Evidentialism, Inertia, and Imprecise Probability

1 Introduction

Evidentialists say that a necessary condition of sound epistemic reasoning is that our beliefs reflect only our evidence. This thesis arguably conflicts with standard Bayesianism, due to the importance of prior probabilities in the latter. Some evidentialists have responded by modelling belief-states using imprecise probabilities (Joyce 2005). However, Roger White (2010) and Aron Vallinder (2018) argue that this Imprecise Bayesianism is incompatible with evidentialism due to "inertia", where Imprecise Bayesian agents become stuck in a state of ambivalence towards hypotheses. Additionally, escapes from inertia apparently only create further conflicts with evidentialism.

This dilemma gives a reason for evidentialist imprecise probabilists to look for alternatives without inertia. I shall argue that Henry E. Kyburg's approach offers an evidentialist-friendly imprecise probability theory without inertia, and that its relevant anti-inertia features are independently justified. I also connect the traditional epistemological debates concerning the "ethics of belief" more systematically with formal epistemology than has been hitherto done.

In Section 2, I explain what I shall mean by "evidentialism". In Section 3, I describe imprecise probability's appeal for evidentialists, plus the challenge from inertia. In Section 4, I introduce Kyburgian probability. In Section 5, I argue for its felicitous fit with evidentialism. In Section 6, I examine how Kyburgian probability avoids inertia; I argue that it does so under surprisingly general conditions. Yet I finish by giving reasons why even convinced Kyburgians should not discard Imprecise Bayesianism entirely.

2 Evidentialism

in the sense I am using it, 'evidentialism'[1] is a claim about doxastic rationality, i.e. about when a person's beliefs are sound or defective. A belief-state *B* is defective if *B* fails to satisfy some (genuine) epistemic norm; *B* is sound if it satisfies all the norms; therefore, *B* is sound if and only if *B* is not defective. 'Belief-states' include belief, disbelief, or equivocal states of belief, plus aspects of these states like belief ambiguity (Runde, 1990; Al-Najjar and Weinstein, 2009) and strength of credence.

[1] I shall use 'single inverted commas' to refer to unnamed words/sentences, "double inverted commas" for quotation, and **bold** for sentence names. I shall not distinguish between statements and sentences.

Another preliminary concept is the notion of 'total evidence'. Its definition is contentious, but I shall define it this way: *S*'s total evidence with respect to a statement *H* at a time *t* is the set of statements that (1) *S should* be aware of at *t* and (2) provide epistemic, as opposed to pragmatic, reasons to adopt some belief-state with respect to *H*. If I should be aware of *E* but I ignore it due to prejudice, laziness, or cowardice, *E* is nonetheless part of my total evidence. This definition of 'total evidence' obviously leaves many questions unanswered, but these are unimportant here.

For any subject *S*, any statement *H*, a belief-state *B*, and total evidence E_T, evidentialists make the following claim about which states are sound:

Evidentialism: *S*'s *B* for *H* is sound if and only if *B* is purely and properly determined by E_T.

Evidentialism is "strict" or "moderate" depending on how this norm combines with other norms, e.g. the extent to which moral, prudential, or other considerations can override our obligation to have sound belief-states. An extremely strict evidentialism would say that **Evidentialism** is an indefeasible norm of rationality. For moderates, it is just a *pro tanto* duty. Insofar as moderates regard this duty as more easily defeasible, they are less strict in their evidentialism. This spectrum of variations means that it is not neatly demarcatable from non-evidentialism, but I shall simplify my discussion by only discussing strict evidentialism. If an imprecise probability theory is compatible with this uncompromising variant, then *prima facie* it is compatible with more moderate versions.

The influential theses below will recur in my discussion. They follow from some natural analyses of what 'purely and properly determined' might mean in **Evidentialism**:

Proportionality: *S*'s *B* is sound with respect to a statement *H* if and only if *S*'s strength of belief in *H* is proportionate to E_T. In other words, how strongly you believe *H* should be justified by E_T.

Uniqueness: If two subjects S_1 and S_2 possess the same E_T for *H*, then S_1 and S_2 should have the same type of *B* for *H*[2].

Representation: If there is a contrast between two bodies of total evidence *Ei* and *Ej* in some aspect of evidential support, then sound belief-states will represent this difference. For example, some philosophers have argued that if *Ei* provides more information with respect to *H* than *Ej* provides for *H*, then this should somehow be reflected in belief-states (Joyce, 2005).

Exclusion: If anything other than E_T (moral considerations, aesthetic considerations, wishful thinking, caprice etc.) affects the determination of *B*, then *B* is defective[3].

---

[2] Naturally, they will not have the same token of *B*; what they should share is a type of psychological state.

[3] **Uniqueness** is distinct from **Exclusion**: for instance, if *objective* moral values play a role in determining one's doxastic obligations, **Uniqueness** could still be true (the moral values and evidence would jointly require unique belief-states) but not **Exclusion**.

Uniqueness indicates a close relationship between objectivity and evidentialism. However, this connection should not be exaggerated, because evidentialism (as I have defined it) leaves open the possibility that subjective considerations partly determine what constitutes *S*'s total evidence. When we reason, we face a choice of how much epistemic risk to tolerate. Evidentialists may say that subjective judgements are involved in trading-off our desire to acquire true evidence with our desire to avoid accepting false evidence. Yet, insofar as our total evidence is subjective, evidentialism loses much of its purpose, since a significant part of its appeal for its adherents is minimising the subjective factors in our reasoning. Therefore, in practice, evidentialists also think that there are strong objective constraints on acceptable total evidence.

3 Imprecise Probabilities

By 'probability theory', I mean (a) an interpretation of the meaning of ordinary language probabilistic statements like 'The hypothesis *H* is probable' and/or 'The event *A* is probable', plus formal statements like '$P(H) = 0.75$', and (b) a set of rules for determining when such statements are true. By 'imprecise', I mean any approach to formalising probabilities where the probability values are *sometimes* intervals, fuzzy numbers, or similar values, rather than real numbers or their equivalents. In the imprecise theories that I shall discuss, the probabilities in question are interpreted as epistemic, not physical, probabilities.

My description of evidentialism had many terms requiring elucidation, e.g. "proportionate" belief. Imprecise probability theories have several propitious features for the purpose of this elucidation. Firstly, while probabilities seem promising for analysing evidentialism and other epistemological concepts, **Uniqueness** has an uneasy and contentious fit with precise probabilism. Suppose that $P(. \mid .)$ measures a Bayesian subject *S*'s credence in each of a Boolean algebra of statements, conditional on their evidence. The calculus of probabilities provides few constraints upon $P(H \mid E_T)$, the conditional probability of *H* given a body of total evidence E_T. Unless these probabilities are somehow a part of their total evidence (weakening **Uniqueness** to merely require people with identical credences to respond identically to the same evidence) or the rational credences are unique (as Objective Bayesians argue) this means that sound belief-states are much less constrained than evidentialists desire. Finding constraints that identify unique but imprecise degrees of belief seems easier.

Secondly, some evidentialists worry that standard precise Bayesianism conflates some importantly distinguishable belief-states. For example, imagine that you encounter a two-sided alien coin with an apparently symmetric shape, but with dimensions that are very unfamiliar to you. You do know that the tosses are exchangeable. Additionally, for all you know, it might also have very different dynamics from normal coins. H_{1001} says that the coin will land on Side-1 on the toss after 1,000 tosses. E_c is your scanty total evidence about the coin, including the knowledge that any tosses not landing on Side-1 or Side-2 are re-tossed.

Suppose that $P(H_{1001} \mid E_c) = 0.5$. Now imagine you toss the coin 1,000 times and it lands on Side-1 in 500 tosses and Side-2 in 500 tosses. The results of these tosses are reported in $E_{1/2}$, so that your total evidence is now $(E_{1/2} \wedge E_c)$. With suitable priors, $P(H_{1001} \mid E_{1/2} \wedge E_c) = 0.5$. However, some have argued that there is a relevant asymmetry between your total evidence before and after learning $E_{1/2}$ that your belief-states should somehow reflect (Popper, 1980 pp. 407–408). If so, precise probabilism violates **Representation**.

Thirdly, there is the problem of having equivocal beliefs in all hypotheses for which your evidence is evenly balanced. **Proportionality** requires equivocal beliefs given total evidence that is neutral towards any hypothesis. Consider the alien coin. Let H_{n2} be the hypothesis that the coin will land on Side-1 in two consecutive tosses. Suppose that you believe that the coin tosses are exchangeable. Assume an apparently equivocal prior distribution in which the probability that every coin toss *i* lands on Side-1 is 0.5 for each toss described in your algebra. This prior seems to give us a satisfyingly neutral belief-state. Suppose that you think that the tosses are physically independent (e.g. one toss landing on Side-1 does not affect the chance that another toss will do so) and your data model reflects this assumption. Yet, suppose you also believe that evidence of coin toss outcomes is relevant to the outcomes of future tosses; in particular, 0.51 is your credence in the coin landing on Side-1 in the one toss given that it has landed on Side-1 in the only preceding toss. Let H_1 be the hypothesis that it lands on Side-1 in the first toss and H_2 that it lands on Side-1 in the second toss. From the probability calculus, $P(H_{n2}) = P(H_1)P(H_2) = (0.5)(0.51) = 0.255$. Additionally, it follows that $P(\neg H_{n2}) = 1 - P(H_{n2}) = 0.745$. Hence, while your beliefs are neutral with respect to the result of any individual toss *i*, they commit you to strong opinions about other hypotheses, even though your total evidence seems neutral towards them too. Intuitively, this is far from **Proportionality**: despite your symmetric evidence with respect to H_{n2} and $\neg H_{n2}$, your belief-states favour the latter. In general, there is no way of using precise probabilities to represent a belief-state that is equivocal with respect to all the statements of a complex Boolean algebra of statements (Cox, 2006 p. 73).

Instead, philosophers like James M. Joyce (2005) have argued for using a *set* of probability functions. For my purposes, the salient claims in Imprecise[4] Bayesianism are these:

Rule (1): If *S* has sound belief-states, then we can model them by the cover of values given by a set $\mathcal{C}$ consisting of probability functions, such that (a) each function provides a coherent additive probability measure and (b) every function assigns a value of 1 to each statement in *S*'s total evidence. For a hypothesis *H*, the cover of values for *H* from the different functions in $\mathcal{C}$ usually[5] provides an interval-valued[6] representation of belief-states $[x, y]$ where *x* is the

[4] Note that interval-valued beliefs are precise, just not real-valued. For reasons I shall not discuss, I think that this is a virtue, not a fault.

[5] Imprecise Bayesians vary about $\mathcal{C}$'s convexity and closure. An interval representation is not always possible (Joyce, 2005 p. 177).

[6] I shall use 'interval' as an abbreviation of 'closed unit interval', i.e. intervals with limits $[x, y,]$ where $x \geq 0$ and $y \leq 1$. A closed interval contains its limits, an open interval does not.

lowest value that any function in $\mathcal{C}$ gives to *H* and *y* is the highest value. The same holds, *mutatis mutandis*, for conditional probabilities. In special cases, this interval may be degenerate, meaning that $x = y$, as in the interval [0.5, 0.5]. To describe $\mathcal{C}$, I shall use an auxiliary function *C*, so that $C(H, E) = [0.1, 0.8]$ means "The lowest value that any function in $\mathcal{C}$ assigns to *H* conditional upon *E* is 0.1, while the highest value is 0.8."

Rule (2): If *S* learns *E* and nothing else, then her subsequent belief-states should correspond to a set $\mathcal{C}_E$, where $\mathcal{C}_E$ is the set that results from updating each function in $\mathcal{C}$ by Bayesian conditionalisation. Any function in $\mathcal{C}$ that assigns 0 to *E* is not included in $\mathcal{C}_E$ (Joyce 2010 p. 287).

To these, an "evidentialist" Imprecise Bayesian adds a principle like the following:

Chance Grounding Thesis: Suppose that *S* is rational and $\mathcal{O}$ is the set of objective chance hypotheses whose numerical chance for each event *A* described in *S*'s algebra of beliefs is consistent with *S*'s total evidence. Then, $\mathcal{C}$ includes all probability functions whose expected values for each *A* are those given by a probability density function (PDF) defined over the members of $\mathcal{O}$. In general, this requirement will simply require that any sharpness in *S*'s beliefs should be justified by *S*'s beliefs about objective chances (White, 2010 p. 174; Vallinder, 2018 p. 1210)[7].

In the alien coin example, there should be a function in $\mathcal{C}$ for each real value from 0 to 1 that could be assigned to H_{1001}, because $\mathcal{O}$ will have a hypothesis assigning a chance for each such value to the 1001st toss lands on Heads. This is even true for 0 and 1, because (for all you know) the coin's shape might cause it to always roll onto one particular side.

A significant advantage of Imprecise Bayesianism is that we can often model a genuinely neutral set of belief-states. Consider the sequence of coin tosses. Let H_i, E_c etc. be as before. If $C(H_i, E_c) = [0, 1]$ for each individual toss H_i,, then it can also be the case that $C(H_{n10}, E_c) = C(\neg H_{n10}, E_c) = [0, 1]$. As we saw, this neutrality is impossible when assigning real values. Additionally, if our total evidence excludes some particular chances, then evidentialist Imprecise Bayesianism can still assign precise probabilities to the coin toss predictions.

Yet there is a problem with this attempted marriage of evidentialism and imprecise probability. Vallinder (2018) argues that Imprecise Bayesianism will often require us to adopt

---

[7] This version is Vallinder's, which is more precise than White's. In Imprecise Bayesianism 'consistency with total evidence' normally means that each statement in the total evidence has a non-zero probability (Bradley, 2019 Section 1.1). An anonymous referee notes that this definition will require some finessing for infinite state spaces, where logically and metaphysically possible events sometimes must be assigned a probability measure of zero. I do not know how an evidentialist Imprecise Bayesian would handle that point, but perhaps non-probabilistic modal formalisms could be used to define 'consistency'.

sets of functions with a feature called "inertia". In Imprecise Bayesianism, maximally imprecise belief-states for a hypothesis remain imprecise on learning evidence that neither contradicts nor implies that hypothesis. To see why, consider a feature of standard Bayesianism: if $P(H) = 0$ or $P(H) = 1$, then conditioning on any evidence E such that $P(E) > 0$ will result in another extreme value $P_E(H) = 0$ or $P_E(H) = 1$ respectively. For example, consider the alien coin example. Recall that, in this example, you do not know if the tosses are physically independent (you just assumed they were in the precise Bayesian reasoning I discussed earlier) so your set $\mathcal{C}$ includes functions for which they are dependent[8]. Consequently, if $C(H_{1001}, E \wedge E_c) = [0, 1]$, the extreme probability functions P_{min} and P_{max} in $\mathcal{C}$ that give, respectively, probabilities of 0 and 1 to H_{1001}, will retain this value upon learning E. Given these assumptions, the only exceptions are the special cases where:

(a) E contradicts $H_{1001,}$ which eliminates P_{max} and results in $C(H_{1001}, E \wedge E_c) = [0, 0]$ because $P(H_{1001} \mid E) = 0$ for every other function.

(b) E entails H_{1001}, which eliminates P_{min} and results in $C(H_{1001}, E \wedge E_c) = [1, 1]$ because $P(H_{1001} \mid E) = 1$ for every other function.

Otherwise, P_{max} and P_{min} remain in $\mathcal{C}$, so your belief-state for H_{1001} stays maximally imprecise[9].

This result holds very generally. For example, given the earlier assumptions, your beliefs stay maximally imprecise even if your new evidence consists of a massive sample report of coin tosses all landing Heads, with no evidence that the sample is unrepresentative. Given the **Chance Grounding Thesis**, maximally imprecise belief-states will be common. Therefore, Imprecise Bayesianism will often force us to respond in an apparently disproportionate way to non-deductive ("ampliative") evidence (see also White, 2010 pp. 173–174)[10].

---

[8] An anonymous referee points out that if functions in $\mathcal{C}$ assign values to the individual coin toss hypotheses such that they are probabilistically independent (in your credences) then your belief-interval *can* narrow upon learning non-deductive evidence. For example, suppose that your new evidence H_1 asserts that the first toss landed Heads. This evidence coin toss eliminates functions assigning $P(H_1 \mid E) = 0$, and therefore $P(H_1 \mid E) > 0$ for every $P \in \mathcal{C}$. Since the remaining functions also treat tosses as independent, it follows that $P(H_{1001} \mid E) > 0$. Thus, $C(H_{1001}, E \wedge E_c) = (0, 1]$, i.e. a half-open interval, differing from [0, 1]. However, in the alien coin example, we are assuming (epistemic) probabilistic dependence for the tosses.

[9] An anonymous referee worries that it might be unfair to Imprecise Bayesians (relative to precise Bayesians) to require them to not assume that the tosses are independent and identically distributed (i.i.d). However, for evidentialists, the possibility of not making this assumption in Imprecise Bayesianism is one of its attractions. Furthermore, merely making the i.i.d. assumption will not help: Vallinder (2018) demonstrates how inertia can occur with sets of beta priors where the tosses exchangeable in each element of $\mathcal{C}$.

[10] Vallinder (pp. 1214–1216) proves that the presence of dogmatic functions from $\mathcal{C}$ is unnecessary for inertia; it can occur even without maximally imprecise belief-states.

I shall not discuss all of the possible evidentialist Imprecise Bayesian responses, such as disputing whether inertia is genuine problematic (Joyce, 2010 p. 291). However, I shall discuss one good answer, to illustrate how evidentialism constrains such responses. Some Imprecise Bayesians think that we should not have [0, 1] belief-states. Susanna Rinard (2013) argues against the use of the [0, 1] interval because of inertia. (She gives other reasons too.) If we also narrow the **Chance Grounding Thesis** in some way that excludes probability functions from $\mathcal{C}$ that regard the evidence as irrelevant, then the issue is solved. However, while Rinard's arguments are promising for non-evidentialists, evading the [0, 1] interval has least two problems for evidentialists[11]. Suppose that we model maximal ignorance in the coin tossing case via a set $\mathcal{C}$ such that $C(H_{1001}, E_c) = [0 + \delta, 1 - \delta]$, where δ is a tiny number. This conflicts with evidentialism, because **Uniqueness** is threatened: our evidence apparently underdetermines δ (Vallinder, 2018 p. 1214). Moreover, **Proportionality** is threatened, because giving up the [0, 1] interval means that we no longer have an equivocal belief-state that can offers a generally symmetric distribution across such a wide range of hypotheses. If your belief state in H_i given your evidence is a half-closed interval $C(H_i, E_c) = [0 + \delta, 1 - \delta)$ for each toss i, then your strength of belief in the hypothesis H_L that a very large sequence of coin tosses (where the value of "large" depends on δ) will all land on Side-1 must be very low, despite the complete absence of evidence against H_L. Whatever its other virtues, Rinard's response to inertia takes would not help *evidentialist* Imprecise Bayesians.

I stress that non-evidentialist Imprecise Bayesian epistemologies are untouched by these points. Additionally, I have not argued that inertia troubles Imprecise Bayesianism as such. However, I have provided some reasons for evidentialists to consider an alternative approach. An obvious target for modification is the update method **Rule (2)** (White 2010, pp. 184–185). This rule expunges probability functions in $\mathcal{C}$ that assign 0 to newly accepted evidence and conditionalizes each surviving function. However, evidentialists should be wary of avoiding inertia via an *ad hoc* modification. They should want an independently justified alternative, within a systematic theory of reasoning.

4 Kyburgian Probability

I shall begin with the interpretation of Kyburgian probability[12] statements, before explaining the system's rules for determining when the statements are true or false. I shall return to inertia in Section 5. As in Imprecise Bayesianism, Kyburgian probabilities are epistemic rather than physical. However, there are three important differences of interpretation.

Firstly, Kyburg adopts a logical interpretation of probability, which I shall call

[11] Rinard (2013, p. 162) rejects strict evidentialism.

[12] I use 'Kyburgian probability' rather than Kyburg's 'Evidential Probability' to avoid confusion with evidentialism.

'logicism' for short[13]. It is distinct from logicism in the philosophy of mathematics. Logicism, as I shall mean it, claims that epistemic probabilities are objective features of arguments which have logically consistent premises[14]. The restriction to consistent premises is because it is hard to define probabilities of arguments with logically inconsistent premises, e.g. $P(H \mid E)$ is undefined for ordinary probability functions if E is inconsistent. The meaning of 'argument strength" will vary with the logicist. Yet they agree that statements about (epistemic) probability are metalinguistic, just like '*H* implies *E*' or '*H* has more words than *E*'. Since probability statements are metalinguistically true, if true, their truth value is knowable *a priori* by anyone who understands the relevant statements and the concept of probability[15]. Logicists interpret probability statements that do not refer to premises, like "It will probably rain tomorrow", as implicitly referring (perhaps vaguely) to background premises. The most famous logicists are John Maynard Keynes (1921) and Rudolf Carnap (1962).

I shall illustrate the logicist analysis of probability with some examples. In the logicist conception of probabilistic notions, '*H* is probable given *E*' is equivalent to 'The argument from *E* to *H* is a strong argument'. Similarly, '*H* is improbable given *E*' is equivalent to 'The argument from *E* to ¬*H* is a strong argument'. Meanwhile, '*H* is neither probable nor improbable given *E*' is equivalent to 'Neither the argument from *E* to *H*, nor from *E* to ¬*H*, is a strong argument'. Numerical equalities like $P(H \mid E) = r$ are true if and only if r correctly measures the strength of an argument from E to H.

The connection between logical probabilities and beliefs is controversial (Popper, 1980). I shall assume that the *epistemically* sound belief-states for *S* correspond to the logical probabilities of statements given *S*'s total evidence[16]. So, if $P(H \mid E) = r$ and E is *S*'s total evidence, then *S* should believe *H* to degree *r*. Therefore, for most logicists, while probability is not identical to rational degree of belief, it nonetheless guides us to epistemically rational beliefs.

A second difference with most Imprecise Bayesians is Kyburg's idea that logical probabilities always involve statements about physical probabilities. Kyburg thinks that, generally, assertions of (actual or hypothetical) relative frequencies[17] suffice for these physical probability statements (Kyburg, 2002) but his theory's spirit is compatible with a larger role for propensities. I shall stick to frequencies to avoid unnecessary complications,

---

[13] In principle, one could also combine such an interpretation with a set-based approach to Bayesianism, perhaps similar to (Carnap, 1962 Chapter VII).

[14] "The logical interpretation of probability" sometimes used very broadly to mean *any* objective interpretation, but this smudges important differences (Rowbottom, 2008).

[15] In principle, a logicist who believed that all logical knowledge is *a posteriori* could say logical probabilities are also *a posteriori*.

[16] This makes sense for evidentialists, but a non-evidentialist could use logical probabilities for the particularly evidential reasons for belief.

[17] I shall subsequently use 'frequencies' as shorthand for '*relative* frequencies'.

but a propensity-based approach would not affect my conclusions.

The third difference with Imprecise Bayesianism is that Kyburgian intervals are not descriptions of a set of functions. Instead, as we shall see, the intervals are the outputs of a single function relating an argument's premises to its conclusion.

Kyburgian probabilities can be determined by a sequence of formal rules. These are neatly defined for particular formal languages, but also (less neatly) applicable to reasoning in natural languages. Like the formalization of deductive logic, it is unrealistic to expect this formal procedure to be identical to our intuitions. Instead, he aims to model our intuitions about probabilistic reasoning in a way that is more precise and systematic than our ordinary conceptions. Implicitly, Kyburg's metaphilosophy is similar to Carnapian "explication" (Carnap, 1962 Chapter I).

The core intuition behind the determination of Kyburgian probabilities is the notion that arguments of the form:

(p1) *i* is a member of the reference class $\mathcal{K}_1$.

(p2) The frequency of satisfying some expression (e.g. a predicate) Φ among members of $\mathcal{K}_1$ is somewhere in the interval [*x*, *y*][18].

Therefore, (c) *i* possesses Φ.

– has a strength measurable by [*x*, *y*][19]. If the interval in (p2) is [1, 1], then the argument is maximally strong; if it is [0, 0], then the argument is maximally weak. Intermediate intervals like [0.25, 0.5] or [0.9, 1] measure intermediate argument strengths[20].

Hence, if (p1) stated that *i* is a swan and (p2) stated the frequency of whiteness among swans is between 90% and 95%, then this argument is strong to the extent [0.9, 0.95]. This interval is also our epistemically sound belief-state if our total evidence with respect to (c) were just (p1) and (p2). If (p2) were 'All swans are white', then the argument would be maximally strong and the belief-interval would be [1, 1].

In real life, our total evidence is never so exiguous: we know that any *i* is a member of many reference classes. Our statistics for these classes will often conflict. Additionally, such arguments are "non-monotonic": adding new premises can affect their strength. For example,

[18] An interval, rather than a fraction, because our statistical methods might only provide imprecise population frequency estimates of reference classes.

[19] Unlike Isaac Levi (1977) Kyburg does not require an additional premise that *i* is randomly selected from $\mathcal{K}$. White (2010, pp. 170–171) succinctly presents a case against Levi in this controversy. However, Kyburg does require that Φ and $\mathcal{K}$ are classes/properties that we regard as genuine, unlike 'grue', 'grue emeralds' etc.

[20] In some cases, one argument might have more informative premises than another, but equal strength, e.g. [0.5, 0.5] relative to [0.1, 0.9].

Kyburg argues that this argument is intuitively strong:

(p1) *i* is a swan.

(p2) The frequency of whiteness among swans is [0.9, 0.95].

Therefore, (c) *i* is white.

– but adding these premises creates a weak argument:

(p3) *i* is an Australian swan.

(p4) The frequency of whiteness among Australian swans is [0.01, 0.05].

In the absence of further relevant information, it seems as though we should ignore the information from (p1) and (p2), because their significance is undercut by (p3) and (p4). For this reason, Kyburg focuses on determining (a) when we can rationally ignore information in our premises and (b) how to combine any conflicting information that we cannot rationally ignore.

Kyburg uses the term "Sharpening" for this process of ignoring information in our premises. The details of Sharpening evolved over time; I shall use his last version (Kyburg, 2006). To have a Kyburgian probability, our hypothesis of interest must be equivalent to a class-membership statement for a single-case. (This single-case can be a population or a sample.) I shall continue to use *i* and Φ, so that our hypothesis is equivalent to '*i* is a member of the class of things that satisfy Φ'. Let $\mathcal{R} = \{R_1, R_2, \ldots R_n\}$ be the set of all the statements in our total evidence E_T such that each element in $\mathcal{R}$ asserts two things: that *i* is a member of a reference class $\mathcal{K}$ and that the frequency of Φ in $\mathcal{K}$ is somewhere in [*x*, *y*].

Our reasoning's premises should be our relevant total evidence, so Sharpening begins with rules for determining 'relevant' statements in $\mathcal{R}$:

Richness: Suppose that (i) neither R_1 nor R_2 asserts an interval that is a proper subinterval of the other's interval, (ii) the reference class described by R_1 is $\mathcal{K}_1$, the reference class described by R_2 is $\mathcal{K}_2$, and our total evidence implies that $\mathcal{K}_1$ is the cross product of $\mathcal{K}_2$ and another reference class $\mathcal{K}_3$. **Richness** requires that we ignore the information from R_2 in favour of the information from R_1. The intuition is that R_1 is based on higher-dimensional evidence than R_2, so it uses more of our data. R_1 is 'richer' in the sense of stating a frequency that is conditional upon more of our statistical information.

Example: Imagine that you are selecting a card from one of two piles. You wonder whether you will select a red card. You have checked each pile before shuffling them, and you have seen that Pile 1 has 13 reds out of 52 cards, whereas Pile 2 has 39 reds out of 52 cards. You

know that, once you have selected a pile, the card will be chosen randomly, but you also know that you will make your selection based on the outcomes of two tosses of a fair coin: Pile 2 if you get Heads twice, Pile 1 otherwise. Let R_1 be the statement that the relative frequency of reds in the piles is $\frac{52}{104}$ and thus in the degenerate interval [0.5, 0.5]. Let R_2 be the statement that the (hypothetical) frequency of drawing reds in the overall selection set-up is $(\frac{13}{52})(\frac{3}{4}) + (\frac{39}{52})(\frac{1}{4}) = (\frac{39}{208}) + (\frac{39}{208}) = (\frac{78}{208})$ = 0.375 and therefore in the interval [0.375, 0.375]. **Richness** requires ignoring R_1 in favour of R_2, because R_1 provides richer information about card selections than the merely proportional data from R_2.

Specificity: This rule is applied sequentially from less to more general members of $\mathcal{R}$ that survive **Richness**. Suppose that (i) neither R_1 nor R_3 asserts an interval that is a proper subinterval of the other and (ii) your total evidence E_T implies or states that the reference class described by R_3 is a proper subset of the reference class of R_1. **Specificity** requires ignoring the information provided by R_1, because R_3 describes a narrower reference class.

Example: Imagine you have drawn the card, but not looked at it. A trustworthy friend playfully snatches the card and examines it. She comments that "Wow, 9 out of 10 cards I've seen drawn today from these piles are red!" You believe her. You do not know the outcomes of the other selections she's seen, but you do know that the selections she has seen are a proper subset of this card-drawing set-up in general. It is the latter, broader class that R_1 describes. Let R_4 be the statement that [0.9, 0.9] of selections today were red. **Specificity** requires ignoring R_1 in favour of R_4[21].

The members of $\mathcal{R}$ that survive **Richness** and **Specificity** are 'relevant'. Sharpening next requires that we separate the intervals in the relevant statements into sets such that each set $\mathcal{F}$ contains any intervals that are neither identical nor a proper subinterval of any other member of $\mathcal{F}$. Thus, in the card selection case, if the intervals of our relevant premises are [0.1, 0.96], [0.8, 0.85], [0.9, 0.9], [0.4, 0.95], and [0.6, 0.96], then:

$\mathcal{F}_1$ = {[0.1, 0.96]}

$\mathcal{F}_2$ = {[0.4, 0.95], [0.6, 0.96]}

$\mathcal{F}_3$ = {[0.8, 0.85], [0.9, 0.9]}

The intuition behind the next rule of Sharpening is that, *once* the 'relevant' premises are identified by **Richness** and **Specificity**, if one surviving interval is more precise than another, then it should be favoured, due to its greater content (in an informal sense). For

[21] An anonymous referee raises this objection: suppose that we know that coin tosses in general land Heads with a frequency of a particular coin in general [0.49, 0.51], but that the proportion in tosses today is [1, 1]. Shouldn't we use the more general frequency [0.49, 0.51]? We should not if our hypothesis is that a particular toss *i*, which we know will occur today, will land Heads. In fact, we can determine the degenerate interval for the unit set consisting of *i*, which Sharpens conflicting reference class information via **Specificity**. If we do not know that *i* occurred today, then we should use the more general information.

example, if we have vague statistical data about the tosses of a particular £1 coin today (e.g. that 0.05 to 0.99 of them landed Heads) but relatively precise information about tosses of £1 coins in general (that about 0.49 to 0.51 of them land Heads) then we should favour the more precise information, even though it is less specific.

The next and final rule of Sharpening is an attempt to systematise this intuition:

Precision: The probability of *H* relative to our evidence is the shortest interval that covers all the intervals of an $\mathcal{F}$.

Example: In the case of $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$, the shortest cover is $\mathcal{F}_3$'s cover, which is [0.8, 0.9], and therefore this is the Kyburgian probability of selecting a red card given your total evidence. Formally, using *KP* to denote the function determining the Kyburgian probability of an argument from E_T to a conclusion *H*, we have $KP(H, E_T) = [0.8, 0.9]$. In the £1 coin tossing case, where the $\mathcal{F}$ sets are just {[0.49, 0.51]} and {[0.05, 0.99]}, the Kyburgian probability of the £1 coin landing Heads is [0.49, 0.51]. Note that *KP* is not an additive conditional probability function.

Sharpening does not necessarily determine a unique reference class statement for a probability. It does so in the £1 coin example: the probability interval comes from just our evidence about tosses of £1 coins in general. However, in the red card example, evidence about two reference classes jointly contribute to the probability. Thus, there are two potential sources of imprecision in Kyburgian probability: (i) our premises might contain approximate estimates of frequencies in reference classes and (ii) **Precision** might require an interval-valued probability.

Furthermore, $KP(H, E_T)$ will not have a value for every possible E_T: without relevant statistical information (including information derivable from mathematics and logic) in E_T, there are no relevant premises, and hence no probability. However, if E_T is rich enough to include the statistical facts derivable via mathematics and logic, then $KP(H, E_T)$ will always exist, because [0, 1] will always be a candidate interval for logico-mathematically contingent statements[22]. This is because we know, from deductive logic, that *i* is a member of the unit set consisting of just *i*, and we know (assuming classical logic) that *i* is either a Φ or not a Φ. Hence, the frequency in the unit set is 0 or 1, and therefore the frequency must be in the interval [0, 1].

Kyburg's theory raises many issues. I cannot discuss them all here, but I shall make a few clarificatory points. Since Kyburgian probabilities are logical probabilities, they are never identical to belief-states. Kyburg thinks that probability is "legislative" for belief (Kyburg, 2006 p. 48) so an epistemically rational person's beliefs should correspond to the logical probabilities for their total evidence, but this normative claim is not an analysis of

[22] For logico-mathematical truths/falsehoods, their Kyburgian probability for such total evidence will be [1, 1] and [0, 0] respectively.

probability.

Another issue is Kyburgian probabilities' relation to decisions. The significant divergence from Bayesian decision theory is regarding decisions and updating, i.e. diachronic coherence. In standard Bayesianism, we have a pleasingly simple relationship (aside from some controversial special cases[23]) between conditional probabilities and rational betting odds for H conditional on learning a sequence of propositions $\{E_1, E_2 \ldots E_n\}$. The rational betting odds are given by for each E in the sequence. In contrast, because Kyburgian probabilities are not conditional probabilities, we cannot determine odds in this way. Furthermore, it is possible that there is no conditional probability function that always assigns values to hypotheses that are within the Kyburgian intervals, due to Kyburgian updating's divergence from conditionalization (Levi, 1977). Consequently, the probabilities are unattractive as a means of guiding ideally rational decisions, although Kyburg tries to argue otherwise (Kyburg, 1990 Chapter 8)[24].

I have talked about statistical statements as if they fall like manna from heaven. Clearly, we need to infer them. While Kyburg's rules of Sharpening evolved, his theory of inference remained fairly stable from his principal treatise on statistical inference (1974) to his later work (2001); see also (1990) for many details. I shall not discuss this step in detail: firstly, because it is logically independent of his probability theory; the same probabilities could be combined with a different theory of ampliative inference. Secondly, my focus is the probability relation between a hypothesis and some total evidence, not that evidence's origin or contents.

5 Kyburgian Evidentialism

Before discussing inertia, I shall examine Kyburgian probability's *prima facie* fit with **Evidentialism**.

5.1 Evidentialism and Its Corollaries

According to **Evidentialism**, a subject's epistemically sound belief-states with respect to H are purely determined by their total evidence. Kyburgian probabilities are purely determined by a sequence of formal rules using the statistical information in our total evidence as the premises in an argument for H. Thus, if there really is a correspondence between Kyburgian

[23] For example, if you are confident in humanity's extinction via nuclear war at time t, is it rational to bet on this event if the bet matures after t? (Talbott, 2016).

[24] For a synchronic set of bets, there is a simpler relationship between coherent betting and Kyburgian probability. Suppose that E_T is a fairly rich evidential corpus and suppose we are betting about a finite set of statements. Kyburg proves that coherent odds can be generated directly from the coherent precise probability function(s) whose values for a hypothesis H given some total evidence E_T are within the corresponding Kyburgian probability interval for $KP(H, E_T)$ (Kyburg, 1974 pp. 320–322).

probabilities and rational beliefs, then rational beliefs are also purely determined by evidence. Hence, Kyburgian probability is evidentialist. To elucidate this point, I shall return to these ideas from Section 2:

Proportionality: If Kyburgian probabilities correspond to the proportions of belief warranted by the evidence, then a Kyburgian's belief-states will be proportionate to their evidence.

Uniqueness: The Kyburgian probability of any particular argument is unique. The probability changes only in the sense that, when our total evidence changes, a new probability is legislative for our beliefs.

Representation: Kyburgian probability provides a flexible formalism for representing belief-states and their features. For instance, in a simple binomial trials set-up, such as conjecturing the result of the next toss in a series of exchangeable tosses of a coin with unknown bias, Kyburgian probability intervals will narrow as evidence accumulates, which represents a reduction in uncertainty (Kyburg, 1968 p. 63). The intervals can widen, but only when our total evidence becomes more ambiguous (Seidenfeld, 2007) or potentially when it contracts.

Exclusion: Prior probabilities are a major issue for standard Bayesianism's compatibility with evidentialism (Kelly, 2016). There are no priors in Kyburgian probability, so this worry is inapplicable[25]. The rules of Sharpening are purely mechanical for given total evidence[26].

5.2 Some Challenges

Kyburgian probability differs in many ways from Imprecise Bayesianism, so there might be problems for its compatibility with evidentialism that I have not yet considered. I shall briefly outline some potential issues and argue that these do not constitute problems for Kyburgian evidentialism.

Frederick Benenson (1984, p. 203) argues that because Kyburgian probability functions are determined by reference class statements about a particular domain of discourse, there is an arbitrary aspect of their determination. Statements about frequencies make different assertions depending on the domain of discourse: to assert that '0.1 of a googolplex-fold domain satisfy Φ' is different from asserting that '0.1 of a ten googolplex-fold domain satisfy Φ'. More generally, Benenson argues that any definition of probability using finite frequencies (in Kyburg's case, statements about finite frequencies) in a particular domain involves an arbitrary choice of domain size, which affects e.g. statistical inferences.

[25] The closest things to priors is background statistical knowledge (e.g. that a phenomenon is approximately normally distributed in a population) but this information is not a 'prior' in sense of ideal or actual credences, nor Carnapian "logical" measures. Additionally, there are Kyburgian probabilities given evidence just consisting of all relevant logico-mathematical truths: [1, 1] for logical truths, [0, 0] for logical falsehoods, and [0, 1] otherwise.

[26] A critic of Kyburg's theory might argue that Sharpening is arbitrary, but I am denying subjectivity *in* Kyburgian probability, not subjectivity *of* Kyburgian probability.

Implicitly, this raises an incompatibility of Kyburgian probability with **Uniqueness**, and therefore evidentialism.

However, this decision is only arbitrary if our evidence is silent on the size of our domain of discourse. Benenson does not substantiate this claim, while it is hardly obvious. Consider the size of the universe. Before Aristarchus, the Ancient Greeks had apparently evidence-based reasons to believe that the universe was small, since the stars appeared to be nearby. In contrast, given Aristarchus's heliocentric theory, there were good reasons to think that the stars are far away and that the universe is relatively large, to explain the absence of stellar parallax. Similarly, the observations at the Hooker Telescope by Edwin Hubble and others provided evidence that the universe is much bigger than the Milky Way, by determining that galaxies like Andromeda are separate and extremely distant from the Milky Way[27]. Of course, our estimates are approximate (we should not accept any *exact* estimate of the size of the universe) but approximation does not imply subjectivity. Benenson is right that estimates of domain size are important in Kyburgian probability, but these estimates can be based on evidence.

White (2010, pp. 169–171) offers an intriguing argument for why philosophers like Kyburg, who think that probabilities should be determined using evidence about frequencies, should be precise probabilists. White proves that this doxastic norm:

Frequency-Credence: If our total evidence includes statements of the form '*i* satisfies predicate *N*' and 'The frequency of *M* among *N*'s is *r*' and no further relevant evidence about whether *i* is an *N*, then our credence in *Ni* should be *r*.

– entails:

The Principle of Indifference: Given an *n*-fold set $\mathcal{V}$ of statements for which we know that one and exactly one statement is true, plus symmetric evidence with respect to each statement, the probability of each statement in $\mathcal{V}$ is $\frac{1}{n}$.

Frequency-Credence thus commits us to unique precise probabilities given symmetric evidence for each element of $\mathcal{V}$. White's argument works because we know that the relative frequency of truth in $\mathcal{V}$ is $\frac{1}{n}$. Therefore, if our only relevant evidence regarding *Ni*'s truth is that *Ni* is a member of $\mathcal{V}$, then **Frequency-Credence** entails that our degree of belief in *Ni* should also be $\frac{1}{n}$. which is also the value that **The Principle of Indifference** mandates for *r*. Kyburgian probability might seem to commit us to something like **Frequency-Credence**, leading us into precision via **The Principle of Indifference**. If evidentialists have good reasons for adopting imprecise belief-states, this precision would be a problem for Kyburgian evidentialism.

[27] The details of such reasoning depend on one's general epistemology. Donald C. Williams (1947, pp. 110–112) suggests how enumerative induction could provide the necessary inferential steps.

Kyburg anticipates this line of reasoning. To avoid it, he limits the application of Sharpening to exclude metalinguistic terms like 'logically possible' and 'true' from those that can demarcate reference classes (Kyburg, 1974 pp. 176–180 and 314–316). Hence, the relative frequency of $\mathcal{V}$ cannot be used for a Kyburgian probability. This limitation is justified by an antecedent opposition to **The Principle of Indifference**. Whether this opposition is justified is beyond my scope. The point is that Kyburgian probability does not collapse into precise probabilism, and thus permits (and requires!) imprecision in some belief-states, as some evidentialists want.

6 Inertia

6.1 Inertia and Kyburgian Probability

When do Kyburgian probabilities avoid inertia? Although Kyburg (Kyburg and Teng, 2001 pp. 254–259) and others[28] note that Kyburgian probabilities can sometimes escape the [0, 1] interval, the conditions under which this occurs are mostly unexamined. I shall argue that they are surprisingly general.

Let H be the hypothesis that i is a Φ. Let E_1 be our initial total evidence. E_1 contains all logico-mathematical truths[29], plus the assertion that i exists, and otherwise the evidence in E_1 is irrelevant to H. Let R_i be the statement that i is in $\mathcal{K}_i$, where $\mathcal{K}_i$ is the unit set consisting of i, and that the frequency of Φ in $\mathcal{K}_i$ is in the interval [0, 1]. R_i is a logico-mathematical truth, so by an earlier remark R_i is in E_1. By supposition, there are no other reference class statements in E_1 with respect to i whose interval differs from [0, 1]. Given such austere total evidence, R_i wins by a walk-over, so that $KP(H, E_1) = [0, 1]$.

We now consider a new body of evidence E_2 formed by adding a reference class statement to E_1. For example, on Kyburg's theory of statistical inference, we can provisionally accept that a population frequency is somewhere within a confidence interval $[x, y]$ if the confidence level for a sample exceeds some contextually determined level α, in the absence of defeating background information. By assumption, there is no *relevant* background knowledge. Suppose E_2 contains both sample data and a statistical statement R_j inferred from this data. R_j asserts that i is a member of $\mathcal{K}_j$ and that the frequency of Φ in $\mathcal{K}_j$ is in the interval $[z, w]$ where $z \neq 0$ or $w \neq 1$. ('Or' is inclusive here.) Either R_i and R_j can only be favoured over the other by **Richness**, **Specificity** or **Precision**.

[28] See Levi (2007, p. 265) although he misdescribes this avoidance of inertia as "*creatio ex nihilo*". In Kyburg's system, expanding or contracting our total evidence is a necessary condition for more precise beliefs. What Kyburgian probability allows is *creatio ex notitia* – creation from information.

[29] This condition could be weakened, but it removes some unimportant complications.

Neither statement can defeat the other by **Richness** or **Specificity**, because those rules only allow us to Sharpen away conflicting intervals (neither interval is a subinterval of the other) and no interval can conflict with the vacuous [0, 1] interval[30]. There is no defeat by these rules even if $\mathcal{K}_i$ is a narrower reference class than $\mathcal{K}_j$ (which will generally be the case, since $\mathcal{K}_i$ just consists of *i*) because **Specificity** requires conflict.

Yet R_j will Sharpen away R_i by **Precision**, because any interval other than [0, 1] will be more precise. In detail, the only relevant reference class statements in E_2 are R_i and R_j, so that there are just two $\mathcal{F}$ sets:

$\mathcal{F}_1$: {[0, 1]}

$\mathcal{F}_2$: {[*z*, *w*]}

[*z*, *w*] must narrower than [0, 1] if $z \neq 0$ or $w \neq 1$. Hence, the cover of $\mathcal{F}_2$ must be shorter than the cover of $\mathcal{F}_1$. By **Precision**, $KP(H, E_2) = [z, w]$. Therefore, if we infer any reference class statement that is narrower than *Ri*, we obtain a non-vacuous rational belief-state. Provided it is possible to ampliatively infer a non-vacuous reference class statement about the hypothesis's subject, there will be no inertia.

For example, suppose that you are about to select a card from a Tarot deck. You have no prior evidence about the composition of Fool cards in this deck, nor about the selection procedure, so [0, 1] is your initial Kyburgian probability for the argument from your total evidence to the hypothesis that you will draw a Fool card. Suppose that I secretly saw the Tarot reader compiling the deck, and I observed that 90% of the cards are Fools. Imagine I tell you what I saw. Assume also that you believe that I am perfectly reliable on this matter. You now have statistical information about the card drawing that is non-deductive evidence, favouring the prediction that the card will be a Fool, which modifies the Kyburgian probability from [0, 1] to [0.9, 0.9]. Consequently, there is no inertia in this example[31].

Note that, in this example, you are agnostic about the (physical) randomness of the selection procedure. An Imprecise Bayesian *could* reason in a similar way, if they made an assumption of random selection. The problem is that, if we assume evidentialism, then this assumption cannot be made, since there is no evidence for it. In Kyburg's theory, without evidence regarding the selection procedure, it is information about the proportions (the composition of the deck) that provides the posterior belief state, due to **Precision**. Of course, if you had background knowledge about non-randomness in the selection procedure, then that would dominate your merely proportional beliefs by **Richness** or by **Specificity**. Our awareness of the potential relevance of such background knowledge is perhaps why the

[30] Since I have confined 'interval' to unit intervals, reference class statements must assert that a frequency is somewhere in [0, 1].

[31] For an extended example using enumerative induction, see (Kyburg and Teng, 2001 pp. 254–259).

reasoning feels somewhat "insecure" – our evidence could plausibly be defeated by subsequent learning.

The Kyburgian might seem to also be making an assumption of randomness. However, there are a few non-trivial differences. Firstly, a belief that the selection procedure was random is not the only belief that would warrant believing that one's information about the deck's proportions is the appropriate reference class statement. For example, the selection procedure could be *more* likely to produce representative samples than a random selection. Imagine a magician's trick selection technique, which had a lower standard deviation of representative samples than occur with random selection, and yet [0.9, 0.9] was still the long-run hypothetical frequency of Fool cards given the deck's composition[32]. Secondly, consider the case where the Kyburgian assigns the probability, but has not yet drawn the card, and discovers that the selection will be selected with a procedure that is strongly biased against Fools. If the Kyburgian was assuming a random selection, then they would have to weigh their new discovery against this background assumption. However, in Kyburgian probability, the evaluation of the new information about selections must be weighed against the proportional data, rather than against any rival belief about selections. We should understand this shift as a change from (1) agnosticism about the selection procedure and use instead of proportional data, to (2) a definite belief about that procedure that is evidentially superior to the proportional data. A change in assumptions would be a shift from one *definite* belief about selections to another such belief.

Kyburg advocates these the rules of Sharpening by arguments other than the avoidance of inertia. He argues that they provide intuitive results in various examples with different sorts of background knowledge (e.g. in Kyburg and Teng, 2001 Chapters 9 and 11). Additionally, he argues that reasoning with Sharpening provides a plausible reconciliation of classical statistics and Bayesian statistics, plus resolutions to various puzzles about inductive inference (Chapters 11-12; see also Kyburg, 1974). Finally, he also makes some model-theoretic arguments in favour of the relatively controversial **Precision** rule (Chapter 10; see also Kyburg 1997). While **Richness** and **Specificity** are formalisations of Hans Reichenbach's intuitively plausible injunction to use the narrowest known reference class data (Reichenbach, 1961 p. 316), **Precision**'s status is more questionable, especially since it is the source of divergences from Imprecise Bayesian conditionalisation (Seidenfeld, 2007). Kyburg proves that the **Precision** rule, applied to a rich but finite formal language, will produce probability intervals that match the minimum and maximum frequency or measures in any model that (model-theoretic analogues of) the other rules identify as relevant. Kyburg also proves the converse of this result. These model-theoretic properties arguably correspond to adaptations of deductive soundness and completeness to ampliative reasoning, and thus seem to count in **Precision**'s favour. Of course, all of his arguments could be contested. After all, no theory of probabilistic inference seems to have any knock-down arguments; they are

[32] Should we then interpret the assumption as a disjunction of all such sampling hypotheses such that the relative frequency of selecting a Fool is the same as the proportion in the deck? Psychologically, this interpretation is implausible as an explanation of why the Kyburgian feels comfortable using the proportional data in such cases, since such a disjunction would be very large and perhaps even infinite. Epistemologically, it would be very uncharitable as a reconstruction of the Kyburgian's reasoning.

lacking in such glory. Nonetheless, the absence of inertia in Kyburgian probability is not an *ad hoc* response, because there are other reasons to be a Kyburgian. Thus, at least one imprecise probability theory avoids inertia, while not being a merely technical response to it.

6.2 Kyburgian Probability and Imprecise Bayesianism

What is the source of these differences between Kyburgian probability and evidentialist Imprecise Bayesianism? Recall the **Chance Grounding Thesis**. This principle makes the elimination of physical probability statements from $\mathcal{O}$ into a *sine qua non* of greater precision in belief-states. In essence, our belief-state is formed using every probabilistic *possibility* given our total evidence. This is a very epistemologically cautious stance. To add to the cautiousness, Imprecise Bayesianism (as I have defined it) proceeds by an extension of conditionalisation to updating probability functions; see **Rule (2)** in Section 3. Conditionalisation is also a very cautious rule, e.g. if $0 < P(H) < 1$, then the conditional probabilities $P(H \mid E) = 0$ or $P(H \mid E) = 1$ only occur when E implies or contradicts H. Therefore, it is unsurprising that combining the **Chance Grounding Thesis** and conditionalisation would result in numerous cases of inertia: the former requires us to take seriously the possibilities of extreme physical probabilities, while the latter means that only deductive inference can stop us from taking the most extreme possibilities seriously.

Kyburgian probability also begins from an epistemically cautious stance. In the absence of relevant evidence, we know that unit frequencies of 0 or 1 for a contingent event are possible. Therefore the [0, 1] interval is part of our total evidence and *ex hypothesi* it will not be Sharpened away by our initial evidence. Prior opinion, unless substantiated by evidence, must be disregarded. Despite genuine differences, the basic epistemological starting point is the same as in evidentialist Imprecise Bayesianism.

The difference thus lies in the updating methods. Space considerations preclude a full discussion, but Kyburg's key idea is that if H is highly probable to at least a level α relative to E_T and H is the most informative hypothesis about its subject matter to exceed α, then H should be incorporated into our beliefs at a standard of acceptance α. The simplest case is when the probability (relative to our total evidence) exceeds α that a sample is approximately representative of a population. However, more sophisticated cases of acceptance are possible with richer statistical information, including Bayesian statistical inferences (Kyburg and Teng, Chapter 11). Kyburgian acceptance is tentative: new evidence might undermine our inference of H. Nonetheless, this is a less epistemologically cautious procedure than conditionalisation, because H can be accepted into our total evidence without being deductively implied by it. Depending on the particular evidentialist epistemology, α could be determined by contextual factors or perhaps have some fixed value. For example, $\alpha > 0.5$ would correspond to the intuition that we should believe hypotheses that are "more probable than not" given our evidence.

Once it is accepted, H can influence other probabilities via Sharpening. In particular,

even though it is possible that e.g. a particular coin toss might have a physical probability of 1 or a physical probability of 0, these mere possibilities will be ignored under the conditions I described in Section 6.1. The most important Sharpening rule here is **Precision**, which favours (*ceteris paribus*) informative reference class statements. Again, this precedence is less epistemologically cautious than Imprecise Bayesian updating, insofar as a premium is being placed on informative statements.

Whether Kyburgians' boldness is justified is beyond my scope. I shall only note that the inclusion of *some* ampliative inference in any plausible epistemology seems unavoidable: not all of our apparently good arguments are deductively valid. Therefore, the mere fact that Kyburgian updating involves the possibility of erroneous acceptance and ignoring certain physical probabilities when forming belief-states does not prove that it is flawed.

Still, the divergences of Kyburgian updating from conditionalisation are concerning. There are many good (but I think inconclusive) arguments for conditionalisation. Kyburg was critical of the pragmatic arguments, like diachronic Dutch Book Arguments (Bacchus *et al*, 1990). However, there are also epistemological arguments, e.g. accuracy-based justifications (Greaves and Wallace, 2006; Pettigrew, 2016) which are generally very interesting for evidentialists[33]. Kyburg does not discuss accuracy-based arguments for conditionalisation, which flourished late in his life. The successful extension of such arguments to imprecise beliefs (an ongoing project) would undermine Kyburgian updating.

As I cannot do justice to that debate, I shall instead present a small olive branch between Kyburgians and Imprecise Bayesians. Even if we suppose that the former is a better ideal epistemology, it has some pragmatic defects compared to Imprecise Bayesianism. In particular, Kyburgian updating requires constant recalculation of the probability (via Sharpening) with reference to the *whole* of one's new total evidence to check for newly relevant statistical information, because there is not necessarily a sequence of conditional probabilities that correspond to Kyburgian updating. In contrast, Imprecise Bayesians can use their old conditional probabilities to calculate new probability intervals. This pragmatic consideration is connected to epistemic aims, as a less arduous procedure would increase our ability to acquire true beliefs.

Here, Kyburgians can take inspiration from scientists, who often use idealized models for computational tractability, and thus sometimes abandon the aim of exactly mirroring nature in favour of other epistemic objectives. Kyburgians might likewise sometimes adopt Imprecise Bayesian updating as a useful simplification of ideal updating. Kyburgians would thereby abandon the aim of exactly mirroring an ideal reasoner, but that is a familiar necessity when we leave epistemology seminars.

From a Kyburgian perspective, there are special advantages of Imprecise Bayesianism for such a simplification. Firstly, conditionalisation-style reasoning often occurs in Kyburgian probability, since reasoning in accordance with **Richness** is often very similar to imprecise

[33] But not always: see the discussion of Joyce's arguments for Probabilism in (Easwaran and Fitelson, 2012).

conditionalisation. The lexical priority given to **Richness** in Sharpening means that this conditionalisation-style reasoning generally dominates the less Bayesian aspects of Kyburgian probability when we have rich multi-dimensional statistical information, as we generally do in science (Kyburg and Teng, 2001 pp. 262–264). Consequently, if we assume richer statistical information than we actually have, we can generally make Kyburgian updating match Imprecise Bayesian reasoning. This suggests a basic epistemological affinity between the two approaches, which is not shared between (say) Popperian learning methods and Kyburgian probability. Secondly, Imprecise Bayesians allow interval-valued beliefs, unlike standard Bayesians. Kyburgians also favour interval-valued beliefs, albeit without convexity (Kyburg 1990 Chapter 14). From Kyburgian viewpoint, the "imprecision" in Imprecise Bayesianism is a positive feature compared to standard Bayesianism, even if the latter is computationally easier.

This olive branch does not dissolve the differences between the two systems. However, it shows how the choice between Kyburgian reasoning and Imprecise Bayesianism for *all* epistemic purposes is a false dichotomy.

7 Conclusion

Clearly, ampliative inference is crucial for Kyburgians. To justify their escapes from inertia, I would have to justify ampliative inference, which is more than a little beyond my scope. The relationship between Kyburgian probability and inertia can only be fully analysed in the context of a theory of ampliative inference. I stress that Kyburgian probability is compatible with many approaches to ampliative inference and scientific reasoning in general. For example, Kyburg rejected abductive inference, but his probabilistic acceptance rules might be supplemented by abductive rules. Overall, Kyburg's theory of probability is extricable from his broader philosophy of science.

My arguments do not contradict philosophers like Vallinder and White who contend that inertia afflicts *some* prominent imprecise probability theories. I have instead defended the compatibility of evidentialism and Kyburgian probability. My conclusions demonstrate that an evidentialist imprecise probabilism is still a promising direction for inquiry.

Acknowledgements

This article was written while working under the Horizon2020 project Grant ID 639276. I am also grateful to Rune Nyrup, Donal Khosrowi, Richard Williams, John Shepard, and Wendy Parker for their feedback on this article.

References

Al-Najjar, N. I. and Weinstein, J. [2009]: ‘The Ambiguity Aversion Literature: A Critical Assessment’, *Economics and Philosophy*, **25**(3), pp. 249–284.

Bacchus, F., Kyburg, H. E., and Thalos, M. [1990]: ‘Against Conditionalization’, *Synthese*, **85**(3), pp. 475–506.

Benenson, F. C. [1984]: *Probability, Objectivity, and Evidence*. London: Routledge and Kegan Paul.

Bradley, Seamus, "Imprecise Probabilities", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities/>.

Carnap, R. [1962]: *The Logical Foundations of Probability*. Chicago: University of Chicago Press.

Cox, D. R. [2006]: *Principles of Statistical Inference*, Cambridge: Cambridge University Press.

Easwaran, K. and Fitelson, B. [2012]: ‘An ‘Evidentialist’ Worry About Joyce's Argument for Probabilism’, *dialectica*, **66**(3), pp. 425–433.

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

James, W. [1896]: ‘The Will to Believe’ *The New World*, **5** pp. 327–347.

Joyce, J. M. [2005]: ‘How Probabilities Reflect Evidence’, *Philosophical Perspectives*, **19**(1), pp. 153–78.

Joyce, J. M. [2010]: ‘A Defence of Imprecise Credences in Inference and Decision Making’, *Philosophical Perspectives*, **24**(1), pp. 281–323.

Kelly, Thomas, "Evidence", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/evidence/>.

Kyburg, H. E. [1968]: ‘Bets and Beliefs’, *American Philosophical Quarterly*, **5**(1), pp. 54–63.

Kyburg, H. E. [1974]: *The Logical Foundations of Statistics*, Dordrecht: Reidel.

Kyburg, H. E. [1990]: *Science and Reason*, Oxford: Oxford University Press.

Kyburg, H. E. [1997]: ‘Combinatorial Semantics: Semantics for Frequent Validity’, *Computational Intelligence*, **13**(2), pp. 215–257.

Kyburg, H. E. [2002]: ‘Don't Take Unnecessary Chances!’, *Synthese*, **132**(1/2), pp. 9–26.

Kyburg, H. E. [2006]: ‘Belief, Evidence, and Conditioning’, *Philosophy of Science*, **73**(1), pp. 42–65.

Kyburg, H. E. and Teng, C. M. [2001]: *Uncertain Inference*, Cambridge: Cambridge University Press.

Levi, I. [1977]: ‘Direct Inference’, *The Journal of Philosophy*, **74**(1), pp. 5–29.

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

Popper, K. [1980]: *The Logic of Scientific Discovery*. London: Hutchinson.

Reichenbach, H. [1961]: *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.

Rinard, S. [2013]: ‘Against Radical Credal Imprecision’, *Thought: A Journal of Philosophy*, **2**, pp. 157–165.

Rowbottom, D. P. [2008]: ‘On the Proximity of the Logical and ‘Objective Bayesian’ Interpretations of Probability’, *Erkenntnis* **69**(3), pp. 335–349.

Runde, J. [1990]: ‘Keynesian Uncertainty and the Weight of Arguments’, *Economics and Philosophy* **6**(2), pp. 275–292.

Seidenfeld, T. [1978]: ‘Direct Inference and Inverse Inference’, *The Journal of Philosophy* **75**(12), pp. 709–730.

Seidenfeld, T. [2007]: ‘Evidential Symmetry and Mushy Credence’, in William Harper and Gregory Wheeler (eds.), *Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr.*, London: College Publications, pp. 267–279.

Talbott, William, "Bayesian Epistemology", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/>.

Vallinder, A. [2018]: ‘Imprecise Bayesianism and Global Belief Inertia’, *British Journal for the Philosophy of Science*, **74**(4), pp. 1205–1230.

White, R. [2010]: ‘Evidential Symmetry and Mushy Credence’, in Tamar Szabó Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology Volume 3*, Oxford: Oxford University Press, pp. 161–186.

Williams, D. C. [1947]: *The Ground of Induction*, Cambridge, Mass.: Harvard University Press.