effect on what you do. Let's focus on Bonbon.

If the dilation picture is right, then your credal state contains committee members who assign *cherry* every probability between zero and one, conditional on X. But, focusing only on the range of values for  $c(\text{cherry} | X)$  leaves out a crucial part of the story. One would have  $C_X(\text{cherry}) = [0, 1]$  no matter what the ratio of cherries to caramels ( $>0, <1$ ) in the box. The fact that cherry bonbons outnumber caramel 8:1 only shows up when we look at the relationship between a committee member's credence for *cherry* given X and her estimate of the degree to which X tells for or against *cherry*. The important thing for current purposes is that it will be true across your credal state that

$$(\$) \quad c(\text{cherry} | X) = \frac{8}{9} \cdot R_c(X, \text{cherry}), \text{ where } R_c(X, \text{cherry}) = \frac{c(\text{cherry} \& X)}{c(X) \cdot c(\text{cherry})}$$

$R_c$  is a well-known measure of evidential relevance<sup>18</sup> that construes (incremental) degree-of-support in terms of posterior-to-prior probability ratios. Values of  $R_c(X, \text{cherry})$  above/at/below 1 indicate that, by  $c$ 's reckoning, X

provides positive/neutral/negative evidence for *cherry*, with values further from 1 indicating stronger levels of confirmation or disconfirmation. The  $\frac{8}{9}$  initial probability for *cherry* is significant in (\$) because it means that only committee members who definitely regard X as evidence *against* *cherry* will assign *cherry* a probability of less than  $\frac{8}{9}$  conditional on X. To get a probability less than  $\frac{1}{2}$  it takes an  $R_c$ -value below  $\frac{9}{16}$ , which signals that X moderately disconfirms *cherry*. For a  $c(\textit{cherry} | X)$  below  $\frac{2}{9}$ ,  $R_c$  has to be less than  $\frac{1}{4}$ , fairly significant disconfirmation by this standard. So, more moderate committee members who don't rate X's evidential impact all that highly will find themselves with high credences for *cherry* when they learn X, e.g., those with  $0.9 < R_c(X, \textit{cherry}) < 1.1$  will end up with  $0.80 < c(\textit{cherry} | X) < 0.97$ .

This can matter in decision making. An unfortunate effect of having imprecise beliefs is that one is so often forced to make choices among prospects of indeterminate value. Some of your credal committee members pronounce one option better than another, others claim the reverse, which makes it unclear what you should do. The solution is *not* to have precise beliefs: the precision of your opinions should be tied to the nature of your evidence, not the choices you face. What you need is some decision rule that will tell you how to make choices when expected utility assessments are equivocal. Such rules should never recommend one act over another when every member of your committee says that the utility of the latter exceeds that of the former. Beyond that, there is no consensus among proponents of the imprecise model about what features a decision rule should have. I will say a little about this in the next section, but two things are noteworthy in the current context. First, in contrast with precise credences, where it is sometimes possible to infer a sharp subjective probability for an event on the basis of a person's choices, choices do not reveal imprecise credences in such a direct way. If, say, a person chooses a bet that pays \$1 if  $X$  and \$0 if  $\sim X$  over a straight payment of fifty cents, we cannot thereby infer that she determinately believes that  $X$  is more likely than  $\sim X$ . All we can say, without knowing more about her decision rule, is that some members of her credal committee set  $c(X) \geq \frac{1}{2}$  (i.e., we know only that her upper probability for  $X$  is at least  $\frac{1}{2}$ ).

A second and related point is that, to facilitate choice, decision rules for imprecise probabilities may appeal to considerations similar to those invoked by precise Bayesians, but they will do so for pragmatic rather than epistemic reasons. To take an extreme example, consider what Elga [2010, p. 6] calls the *midpoint rule*, "according to which agents should evaluate bets according to the midpoints of their probability intervals. . . Your evidence nails down the interval and the midpoint rule says that you should evaluate bets in the same ways as someone whose probability is in the *exact* center of the interval." It is hard to know how this rule is to be applied, since there can be a lot of "probability intervals" in play and the center of one need not correlate to that of another. But, let's suppose that in Bonbon the relevant interval concerns your credence

for  $cherry \equiv X$ . If so, then the midpoint rule has you bet as if  $c(cherry \equiv X) = 1/2$  and  $c(cherry | X) = 7/8$ . You will behave exactly as White thinks you should!

Both Elga and White see the midpoint rule as a capitulation to the precise model.

The midpoint rule robs unsharp probabilities of their point... if it is always OK to have point-valued constraints on betting odds, there is no good reason for objecting to point-valued probabilities in the first place. Bottom line: the midpoint rule undermines the main motivation for [the imprecise] model. Elga [2010, 6]

If the answer is that one should bet as if one has sharp credence we might wonder... what the difference between sharp and mushy credence really consists in. White [2010, 180]

I have heard similar sentiments from others, but am unmoved. Unlike a person with sharp credences, when someone with imprecise credences invokes a rule like *midpoint* her goal is not to choose an act that is determinately among the best in light of her assessment of the evidence. Rather, since she feels that the evidence leaves the matter open, and since she must do something, she looks to the decision rule to settle matters that outrun her evidence. The decision-making principles she applies are thus not epistemically motivated. Instead, they are pragmatic principles, intended to facilitate choice, which often explicitly involve acting in ways the agent knows are *not* mandated by her beliefs. For example, using *midpoint* means ignoring aspects of your credal state not reflected in credence intervals (like complementarities) and treating all your committee members as equally reliable (by taking a straight average of their views). A person might recognize that neither of these things are warranted by their evidence, and still find them to be a useful way to make a decision. It would be a mistake, therefore, to conflate the (typically much sharper) ‘pseudo-beliefs’ that can be inferred from an agent’s choices with her actual beliefs, which are based on her assessment of the evidence.

This way of thinking provides a new perspective on the sharpening idea. There is no doubt that people making choices often act as if there is less imprecision in their opinions than there really is. Call this tendency *pragmatic sharpening* to distinguish it from the epistemologically motivated sharpening policies discussed on pp. 291–92. In general, pragmatic sharpening is a matter of proceeding, for purposes of choosing actions, *as if* one’s credal state is a proper subset  $C^\#$  of one’s actual credal state  $C$ . The choice of  $C^\#$  often involves symmetry considerations. When  $C$  generates a credence interval  $(c_-, c^+)$  for an event of interest  $X$ , people tend to adopt a  $C^\#$  that generates  $c(X)$  values that fall within some smaller interval with the same midpoint  $(c_- + \delta, c^+ - \delta)$ . More extreme committee members, relative to the group average, are eliminated, but to preserve the committee’s balance in light of the identical evidential positions of its members, each “left winger” with  $c(X) = 1/2 \cdot (c_- + c^+) - \varepsilon$  is ushered



off accompanied by a complementary right winger with  $c^*(X) = \frac{1}{2} \cdot (c_- + c^+) + \varepsilon$ . This process can lead to definitive decisions even when it stops short of *midpoint*, the limit of such strategies. As we have seen, a little sharpening can provide a rationalization for eating the bonbon. It takes more sharpening to rationalize opening the letter in *Crime & Punishment*. Here too it is true (in virtue of the  $\frac{8}{9}$  chance of freedom) that there are symmetrical sharpenings that recommend opening and symmetrical sharpenings that make no definite recommendation, but there are no symmetrical sharpenings that recommend leaving the envelope closed. The fact that it takes more sharpening to resolve *Crime & Punishment* than in *Bonbon* is a symptom of the fact that the stakes in the former decision are much higher. This also explains why we feel less sure about the decision to open the envelope than the decision to eat the candy: the more sharpening we have to do to resolve a choice problem the further we are going beyond what we take to be our hard evidence, and so the more uncertain we feel about the merits of our choice.

It must be reemphasized, however, that the rationalizations provided by sharpening are entirely pragmatic in nature. Why banish the extremist committee members? Why not eliminate moderates? Instead of moving from  $(0, 1)$  to  $(0.1, 0.9)$ , why not move to  $(0, 0.1) \cup (0.9, 1)$ ? There answer is that there is no good *epistemological* reason for doing the one thing rather than the other. Either way, you end up acting as if you have evidence you do not actually possess. Still, “eliminate the extremists” does have the practical advantage of making consensus easier to achieve while requiring less extreme changes in one’s credal state, e.g., one cannot get a sharpening to recommend eating the bonbon without violating Symmetry.

Sharpening is merely one strategy among many that a person might use to make decisions in imprecise contexts. In all such strategies, the underlying credences do not change.<sup>19</sup> An agent merely acts, in a specific decision-making context, as if she has more precise beliefs than she actually does. She might sharpen in different ways in different contexts, so that different  $C^\#$ ’s are used for different decision problems.<sup>20</sup> But, by adopting some strategy of this sort, she is able to make choices even when her beliefs are too imprecise to definitively recommend any utility maximizing action.

### **Do Imprecise Probabilities lead to Absurd Decisions?**

White [2010] and Elga [2010] allege that the imprecise model has absurd implications for decision making. In particular, they both maintain that the most plausible decision theories for imprecise probabilities will allow agent’s to pass up objectively advantageous bets or to accept objectively disadvantageous bets. The question, as White sees it, is this: “if your credence in  $X$  is  $[x, y]$ , at what odds should you bet?” He considers two “common” answers:<sup>21</sup>

*Liberal.* It is mandatory to take bets that uniquely maximize expected utility for all  $c \in \mathbf{C}$ . It is permissible to take any set of bets that maximizes expected utility relative to some  $c \in \mathbf{C}$ .

*Conservative.* It is mandatory to take bets that uniquely maximize expected utility for all  $c \in \mathbf{C}$ . It is *impermissible* to take any bet that fails to maximize expected utility for all  $c \in \mathbf{C}$ .

I regard Conservative as incoherent. As we have seen, in many decisions there are no options that maximize expected utility across  $c \in \mathbf{C}$ , but one must still choose. In my view, any reasonable decision rule for imprecise probabilities must designate at least one act on an agent's total menu of possible options as permissible. If the agent is constrained to choose off of the menu, then the rule cannot say that *every* choice she might make is definitely impermissible. It might say that she has no mandated options, or even no good options, but it cannot say that every option is prohibited, as Conservative sometimes does. So, let's focus on Liberal.

White seeks to refute Liberal using a twist on his Coin Game example. Suppose that subsequent to learning  $H \equiv B$  or  $H \equiv \sim B$ , you must either accept or decline a wager  $W$  that pays you \$20 in the event of heads and costs you \$10 in the event of tails (where declining the wager means accepting a straight payment of \$0). Note that  $W$  is *objectively favorable* — its objective expected payoff is \$5 — and there is no doubt that you should take it prior to my report. After the report things get dicey since, whether you learn  $H \equiv B$  or  $H \equiv \sim B$ , your credal committee's utilities for  $W$  will end up covering the whole of the interval (20, -10). Supposing, as before, that your credal state has members whose values for  $c(B)$  range over all of (0, 1), then the evaluations of  $W$  look like this:

Table 2

	$Exp_c(W \mid H \equiv B)$	$Exp_c(W \mid H \equiv \sim B)$
$c(H) < 1/3$	Negative	Positive
$c(H) = 1/3$	Zero	Positive
$2/3 > c(H) > 1/3$	Positive	Positive
$c(H) = 2/3$	Positive	Zero
$c(H) > 2/3$	Positive	Negative

Focus on the top and bottom rows. According to White, Liberal makes it permissible, for purposes of action, for you to assign  $H$  any credence represented in your post-learning credal state. Specifically, it is permissible for you to act as if  $H$ 's (and  $B$ 's) probability is less than  $1/3$  upon learning  $H \equiv B$ , and also to act as if  $H$ 's (and  $\sim B$ 's) probability is less than  $1/3$  upon learning  $H \equiv \sim B$ . Since you act as if  $H$  has a probability below  $1/3$  either way, it follows that it is permissible for you to *reject*  $W$  (in favor of \$0) *no matter what*

you learn! White rightly thinks that this would be absurd, but wrongly places the blame on the imprecise model.

The blame really belongs with White's version of Liberal, which presupposes the coarse conception of credal states found in ESET. White's liberal strategy only makes sense if you end up having *the same* maximally imprecise beliefs about  $H$  whatever you learn. If this were so, there would be no problem with acting as if  $c(H) < 1/3$  in both posterior states, since there would really be only one such state to consider. But, as we know, this is not how the imprecise model works. From your prior perspective, your beliefs about  $H$  upon learning  $H \equiv B$  are complementary to your beliefs about  $H$  upon learning  $H \equiv \sim B$ . This complementarity is reflected in the utilities by the fact that the identity  $Exp_c(W \mid H \equiv B) + Exp_c(W \mid H \equiv \sim B) = \$10$  holds for all  $c \in \mathbf{C}$ . I maintain that on any plausible account of rational decision making with imprecise credences, this complementarity prevents you from turning down  $W$  both when you learn  $H \equiv B$  and when you learn  $H \equiv \sim B$ .

To see why let's start with an easier case. In his [2010] Adam Elga argues that the imprecise model cannot satisfactorily explain why you should accept at least one of the following bets, offered in succession:  $V = [\text{Win } \$20 \text{ if } B, \text{ Lose } \$10 \text{ if } \sim B]$ ,  $V^\perp = [\text{Lose } \$10 \text{ if } B, \text{ Win } \$20 \text{ if } \sim B]$ . Elga, like White, considers a "liberal" strategy that allows you to act as if  $c(B) < 1/3$  when deciding whether to accept or reject  $V$ , and then to act as if  $c(\sim B) < 1/3$  when deciding about  $V^\perp$ . He rightly dismisses this strategy, and goes on (successfully) to refute a number of others designed to show how the imprecise model can force you to accept  $V$  or  $V^\perp$ . But, none of the strategies he discusses really gets to the heart of the matter.

The critical fact is that the identity  $Exp_c(V) + Exp_c(V^\perp) = \$10$  holds everywhere in your credal state, owing to the fact that, despite their maximal imprecision,  $B$  and  $\sim B$  are contradictories (so that  $c(B) = 1 - c(\sim B)$  holds across  $\mathbf{C}$ ). It is thus a determinate fact about you that you disprefer  $V$  to a straight payment of  $\$x < \$10$  only if you prefer  $V^\perp$  to a straight payment of  $\$(10 - x)$ , and conversely. It is not determinate whether you disprefer  $V$  to  $\$x$  or whether you prefer  $V^\perp$  to  $\$(10 - x)$ . What is determinate is that your assessments of  $V$  and  $V^\perp$  are related so that  $Exp_c(V) + Exp_c(V^\perp) = \$10$ . As a result, if a decision rule is actually going to reflect your determinate preferences, then it must satisfy the following requirement:

- ☆ The rule permits you to choose  $\$x < \$10$  over  $V$  only if it requires you to choose  $V^\perp$  over  $\$(10 - x)$ . Likewise, the rule permits you to choose  $\$x < \$10$  over  $V^\perp$  only if it requires you to choose  $V$  over  $\$(10 - x)$ .

This idea can be extended in various ways, but for current purposes the key point is that, whatever the final decision theory for imprecise credences looks like, it will surely entail ☆. To omit it is to adopt an overly coarse conception

of credal states that fails to appreciate that your beliefs about  $B$  and  $\sim B$  are, probabilistically speaking, mirror images of one another.<sup>22</sup>

Many potential decision rules satisfy  $\star$ , including all pragmatic sharpening rules that obey Symmetry, but one that definitely does *not* is Liberalism (as used by Elga and White). It should be clear that  $\star$  precludes any rule that lets you assume  $c(B) < 1/3$  in your assessment of  $V$  and also assume  $c(\sim B) < 1/3$  in your assessment of  $V^\perp$ . Instead, it requires you to be consistent: if you do reject  $V$  in favor of  $\$0$  on the basis of  $c(B) < 1/3$  then you must also choose  $V^\perp$  over  $\$10$  (or any smaller sum) on that same basis. More generally,  $\star$  ensures that at least one of Elga’s wagers will always be chosen by any agent using any reasonable decision rule.

To see why one of White’s wagers will always be chosen, we must further explore the consequences of complementarity, particularly as they apply to conditional preferences. Consider the following six wagers for a fixed  $y$ .

Table 3

	$(H \equiv B) \ \& \ B$	$(H \equiv B) \ \& \ \sim B$	$(H \equiv \sim B) \ \& \ B$	$(H \equiv \sim B) \ \& \ \sim B$
$V_{H \equiv B}$	\$20	-\$10	\$y	\$y
$V^\perp_{H \equiv B}$	-\$10	\$20	\$y	\$y
$[\$x]_{H \equiv B}$	\$x	\$x	\$y	\$y
$V_{H \equiv \sim B}$	\$y	\$y	\$20	-\$10
$V^\perp_{H \equiv \sim B}$	\$y	\$y	-\$10	\$20
$[\$x]_{H \equiv \sim B}$	\$y	\$y	\$x	\$x

Everywhere in  $\mathbf{C}$  we have:

$$Exp_c(V_{H \equiv B}) = Exp_c(V_{H \equiv \sim B}) = Exp_c(V) + \$y$$

$$Exp_c(V^\perp_{H \equiv B}) = Exp_c(V^\perp_{H \equiv \sim B}) = Exp_c(V^\perp) + \$y,$$

$$Exp_c([\$x]_{H \equiv B}) = Exp_c([\$x]_{H \equiv \sim B}) = \$x + \$y$$

Thus, every committee member in your credal state will take the same stance toward the  $V_{H \equiv B}$  versus  $[\$x]_{H \equiv B}$  and  $V_{H \equiv \sim B}$  versus  $[\$x]_{H \equiv \sim B}$  questions as she takes toward  $V$  versus  $\$x$ . Likewise she will take the same stance toward the  $V^\perp_{H \equiv B}$  versus  $[\$(10 - x)]_{H \equiv B}$  and  $V^\perp_{H \equiv \sim B}$  versus  $[\$(10 - x)]_{H \equiv \sim B}$  questions as she takes toward  $V^\perp$  versus  $\$(10 - x)$ . This is a determinate fact about your preferences. To respect it any acceptable decision rule for imprecise probabilities should satisfy the following constraint for every value of  $y$ :

- $\star\star$  The rule permits/requires/prohibits the choice of  $\$x < \$10$  over  $V$  if and only if it also permits/requires/prohibits the choice of  $[\$x]_{H \equiv B}$  over  $V_{H \equiv B}$  and the choice of  $[\$x]_{H \equiv \sim B}$  over  $V_{H \equiv \sim B}$ . Likewise, the rule permits/requires/prohibits the choice  $V^\perp$  over  $\$(10 - x)$  if

and only if it permits/requires/prohibits the choice of  $V^\perp_{H \equiv B}$  over  $[\$(10 - x)]_{H \equiv B}$  and the choice of  $V^\perp_{H \equiv \sim B}$  over  $[\$(10 - x)]_{H \equiv \sim B}$ .

Another way to put this is to say that, in virtue of the complementarity inherent in Coin Game, your choices between  $V$  and  $V^\perp$  and sums of money must remain the same, ordinally speaking, whether they are made unconditionally or are conditioned on either  $H \equiv B$  or  $H \equiv \sim B$ .

Combining  $\star\star$  with  $\star$  produces the following interesting result:

$\star\star\star$  If your decision rule permits you to choose \$0 over  $V$  conditional on  $H \equiv B$ , then it also requires you to choose  $V^\perp$  over \$10 conditional on  $H \equiv \sim B$ .

The argument goes like this: If your rule permits you to choose \$0 over  $V$  conditional on  $H \equiv B$ , i.e., if it allows you to choose  $[\$0]_{H \equiv B}$  over  $V_{H \equiv B}$  for any  $y$ , then, by  $\star\star$ , it also permits you to choose \$0 over  $V$  unconditionally. But, if you are permitted to choose \$0 over  $V$  unconditionally, then, by  $\star$ , you are required to choose  $V^\perp$  over \$10 or any smaller sum. But, if you are required to choose  $V^\perp$  over \$10 or any smaller sum, then, by  $\star\star$ , you are also required to choose  $V^\perp$  over \$10 or any smaller sum conditional on  $H \equiv \sim B$ .

For the next step, define conditional versions of White's wager  $W$  as follows:

Table 4

	$(H \equiv B) \ \& \ H$	$(H \equiv B) \ \& \ \sim H$	$(H \equiv \sim B) \ \& \ H$	$(H \equiv \sim B) \ \& \ \sim H$
$W_{H \equiv B}$	\$20	-\$10	\$y	\$y
$W_{H \equiv \sim B}$	\$y	\$y	\$20	-\$10

Notice that  $W_{H \equiv B}$  is identical to  $V_{H \equiv B}$  and that  $W_{H \equiv \sim B}$  is identical to  $V^\perp_{H \equiv \sim B}$ . Thus, your preferences for  $W$  and  $V$  coincide conditional on  $H \equiv B$ , and your preferences for  $W$  and  $V^\perp$  coincide conditional on  $H \equiv \sim B$ . This lets us rewrite  $\star\star\star$  as

$\star\star\star$  If your decision rule permits you to choose \$0 over  $W$  conditional on  $H \equiv B$ , then it also requires you to choose  $W$  over \$10 conditional  $H \equiv \sim B$ .

This is close to what we want. It says that, from the perspective of your prior beliefs, it would be incoherent of you to have a decision-making policy, like Liberal, that allows you to you reject  $W$  no matter what you learn. If your decision rule reflects your actual preferences to the extent that  $\star$  and  $\star\star$  require, then you must deem White's wager choiceworthy either in the event

that heads/tails is correlated with black/grey or in the event that heads/tails is correlated with grey/black.

This answer is unlikely to satisfy White. He was careful to frame his example in terms of your posterior beliefs, not your prior conditional beliefs. He wants to know why it is wrong for you to turn down *W* *ex post*, after you learn  $H \equiv B$  or  $H \equiv \sim B$ , but  $\star\star\star$  is an *ex ante* constraint that speaks only to the question of what decision strategies are coherent *from the perspective of the prior C*. We can bridge the gap by requiring that the decisions you make in your prospective posterior states  $C_{H \equiv \sim B}$  and  $C_{H \equiv B}$  conform to the conditional decisions you make in *C*. More precisely, the requirement is this:

*Dynamic Consistency.* A decision rule should permit/require/prohibit the choice of  $\$x$  over *W* in  $C_{H \equiv B}$  only if it permits/requires/prohibits the choice of  $\$x$  over *W* conditional on  $H \equiv B$  in *C*, i.e., only if it permits/requires/prohibits the choice of  $[\$x]_{H \equiv B}$  over  $W_{H \equiv B}$  in *C* for every possible value of *y*. The same holds with  $H \equiv B$  replaced by  $H \equiv \sim B$ .

If this holds then, on any reasonable decision theory for imprecise credences, it is never allowable for you to reject White's wager *W* no matter what you learn. There are reasonable rules that tell you to accept *W* whatever you learn, but none recommend uniform rejection.

This might seem like rigging the game. How, White might wonder, can a decision rule legitimately restrict choices for wagers involving *H* in  $C_{H \equiv \sim B}$  on the basis of what it permits/requires/prohibits for similar wagers in  $C_{H \equiv B}$ ? After all, your belief state vis-à-vis *H* is *the same* in  $C_{H \equiv B}$  as in  $C_{H \equiv \sim B}$ . While you can view the two states as complementary *from the perspective of your prior*, after you have actually learned  $H \equiv B$  or  $H \equiv \sim B$  you can no longer draw this distinction. Posing the question *ex post*, there appears to be no relevant difference between  $C_{H \equiv B}$  and  $C_{H \equiv \sim B}$  when it comes to *H*'s probability. The only thing that matters, *ex post*, is that your credence values for *H* are spread across (0, 1), no considerations about the complementarity "directionality" of the spread remains. So, on the reasonable principle that a decision rule must produce the same output given the same relevant input, it seems to follow that any rule which permits/requires/prohibits rejecting *W* in  $C_{H \equiv B}$  must also permit/require/prohibit rejecting *W* in  $C_{H \equiv \sim B}$ . Since the imprecise model lacks the resources to mandate acceptance of *W* in both cases, it must permit it in both cases. From this perspective, Dynamic Consistency looks like little more than an *ad hoc* requirement to evaluate present options in the same way that you evaluated past options, even when your epistemic position has changed.

While I agree that a brute prescription to make choices *ex post* just because they seemed best *ex ante* is a nonstarter, this is not the situation here. It is quite true that after learning  $H \equiv B$  or  $H \equiv \sim B$  you no longer inhabit a credal state in which  $c(H | H \equiv B) = 1 - c(H | H \equiv \sim B)$ .<sup>23</sup> Even so, there is an intimate connection between your prior conditional preferences and the preferences you

have after learning. Learning  $H \equiv B$  transforms uncertainty about  $H$  versus  $\sim H$  into uncertainty about  $H \& B$  versus  $\sim H \& \sim B$ , and so the decision you face is whether to accept \$0 or a bet that pays \$20 if  $H \& B$  and costs \$10 if  $\sim H \& \sim B$ . But, you have already taken a stand on this issue in your conditional preferences! Your decision rule, whatever it might be, permits you to choose \$0 over  $W$  conditional on  $H \equiv B$ . Put differently, in a choice between \$0 if  $H \equiv B$  and \$20 if  $H \& B$  and  $-\$10$  if  $\sim H \& \sim B$ , with the outcome when  $H \equiv \sim B$  being entirely independent of what you do, it permits you to choose \$0. This is identical to the decision you face after learning  $H \equiv B$ . While you do know, *ex post*, that  $H \equiv \sim B$  has been eliminated, that should not affect your assessments of \$0 versus [\$20 if  $H \& B$ ,  $-\$10$  if  $\sim H \& \sim B$ ]. Since your decision rule permits you to choose \$0 when the outcomes under  $H \equiv \sim B$  are identical (and so irrelevant to what you choose), it should also permit you to choose \$0 when you know those outcomes will not come about. So, while your decision rule does not allow you to choose \$0 *ex post* merely because it allows it *ex ante*, it should make the same recommendation in the two situations because it should see them, in all relevant respects, as *the same* decision. Similar remarks apply to the situation in which you learn  $H \equiv \sim B$ . You choose differently here because you face a different decision. Instead of comparing \$0 to [\$20 if  $H \& B$ ,  $-\$10$  if  $\sim H \& \sim B$ ] you compare it to [\$20 if  $H \& \sim B$ ,  $-\$10$  if  $\sim H \& B$ ]. For whatever reason, your decision rule considers this a relevant difference, despite the fact that your credence committee leaves the issue maximally undecided. If it did not consider the difference relevant then it would not recommend different choices conditional on  $H \equiv B$  and conditional on  $H \equiv \sim B$ . Given this, Dynamic Consistency asks only that your rule should make the same assessments of the choiceworthiness of acts across situations that it sees as relevantly similar. It should have you choose the same way *ex ante* and *ex post* not because the former choice constrains the latter, but because both choices are responding to the same (pragmatic and epistemic) reasons.

In the end, then, the imprecise model, augmented by a decision rule that obeys  $\star$ ,  $\star\star$ ,  $\star\star\star$  and Dynamic Consistency, has the resources to explain why you must accept White's wager Weither when you learn  $H \equiv B$  or when you learn  $H \equiv \sim B$ .<sup>24</sup> The decision-theoretic objections to the imprecise model do not succeed.

## Conclusion

While I cannot claim to have addressed all the objections that one might conceivably raise to the imprecise model, I do hope to have shown that it is more resilient than one might initially imagine. If we reject the overly coarse conception of imprecise credences as ranges of probability values for events, and embrace a richer conception that lets us distinguish stochastic dependence

from unknown interaction and allows us to exploit complementarities, then many seemingly problematic aspects of the model (e.g., dilation) are revealed as advantages. This expressively richer model is, in my view, the best hope we have of providing a representation of our doxastic states that is capable of honestly recognizing that our beliefs should not be any more definitive or unambiguous than the evidence we have for them.

## Notes

1. For helpful advice and conversation, I'd like to thank David Christensen, Josh Dever, Adam Elga, Branden Fitelson, Dmitri Gallow, Dan Garber, Clark Glymour, Stephan Hartmann, Richard Heck, Chris Hill, Kevin Kelly, Tom Kelly, Berteld Kooi, Sarah Moss, Isaac Levi, Ram Neta, Jim Pryor, Erik Quaeghebeur, Jan-Willem Romeijn, Teddy Seidenfeld, Dan Singer, Brian Skyrms, Eric Swanson, Mike Titelbaum, Brian Weatherson, Jon Williamson, and audiences at Princeton, Rutgers, Carnegie Mellon, Brown, The University of Groningen and the University of St. Andrews.
2. Intuitively,  $\Omega$  is the set of propositions about which the believer is committed to having an opinion. It must contain the necessary proposition, and be closed under negation and countable disjunction.
3. These are the laws of *finitely* additive probability. Many Bayesians require *countable* additivity, but the difference between finite and countable additivity will not concern us here.
4. More generally, if learning alters the believer's confidence in  $D$  without providing new evidence that distinguishes among possibilities that might occur in the event of  $D$ , then her posterior probability  $c^*$  will obey  $c^*(X \& D) = c(X) \cdot c(D | X)$  for all  $X \in \Omega$ . This captures the idea of *Jeffrey conditioning*.
5. It can be shown that this is consistent with the idea that one should update by Bayesian conditioning or Jeffrey conditioning. See Diaconis and Zabel [1982].
6. The probability of black on the first toss is  $\int_0^1 x \, dx = 1/2$ . If you observe one black and update, you obtain a posterior pdf of  $2x \cdot dx$ , and the probability of getting a second black is  $\int_0^1 p \cdot (2x \, dx) = 2/3$ . If you observe a second black and update, you get a pdf of  $3x^2 \cdot dx$  and the probability of a third head is  $\int_0^1 x \cdot (3x^2 \, dx) = 3/4$ . *Et cetera*.
7. I assume, harmlessly, that utility = money.
8. I first heard this analogy from Adam Elga, but have since heard others use it as well.
9. It is unclear what majority rule would mean in this context, and the only reasonable way to implement the differential weight idea would be to introduce a probability  $\mu$  measure over  $C$  and to use it to define an average credence for the whole committee to reflect the person's beliefs. This, however, would be to reject the imprecise approach since these weighted averages amount to precise credences. Proponents of the imprecise interpretation will point out, of course, that in cases like Black/Grey or Game Show nothing in the evidence favors any one choice of  $\mu$  over any other.
10. Indeed, it reduces to **POI** in the finite case.



11. This generalizes **PSET** (and **LP**) since a probability estimate for  $X$  can always be interpreted as the assignment of an expected value to  $X$ 's *indicator* function:  $i_X(X) = 1$  and  $i_X(\sim X) = 0$ . In addition, if  $\mathbf{C}(g)$  is an interval for each random variable  $g$ , and if (as is reasonable in the present context) it is assumed that  $\mathbf{C}_-(g + h) \geq \mathbf{C}_-(g) + \mathbf{C}_-(h)$  and  $\mathbf{C}_-(a \cdot g + b) = a \cdot \mathbf{C}_-(g) + b$  when  $a$  and  $b$  are any constants, then **ESET** is equivalent to the view, defended vigorously in Levi [1974], that identifies a believer's credal state with the *convex* set of probability functions that generates her lower expectations. See Walley [1991] for details.
12. In its convex version **ESET** cannot even capture bare judgments of stochastic independence. Suppose we initially represent your beliefs by the set  $\mathbf{C}$  of all probabilities defined over  $\pm H$  &  $\pm B$  with  $c(H \& B) = c(H) \cdot c(B)$ , so that  $\mathbf{C}$ 's committee members all regard  $H$  and  $B$  as independent. Closing under convex mixtures destroys the consensus: even if  $H$  and  $B$  are independent according to  $c$  and  $c^*$  it does *not* follow that  $H$  and  $B$  are independent relative to  $\mu \cdot c + (1 - \mu) \cdot c^*$  for  $0 < \mu < 1$ . Indeed, the only time this inference goes through is when  $c(H) = c^*(H)$  or  $c(B) = c^*(B)$  or both. Thus, convex **ESET** says that it is impossible to regard two events as independent without having a definite credence for at least one of them. This is implausible on its face, as our earlier example of the weather in Detroit and the Ashes cricket match shows.
13. It is crucial here that the coin is fair, not merely that its bias is known. For a coin of known bias  $\theta$  toward heads the answer one gets is  $c(B \mid H \equiv B) = c(B) \cdot \theta / [(1 - \theta) + (2 \cdot \theta - 1) \cdot c(B)]$ , which only reduces to  $c(B)$  when  $\theta = 1/2$ .
14. As this essay went to press I became aware of Scott Sturgeon's treatment of a similar problem in his [2010]. While I have not had time to fully digest Sturgeon's arguments, he does seem to agree with White on this point.
15. It is no simple matter to say what it is to possess inadmissible evidence for a time- $t$  chance assignment, but a useful sufficient condition is this: learning  $E$  provides one with evidence that is inadmissible with respect to the assignment  $chance_t(X) = p$  when (i)  $0 < c(E) < 1$ , (ii)  $c(chance_t(X) = p \mid E) = c(chance_t(X) = p)$ , and (iii) there is a time  $t^*$  subsequent to  $t$  such that  $c(chance_{t^*}(X) = p \mid E) \neq c(chance_{t^*}(X) = p)$ . In the special case where  $E$  provides direct evidence about  $X$ 's outcome, learning  $E$  ensures that  $chance_{t^*}(X) = 1$  for all  $t^*$  beyond some specified time.
16. Note that those for whom  $c(B) = 1/2$  do not judge that  $H$ 's probability should remain at  $1/2$  as a result of *PP*, i.e., they do not deem  $H \equiv B$  inadmissible and align their beliefs with the known chances. They too regard  $H$ 's initial chance as evidentially irrelevant to its posterior probability. They end up at  $c(H \mid H \equiv B) = 1/2$  because  $c(B) = 1/2$ , not because  $chance_0(H) = 1/2$ . Thanks to Branden Fitelson for helping me to appreciate this.
17. Though superficially very different, this objection is importantly similar to one discussed in van Fraassen [2006]. I would take the same general line of response to that case as I do to this one.
18. Estimated incremental evidential relevance may be measured in a variety of ways, but the differences won't matter here. The same points can be made for every such measure.
19. Let me be careful to distinguish this from what Elga [2010, p. 7] calls the "narrowing proposal," which he characterizes as the idea that "when a rational agent

with unsharp probabilities performs an action her probabilities typically become more sharp or well-defined than they were before.” This is not sharpening! Sharpening does not have agents change their beliefs as a result of what they choose.

20. In localized contexts, where diachronic Dutch books should be avoided, one may want to insist that all the  $C^{\#}$ 's have a non-empty intersection. I think diachronic consistency is a highly local matter, however, and would resist any strong requirement along these lines. It is often *not* a mistake to perform an action that, when conjoined with past acts, produces a Dutch book.
21. Let me note that I think that the categorical notions of a *permissible* act or an *impermissible* act are out of place in an imprecise framework, but I will not press this issue here.
22. In some of his criticisms Elga stresses that  $V$  and  $V^{\perp}$  are presented *in succession*, but I do not see how this affects my argument. To be clear, I am *not* imaging that a decision rule would tell you what to do about  $V^{\perp}$  by relating it to what you already did about  $V$ . It addresses each of your choices individually. So, I am definitely *not* advancing a version of what Elga calls the “sequence proposal,” which hopes to derive the conclusion that you must choose one of  $V$  or  $V^{\perp}$  from the premise that it would be irrational to perform the sequence  $V$ -followed-by- $V^{\perp}$ . Refusing both  $V$  and  $V^{\perp}$  indicates that a mistake has been made, but the mistake is not either that of refusing- $V$ -before-refusing- $V^{\perp}$  or that of refusing- $V^{\perp}$ -after-refusing- $V$ . It is the mistake of having (static) preferences that fail to reflect the fact that  $Exp_c(V) + Exp_c(V^{\perp}) = \$10$ .
23. Some Bayesians (e.g., Joyce [1999]) think it makes sense to have conditional probabilities for conditioning events of probability zero. This would make it possible to preserve complementarity even after one biconditional is learned, in which case the argument of this paragraph would be even more direct.
24. Some people want a theory that requires you to accept the wager whatever you learn. While I do not find this request terribly reasonable, the imprecise model can honor it, e.g., by adopting a pragmatic sharpening strategy that eliminates all the credences for  $H$  that fall outside the interval  $(1/3, 2/3)$ .

## References

- Couso, I., Moral, S. and Walley, P. 2000. “A survey of concepts of independence for imprecise probabilities,” *Risk, Decision and Policy* 5 (2): 165–181.
- Diaconis, P. and Zaball, S. 1982. “Updating Subjective Probability,” *Journal of the American Statistical Association* 77: 822–830.
- Elga, Adam. 2010. “Subjective Probabilities Should be Sharp,” *Philosophers Imprint* 10 (5): 1–11.
- Fine, T. L. 1973. *Theories of Probability*. New York: Academic Press.
- Jaynes, E.T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Joyce, J. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
- Joyce, J. 2005. “How Degrees of Belief Reflect Evidence,” *Philosophical Perspectives* 19: 153–179.
- Joyce, J. 2009. “The Development of Subjective Bayesianism,” in *Handbook of the History and Philosophy of Logic*, 415–476. Amsterdam: Elsevier.

- Kaplan, Mark. 1996. *Decision Theory as Philosophy*. Cambridge: Cambridge University Press.
- Kraft, C., J. Pratt, and A. Seidenberg. 1959. "Intuitive Probability on Finite Sets," *Annals of Mathematical Statistics* 30: 408–19.
- Kyburg, H. 1983. *Epistemology and Inference*. Minneapolis: University of Minnesota Press.
- Laplace, Pierre-Simon. 1825. *Essai Philosophique sur les Probabilités*, fifth edition Paris: Courcier. Translated by A.I. Dale as *Philosophical essay on probabilities*. New York: Springer-Verlag, 1995.
- Levi, I. 1974. "On Indeterminate Probabilities," *Journal of Philosophy* 71: 397–418.
- Levi, I. 1985. "Imprecision and Indeterminacy in Probability Judgment," *Philosophy of Science* 52: 390–409.
- Lewis, D. 1980. "A Subjectivist's Guide to Objective Chance," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, vol. 2: 263–94. Berkeley: University of California Press.
- Seidenfeld, T. and Wasserman, L. 1993. "Dilation for Convex Sets of Probabilities," *Annals of Statistics* 21: 1139–1154.
- Scott, D. 1964. "Measurement Structures and Linear Inequalities," *Journal of Mathematical Psychology* 1: 233–47.
- Sturgeon, S. 2010. "Confidence and Coarse-Grained Attitudes," *Oxford Studies in Epistemology* 3: 126–149.
- van Fraassen, B. 1984. "Belief and the Will," *The Journal of Philosophy* 81: 235–256.
- van Fraassen, B. 2006. "Vague Expectation Loss," *Philosophical Studies* 127: 483–491.
- Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*, London: Chapman and Hall.
- White, R. 2010. "Evidential Symmetry and Mushy Credence," *Oxford Studies in Epistemology* 3: 161–186.