Economics & Philosophy (2026), **42**, 420–456
doi:10.1017/S026626712510059X

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Ambiguity, Coherence and Performance

William Peden[1], Mantas Radzvilas[2], Daniele Tortoli[3] and Francesco De Pretis[3,4]

[1]Institute for Philosophy and Scientific Method, Johannes Kepler University, Altenberger Strasse 69, 4040 Linz, Austria, [2]Department of Philosophy, University of Konstanz, Universitätsstrasse 10, Konstanz 78464 Germany, [3]Department of Communication and Economics, University of Modena and Reggio Emilia, Viale Allegri 9, 42121 Reggio Emilia, Italy and [4]School of Public Health, Indiana University Bloomington, 1025 E 7th St, Bloomington, IN 47405, USA
Corresponding author: Mantas Radzvilas; Email: mantas.radzvilas@uni-konstanz.de

(Received 07 April 2025; revised 21 August 2025; accepted 10 September 2025; first published online 06 November 2025)

Abstract
Imprecise Bayesianism has been proposed as an alternative to Standard Bayesianism, partly because of its tools for representing ambiguity. Instead of representing credences via precise probabilities, a set of probability distributions is used to model belief states. However, there are criticisms of Imprecise Bayesianism's update rule. A recent alternative update rule is Alpha Cut, which evades some of the primary criticisms of Imprecise Bayesian updating. We compare Alpha Cut with Imprecise Bayesianism and another alternative update approach called Calibration. We find that Alpha Cut has problems with respect to ambiguity, coherence, and performance qualities, whereas there are more promising alternatives.

Keywords: Alpha Cut; decisions under uncertainty; formal epistemology; imprecise Bayesianism; imprecise probabilities

1. Introduction

Many economists and philosophers want to represent epistemic ambiguity within a formal framework. For the purposes of this article, we focus on ambiguity in the sense of lacking information about physical probabilities that are relevant to the decision problem or epistemic context. This sense of "ambiguity" is prominent in descriptive as well as normative decision theory, such as in the Ellsberg Paradox (Ellsberg 1961) and the wider ambiguity aversion literature (Bradley 2024). Unfortunately, it has proven to be extremely hard to include ambiguity in a theory of rational decision-making without incurring heavy costs. Many philosophers and economists regard these costs as worse than being unable to represent ambiguity (Al-Najjar and Weinstein 2009; Elga 2010).

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

There have been a number of theories proposed for representing a rational agent's beliefs. One family that shares many details in common are the descendants of what we shall call "Standard" Bayesianism: the view that a rational agent's beliefs must be representable by a unique probability distribution and that this probability distribution should be updated by conditionalization.[1] Standard Bayesianism's capacity to represent ambiguity has often been questioned, but its probabilistic framework is highly attractive, partly because it provides inputs for an appealing decision theory – maximizing expected payoffs. To what degree can these strengths be preserved within a theory that can represent ambiguity?

The literature thus far has focused on comparisons of Standard Bayesianism with Imprecise Bayesianism. The latter uses sets of probability distributions to represent beliefs. However, Imprecise Bayesianism has its own issues, particularly with updating. In this article, we examine a recently proposed modification of Imprecise Bayesianism, called "Alpha Cut". It has also been called "α-cut" (Cattaneo 2014; Bradley 2022). This theory maintains the set-based belief representation framework of Imprecise Bayesianism, but updates by a newly developed rule. Thus, it is prima facie conceivable that Alpha Cut can retain Imprecise Bayesianism's tools for ambiguity representation, while offering a better update rule.

Alpha Cut is one of the most recent implementations of making hierarchical distinctions among probability distributions in the set (Gärdenfors and Sahlin 1982; Pearl 1988; Wilson 2001; De Bock and De Cooman 2015; Moral 2019; Omar and Augustin 2019). This contrasts with the approach of most Imprecise Bayesian philosophers: to regard each member in the set equally when modelling belief states or rational decision-making. The hierarchical approach has recently attracted significant interest in decision theory and formal epistemology, partly due to the challenges for traditional Imprecise Bayesianism (Hill 2013; Bradley 2017; Lyon 2017; Hill 2019; Lassiter 2019).

We evaluate Alpha Cut to assess (1) the extent to which it avoids problems of Imprecise Bayesianism and (2) the new problems it introduces. We also compare Alpha Cut against another approach, which we shall call "Calibration". We argue that, at least with respect to the issues we discuss, someone willing to pay the price of Alpha Cut should view Calibration as a better option.

We begin in section 2 by explaining the desiderata that we shall use for comparisons. In section 3, we briefly explain Standard Bayesianism and some criticisms of it that have motivated Imprecise Bayesians. In section 4, we explain Alpha Cut, then compare it with Imprecise Bayesianism. Finally, in section 5, we compare Alpha Cut and Calibration. We conclude in section 6 that there is hope for a theory that has consistently strong ambiguity, coherence and performance properties, but Alpha Cut is not it.

2. Three Desiderata for Theories

By "theories", we shall mean a formal model of rational beliefs and reasoning, in formal epistemology or decision theory. Standard Bayesianism, Imprecise

[1]As we use the term, Standard Bayesianism includes both "objective" and "subjective" versions of this theory.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Bayesianism, Alpha Cut and Calibration will be the theories that we discuss in this article. There can be considerable variety in different versions of a theory, such as Standard Bayesianism with strict conditionalization or with Jeffrey conditionalization.

There are varying intuitions about rationality that make it hard to compare theories. In general, our approach is to focus on a set of criteria that, while insufficient for precisely defining a "rational" agent, nonetheless capture desiderata that many formal epistemologists and decision theorists share. This approach enables us to make comparisons using a relatively small number of presuppositions about the answers to controversial debates. Thus, we shall not discuss issues like epistemic accuracy (Greaves and Wallace 2006; Norton 2021) or unmeasurable sets (Isaacs *et al.* 2021; Goodsell and Nebel 2024).

2.1 Ambiguity Representation

Philosophers and economists have distinguished between evidence's "balance" and its "weight" (Keynes 1921: Ch. VI). The balance of evidence E tilts towards a hypothesis H if learning E increases the evidence in favour of H (put another way, if learning E confirms H). In contrast, the *weight* of evidence is the total amount of relevant information that E provides with respect to H (Popper 1958; Davidson and Pargetter 1987; Franklin 2001; Reiss 2014; Kasser 2015).

To see how these can come apart, consider the hypothesis that a particular coin toss will land heads. Suppose that, for each subsequent toss, an agent has a prior of 0.5 that the coin will land heads. A random sample of 10 coin tosses, all of which land heads, can (with suitable background knowledge) be balanced in favour of the hypothesis that the coin will land heads, but this evidence has low weight. Conversely, a sample of 1000 tosses, 50% of which landed tails, seems to increase significantly the weight of evidence with respect to the coin toss landing heads (by increasing the information about the coin's bias) even though its balance is neutral towards the hypothesis that the coin will land heads, given the prior.

We shall say that evidence is more "ambiguous" about a hypothesis insofar as its weight for that hypothesis is lower. Representing ambiguity is an alleged advantage of imprecise probabilities (Keynes 1921; Walley 1991; Joyce 2005; Sturgeon 2008; Peden 2018; Bradley 2019).

A closely related issue is the representation of neutrality. Many have thought that one should have a completely neutral doxastic attitude towards a hypothesis H when in a state of complete ambiguity regarding H. But how should that neutrality be represented in a formal model of beliefs? The traditional Bayesian answers are the Principle of Indifference and more generally the Maximum Entropy Principle. However, while these principles can represent neutrality with respect to a partition of the fundamental possible outcomes, this is at the cost of creating strong beliefs for or against many other hypotheses (Kyburg 1968).

A connected putative advantage of ambiguity representation is that it enables the incorporation of ambiguity aversion into decision rules. Ambiguity aversion occurs

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

when an agent has a preference against choices with less precise information about the relevant physical probabilities.[2] The classic example is the Ellsberg Paradox, where many people prefer options with known physical probabilities in two sets of choices between alternative betting situations, even though there are no Standard Bayesian probabilities that permit this set of preferences (Ellsberg 1961). Similar phenomena continue to be studied in experimental psychology and behavioural economics (Fox and Tversky 1995; Halevy 2007; Ahn *et al.* 2014; Jia *et al.* 2020). Due to the significance of ambiguity aversion in economics and decision theory more widely, we shall focus on this type of ambiguity in this article. Additionally, for the purposes of this article, we shall assume that being able to represent ambiguity in an intuitive way is desirable *ceteris paribus*, but only *ceteris paribus* – other considerations can override this quality when choosing between theories. Furthermore, while our discussion is relevant to descriptive decision theory, we shall focus on the normative interpretations of the theories in question.

2.2 Coherence

An agent is synchronically coherent if they avoid accepting any synchronic Dutch Book, which is a set of bets, given particular beliefs and preferences, that would result in a guaranteed net loss of payoffs. In the context of this article, "payoffs" are utility numbers representing a player's preferences over betting outcomes. An agent is diachronically coherent if they are such that, when they make plans for a sequence of decisions, they avoid accepting diachronic Dutch Books – a set of bets and odds, over a sequence of decisions, that would inevitably lose payoffs. The significance of diachronic coherence has been debated (Ramsey 1988; Bacchus *et al.* 1990; Milne 1991; Rowbottom 2007; Williamson 2011; Pettigrew 2020; Gustafsson 2022; Vineberg 2022). However, as with tools for ambiguity representation, we shall assume that it is a desirable property of a theory, *ceteris paribus*.

2.3 Decision-making Performance

Coherence properties for decision rules are possibility results: they prove that a certain flawed combination of decisions is impossible if a theory is used. Thus, by themselves, they imply nothing about the overall performance of decision rules in practice.

Convergence results are more informative. These show that, under certain conditions, using a theory will lead to beliefs that approach the true physical probabilities in the long-run. Unfortunately, all the theories we discuss in this article have roughly equally attractive long-run convergence results.

Instead, one can compare theories with respect to short-run decision-making: an agent's choices during the period in which cumulative sample frequencies can have substantial deviations from the true limiting relative frequencies, due to random fluctuations. Comparing theories' short-run decision-making performances via agent-based modelling was briefly explored by Henry E. Kyburg and Choh Man

[2]A physical probability distribution is a non-epistemic quantity satisfying the probability axioms, such as propensities and long-run relative frequencies.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Teng (Kyburg and Teng 1999) and it has been pursued recently by refining and expanding their tests (Radzvilas *et al.* 2021, 2023, 2024, 2024a).

The focus thus far has been on a decision problem based around sequences of Bernoulli trials – events that are binomial and exchangeable; we explain Bernoulli trials in more detail later. The focus on Bernoulli trials makes sense because of the prominence of exchangeable events in statistics. Furthermore, the relative simplicity of Bernoulli trials improves the capacity of the agent-based modelling to clearly and quickly identify the causes of differences in performance. We shall call this decision problem "the Bernoulli task". Its crucial traits are shared by many other types of statistical problems, such as multinomial generalizations of Bernoulli trials. We shall use this decision problem as the basis for short-run performance comparisons, while acknowledging that it does not provide an exhaustive comparison across all decision problems, which are potential topics for further research.

3. Standard Bayesianism

Our main goal in this article is to evaluate theories that are descendants of Standard Bayesianism. Therefore, we shall not discuss Standard Bayesianism in detail. Instead, we shall review the relevant results for Standard Bayesianism, to prepare the ground for later comparisons.

In our use of the term, Standard Bayesians (1) have credences over a domain and these credences can be represented by a (finitely or σ) additive probability distribution and (2) update by strict conditionalization. We shall also define Standard Bayesians as expected payoff maximizers in their choice behaviour. One difference from many discussions of Bayesianism in epistemology is that we shall assume that Standard Bayesians' credences are distributed over events rather than statements, because it is more felicitous for using relevant terminology from statistics.

3.1 Representation

Criticisms of Standard Bayesianism's representation of ambiguity are numerous, both in decision theory (Ellsberg 1961; Feduzi 2010; Bradley 2017; Bradley 2019; Hill 2019) and formal epistemology (Popper 1958; Kyburg 1974; Joyce 2005; Norton 2011). Firstly, Karl Popper's "Paradox of Ideal Evidence" challenges Standard Bayesianism's capacity to represent changes in ambiguity (Popper 1958). In a wide range of contexts, when a sample frequency is equal to the prior, the reduction in ambiguity will not be registered by a difference in the posterior. There have been efforts to develop formal analyses of ambiguity (often in terms of analysing the weight of evidence) some of which may be consistent with Standard Bayesianism, but there is no consensus method (Carnap 1962; Kuipers 1976; Davidson and Pargetter 1987; Runde 1990; Nance 2016; Lassiter 2019).

Secondly, another challenge is representing a neutral doxastic state, which arguably should occur if ambiguity is maximal. Consider a sequence of 1000 coin tosses, where the coin's bias or fairness is unknown. A Standard Bayesian prior can be neutral in the sense that each coin toss has a 50% probability of landing heads, but not simultaneously neutral towards every intersection, union or generalization

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

of these coin tosses. Since probability assignments to partitions of events will have consequences for partitions of compounds of those events, Standard Bayesian agents cannot avoid having strong beliefs even in situations of extreme ambiguity. In the coin tossing example, a Standard Bayesian who assigns a 0.5 prior to each individual coin toss landing heads will be almost certain that approximately 500/1000 of the coin tosses will land heads, despite the highly ambiguous situation.

Overall, ambiguity representation is not a strength of Standard Bayesianism. The development of all the theories that we subsequently discuss has been partly motivated by a desire to have a theory that has most or all of Standard Bayesianism's virtues, yet with better ambiguity representation properties. In these alternatives, Standard Bayesian reasoning effectively occurs as a special case that occurs in the presence of little or no ambiguity.

3.2 Coherence

As is well-known, Standard Bayesianism is synchronically and diachronically coherent. In the context of this article, the key point to note is that Standard Bayesianism is sufficient but not necessary for diachronic coherence, as we shall explain in the next section.

3.3 Performance

Before describing the performance of the most studied Standard Bayesian player in agent-based models of the Bernoulli task, we provide the broad details of this decision problem. In the Bernoulli task, a player has a choice of whether and how to bet on the result of a Bernoulli trial[3] after observing a sample of preceding trials. Like earlier philosophical literature using this decision problem, we shall talk about "coin tosses", but with the proviso that, unlike real people with real coins, a player begins in a state of complete ambiguity with respect to the true values of the biases.

A player of the game does not have any initial information about the coin toss bias. However, they do know the payoffs and that the Bernoulli task consists purely of Bernoulli trials. The aim in the decision problem is exclusively to maximize payoffs in each bet. There is no player interaction, nor additional player goals.

A "game" consists of 4 new observations of coin toss results, followed by a choice between c_h, to bet on heads, c_t, to bet on tails, or c_a, to abstain from betting. A "test" consists of a sequence of 1000 games, with a fixed coin bias. We tested players using coin biases of 0.1, 0.3, 0.5, 0.7 and 0.9, since players' relative performances vary among different biases.

Players recall earlier observations. For instance, after 10 games, a player adds their knowledge of the previous 50 tosses to their 4 new observations when making their decision in the 11th game. As the test progresses, a player accumulates a "history" – informally, the history is their net sample of observations of the coin

[3]A sequence of Bernoulli trials is an ordered set of events with two possible types of outcome (the events are binomial) such that the probability of a type of outcome on each trial in the sequence is the same (the events are exchangeable).

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

ω_i / c_j	ω_h	ω_t
c_h	$(1-\delta)$	$-\delta$
c_t	$(\delta-1)$	δ
c_a	0	0

Figure 1. Player Payoff Matrix. ω_h and ω_t are respectively the events of the coin landing heads and the coin landing tails.

tosses in that particular test. Depending on their choice and the outcome of the coin toss, they receive a payoff described in Figure 1.

The randomly generated parameter δ takes values in the continuum from 0 to 1. δ determines how favourable the structure of a game is between heads and tails. When comparing players in tests, we always use the same sets of δ values and coin toss sequences.

We provide a more detailed description of the game structure in the Bernoulli task in Appendix 6.1. We do the same for the payoff structure in Appendix 6.2.

Although there are many ways that a Standard Bayesian might engage with the Bernoulli task, the literature thus far has focused on the use of beta distributions. Since beta distributions are a common approach in statistics and decision theory for addressing reasoning tasks like the Bernoulli task, this choice makes sense. Beta distributions are a type of prior that can be summarized by a function $Beta(a, b)$. The parameters a and b are called "hyperparameters": they are not probabilities, but rather numbers for describing the shape of the continuous probability distribution on the interval of $[0, 1]$ representing all the possible biases of a coin.

The best-performing Standard Bayesian agent among those investigated so far is called *Stan*. They have a flat prior, $Beta(1, 1)$, which means that their prior is equivocal (it leads to equal probability assignments for heads and tails) and soft (the posteriors rapidly converge to the sample frequency). The priors and posteriors of beta distributions are conjugate distributions, so as *Stan* observes more coin tosses, their credences continue to be beta distributions. We detail *Stan* in Appendix 6.3.

In games, *Stan* bets on tails when δ exceeds their credence that the next toss (the final toss in a particular game) will land heads. *Stan* will bet heads if δ is less than their credence in heads. They randomize if δ is equal to their credence.

Stan performs at least as well as any other player thus far examined. For the specific implementations investigated thus far, *Stan* performs better than Imprecise Bayesians (Radzvilas *et al.* 2024a), better than Dempster–Shafer reasoners (Radzvilas *et al.* 2024), and better or as well as players using Calibration (Radzvilas *et al.* 2023). We shall discuss the particular criteria for these evaluations in the next subsection.

4. Alpha Cut and Imprecise Bayesianism

Alpha Cut is a modification of the more well-known Imprecise Bayesian theory (Wilson 2001; Cattaneo 2008, 2014). It is alleged to have more attractive learning properties (Bradley 2022). Due to their similarities, we combine our definitions and discussions of these two theories.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

4.1 Definition

We provide detailed formal definitions of an Imprecise Bayesian agent and an Alpha Cut agent in Appendices 6.4 and 6.5. In this section, we provide a general and informal description of each approach.

Imprecise Bayesians use sets of probability distributions to represent belief states. We shall call these sets "credal sets". While credal sets do not have to be convex, we shall only consider convex credal sets in this article.[4] An agent's "belief interval" for a hypothesis is a closed interval that is the shortest cover of the range of values assigned to an event by the probability distributions in the agent's credal set. Sometimes, it will be convenient to discuss an agent's belief in a hypothesis in terms of a belief interval rather than the credal set.

Regarding updating, we shall define Imprecise Bayesianism narrowly. Where each probability distribution assigns non-zero prior probability to the evidence (the statements now assigned probability 1 by the agent) each probability distribution is strictly conditionalized on the evidence. Probability distributions assigning zero probability to the evidence are excised from the set (Gilboa and Schmeidler 1993). We shall refer to this update procedure as "generalized conditioning".

In contrast, Alpha Cut updates by excising probability distributions from the set if their prior for the evidence is less than the product of (a) the highest prior assigned to the evidence by any other probability distribution in the credal set and (b) a constant $\alpha \in (0, 1)$. Alpha Cut is intended to be a generalization of Imprecise Bayesians' excision of those probability distributions that assign zero to one's evidence: this move might be justified by thinking that such probability distributions are "unreliable" (Cattaneo 2008; Bradley 2022).[5] The role of multiplication by the highest prior for the evidence in the agent's credal set is to address the fact that, as evidence grows over time, it will generally have a lower prior (Bradley 2022). The meaning of "reliability" in this context needs more clarification by supporters of Alpha Cut, since it is not supposed to be a more reliable representation of the agent's doxastic states, nor a better correspondence of those doxastic states to objective limiting relative frequencies. However, there are similar approaches to Alpha Cut, whose advocates have provided deeper explanations of comparable parameters (Braithwaite 1953; Gärdenfors and Sahlin 1982; Hill 2013; Lyon 2017; Bradley 2017; Hill 2019; Lassiter 2019).

This process of excising or "cutting" distributions from the credal set is applied symmetrically. For instance, when considering whether a coin toss will land heads in the Bernoulli task, an Alpha Cut agent will excise probability distributions in a way that raises the lower limit of the belief interval for heads. Conversely, their excisions

[4]Roughly speaking, a convex credal set contains all the probability distributions (in this case, specifically beta distributions) between its most extreme distributions. For instance, a convex credal set ranging from $Beta(1, 99)$ to $Beta(99, 1)$ also includes $Beta(2, 99)$, $Beta(50, 50)$ and so on. For details, see Appendix 6.4. Some Imprecise Bayesians think that credal sets should be convex, others see convex sets as useful in some cases but not required (Levi 1980; Cozman 2012; Troffaes and De Cooman 2014; Bradley 2019; Wheeler 2021).

[5]However, this justification of Imprecise Bayesians' excision rule is not unique. If the prior of the evidence is zero and a probability distribution is a Kolmogorov probability function, then it cannot be used in generalized conditioning.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

will raise the lower limit of the belief interval for tails if they are considering whether the coin will land tails.

Why might an imprecise probabilist prefer Alpha Cut to ordinary Imprecise Bayesian updating? Perhaps the most prominent criticism of Imprecise Bayesianism involves a phenomenon called "inertia". If an Imprecise Bayesian agent's credal set includes dogmatic probability distributions with respect to a hypothesis (such as assigning priors ranging from 0 to 1, which is arguably appropriate in situations of maximal ambiguity) then the agent cannot change their belief state for that hypothesis, even given evidence that seems intuitively relevant (Levi 1980; Walley 1991). Note that it is the dogmatism that matters for inertia: sets containing probability distributions assigning all possible values are sufficient but not necessary for inertia. Therefore, simply avoiding credal sets with belief intervals of $[0, 1]$ for hypotheses, as some have suggested (Walley 1991; Rinard 2013; Benétreau-Dupin 2015), does not escape inertia. Other credal sets will also be inert if the distributions at their limits are dogmatic with respect to a hypothesis (Piatti *et al.* 2009; Vallinder 2018). For example, a credal set with distributions assigning 0.1 and 0.9 to a hypothesis H, where those extreme distributions are completely insensitive to evidence regarding H, will also be inert with respect to H. Moreover, avoiding inertia via imposing restrictions on credal sets reduces the ambiguity representation capacities of Imprecise Bayesianism.

One feature of Alpha Cut is that it can generate intervals that are inconsistent. In the context of the Bernoulli trial, the upper probability for heads is not equal to 1 minus the lower probability for tails, nor is the lower probability for heads equal to 1 minus the upper probability for tails. The same applies, *mutatis mutandis*, to the cut credal set used for an Alpha Cut agent's beliefs about the coin landing tails.

We discuss existing Alpha Cut responses to this issue in the last part of this section. However, one possible modification is best explained now. We shall use "MLAC" (Maximum Likelihood Alpha Cut) to refer to a rule that uses the observed sample mean for selecting a unique belief interval for an event, analogous to the belief intervals of an Imprecise Bayesian. In the context of the Bernoulli trial, this rule requires choosing to use the heads belief interval if the coin toss history so far has a majority of heads observations, the tails belief interval if the coin toss history so far has a majority of tails observations, and randomizing between the two if the coin toss history is equally balanced. Thus, an MLAC player effectively uses the sample mean for a maximum likelihood estimate-based choice of among their Alpha Cut intervals. We further define MLAC in Appendix 6.6. We report the results for *MLAC-Optimist* in Figure 3. MLAC has not previously been investigated in the literature, but we found that it has some interesting and distinctive performance properties.

In what follows, we first consider performance in the Bernoulli task, because Alpha Cut does not have any novel problems in this respect. Next, we consider its ambiguity representation properties, where its strengths and weaknesses are mixed. Finally, we consider its coherence properties, where it has very significant problems.

4.2 Performance

An Imprecise Bayesian or Alpha Cut agent cannot use expected payoff maximization as a general decision rule, since their expectations will generally

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

not be unique. Thus, an imprecise probability decision rule is necessary to test their performance in the Bernoulli task. In previous research using the Bernoulli task, the Optimist rule was consistently the best performer among imprecise probability decision rules, regardless of the player type (Radzvilas *et al.* 2023, Radzvilas *et al.* 2024, 2024a). We provide full details of this rule in Appendix 6.8. Informally, an *Optimist* player in our agent-based model sums their maximum and minimum expected payoffs, but weighs their maximum expected payoffs by $3/4$ and their minimum expected payoffs by $1/4$; this is one way of implementing the Hurwicz criterion (Hurwicz 1951). The Imprecise Bayesian Optimist player is called *IB-Optimist* and the Alpha Cut Optimist player is called *AC-Optimist*. Their credal set ranges from $Beta(99, 1)$ (biased towards heads) to $Beta(1, 99)$ (biased towards tails). We refer to *IB* as a generic Imprecise Bayesian player: a player with this initial credal set, but considered independently of the Optimist decision rule. This enables us to refer to what is true of any Imprecise Bayesian's learning behaviour with this credal set, regardless of their decision rule.

One feature of *IB*'s performances in the Bernoulli task is known as the Ambiguity Dilemma (Radzvilas *et al.* 2024a, 2024b). Imprecise Bayesianism's putative ability to represent differences in ambiguity depends on the difference between belief intervals and additive credences. For example, in the Bernoulli task, *Stan* starts with a credence of 0.5. After acquiring a sample of 50 tosses, 25 of which land heads, *Stan* has learnt more about the physical probabilities (the coin bias). Yet their credence is still 0.5. This is an instance of Popper's Paradox of Ideal Evidence. Some Imprecise Bayesians[6] argue that the width of the belief interval for a hypothesis can represent ambiguity, at least in problems like the Bernoulli task; equivalently, the degree of divergence among the credal set's credences for the hypothesis in question is supposed to measure (or at least indicate) ambiguity (Walley 1991; Joyce 2005).

The Ambiguity Dilemma is that these tools for representing ambiguity come at the cost of comparatively weak performance in the Bernoulli task. Regardless of the decision rule used, *IB* is outperformed on the Bernoulli task by *Stan*, even if *IB* uses non-inert sets that only somewhat represent the extreme ambiguity in the Bernoulli task. Some versions of *IB* can match *Stan* for some biases, but all are outperformed given at least one bias none outperforms *Stan* overall (Radzvilas *et al.* 2024a). Since Alpha Cut is a faster update rule than Imprecise Bayesians' generalized conditioning rule, one might expect that they can close this gap and escape the Ambiguity Dilemma.

We set α at 0.75. This value for *AC-Optimist* is a balance between two considerations. On the one hand, higher α results in faster convergence, which is a key selling point of Alpha Cut. On the other hand, higher α reduces the difference between Alpha Cut and approaches like Standard Bayesianism and maximum likelihood estimation (which, in the Bernoulli task, simply estimates that the coin bias is the sample frequency) thus reducing Alpha Cut's tools to represent ambiguity. In the limit, if $\alpha = 1$, an Alpha Cut belief interval would no longer represent differences in ambiguity after the first observation, since the credal set's minimum probability would simply become equal to the maximum probability. However, this problem and similar issues only arise once α is 1 or arbitrarily close to 1. Thus, since α should not be too high or too low, we investigated $\alpha = 0.75$.

[6]Not all Imprecise Bayesians agree (Bradley and Steele 2014b).

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

We created graphs of players' average profits (payoff) per game in the 1000 tests. The important performance differences are large enough to be clear by visual inspection. We report these results in Figures 2 and 3. We also made comparisons using methods from some earlier studies in this literature (Radzvilas *et al.* 2021, Radzvilas *et al.* 2024a) but we found no significant differences in these that were not reflected in the graphs, so we simply report the latter.

To perform the tests, we used Python (version 3.12.4) with the `statsmodels` econometrics and statistics library (Seabold and Perktold 2010). One script generated exchangeable coin-toss outcomes; a second produced δ values; the remaining scripts computed individual game payoffs and averaged the net profits of each player across tests. All tests were executed on an Ubuntu Linux server (64-core/128-thread Intel Xeon @ 1.3 GHz, 128 GB RAM).

As can be seen in Figures 2 and 3, there is no statistically significant difference in performance between *AC-Optimist* and *IB-Optimist*. Therefore, the use of Alpha Cut updating does not solve the Ambiguity Dilemma: *AC-Optimist*'s greater capacity to represent ambiguity comes at the cost of performing less well than *Stan*. Like *IB-Optimist*, the performance gap of *AC-Optimist* would be worse if they used a wider credal set, but this wider credal set would better represent the extreme ambiguity in the Bernoulli task.

This result is surprising, since Alpha Cut is designed to have a faster rate of convergence than Imprecise Bayesianism, but there is an explanation. Although *AC-Optimist* tends to have narrower belief intervals for the events than *IB-Optimist* in early games, it does so by boosting the lower interval for a toss landing heads *and* the lower interval for that toss landing tails. The result is that, when the coin toss has an extreme bias, *AC-Optimist* misses reliable information that *Stan* detects and uses.

For example, given any belief about the probability of coin landing heads ρ, the expected payoff from betting on heads is $\rho(1-\delta)+(1-\rho)(-\delta)=-\delta+\rho$, while the expected payoff from betting on tails is $\rho(\delta-1)+(1-\rho)(\delta)=\delta-\rho$. Assume that $\rho=0.9$ and $\delta=0.7$. The expected payoff from betting on heads is 0.2, while the expected payoff from betting on tails is -0.2. Hence, an expected payoff-maximizing player who knew the true bias would choose to bet on heads under these conditions.

We now imagine a possible second game in a test, assuming that 8/9 coin tosses (approximately 0.9 of coin tosses) have landed heads in the test so far. *Stan*'s posterior for heads can be found by adding this coin toss history of 8 heads tosses and 1 tails toss to the values from their $Beta(1,1)$ probability distribution as follows: $(1+8)/(1+1+9)\approx 0.8181$. (All of our approximations in this article's examples are to four decimal places, but our tests are far more precise.) Substituting this value for 0.9 in the calculations in the preceding paragraph, one can see that *Stan*'s expected payoffs are 0.1181 for betting on heads and -0.1181 for betting on tails. In contrast, with respect to heads, the *IB-Optimist* player has a maximum credence for heads of $(99+8)/(99+1+9)\approx 0.9817$ and a minimum credence of $(1+8)/(99+1+9)\approx 0.0826$. The maximum/minimum values for tails can be found by subtracting the heads minimum/maximum values from 1. *IB-Optimist* multiplies their expected payoffs for betting heads and for betting tails by 3/4 for the maximum expected payoff and 1/4 for the minimum expected payoff,

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

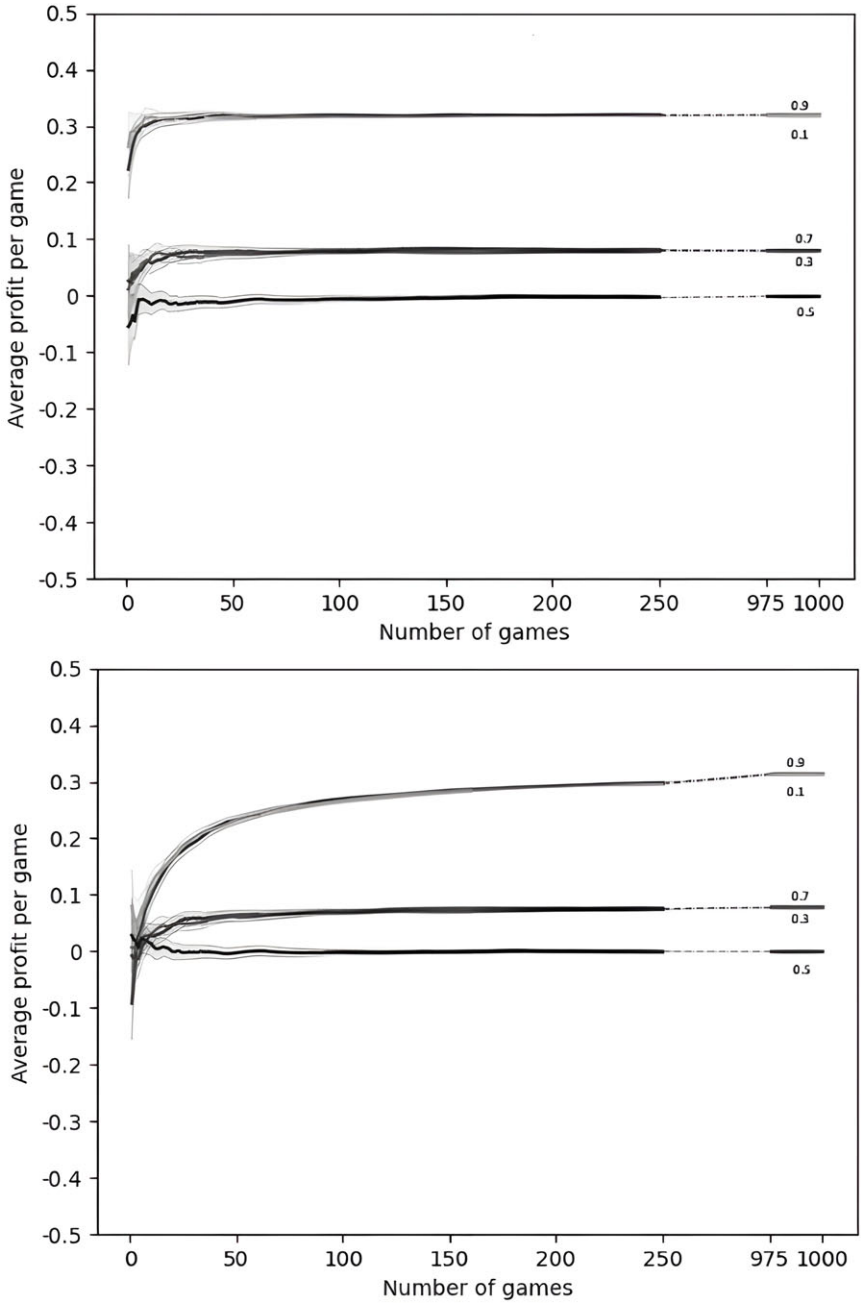


Figure 2. Bayesian Players: Stan (top) and IB-Optimist (bottom). The graphs show the average profit per game (y-axis) against number of games (x-axis). The solid lines are the averages. The confidence intervals around the lines are calculated at the 0.95 level. Numbers adjacent to the lines mark the coin bias.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

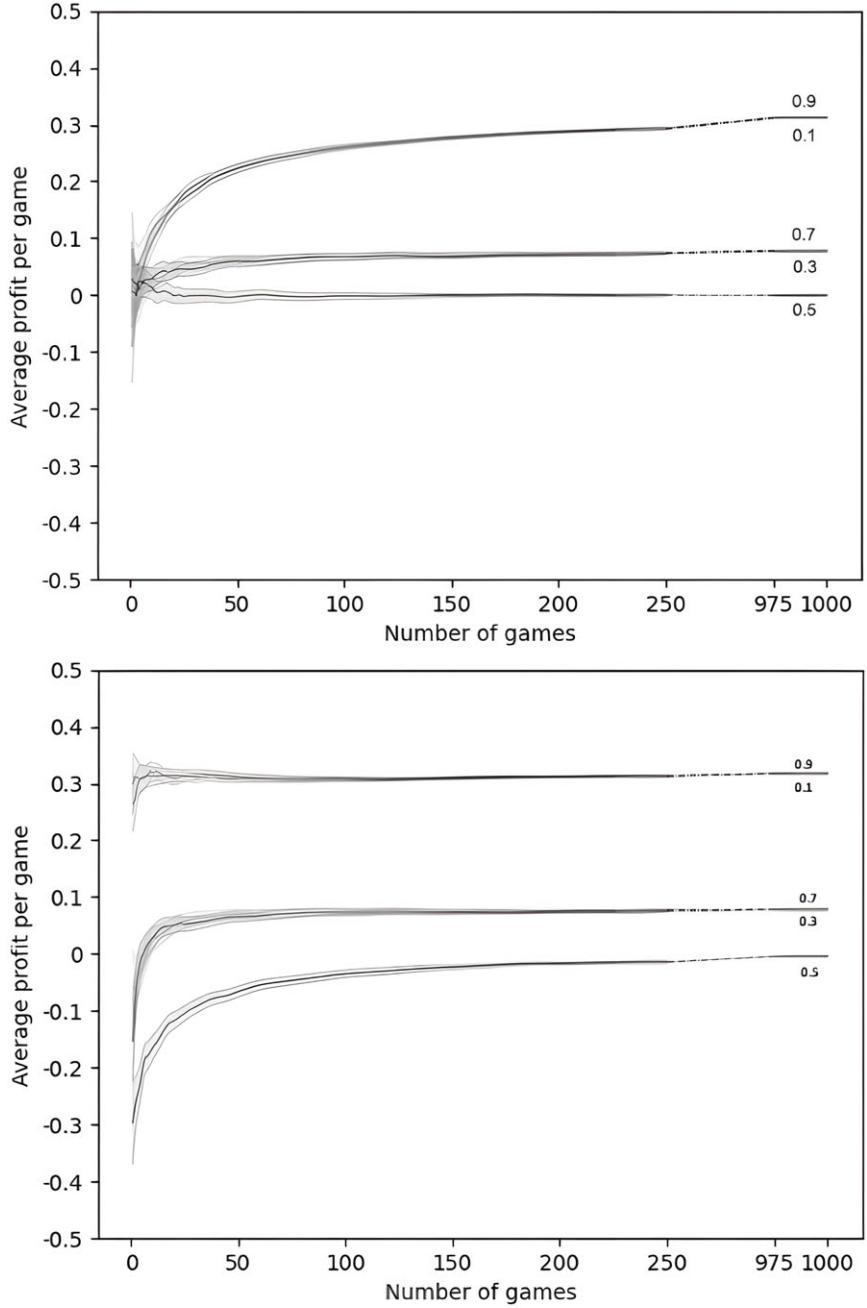


Figure 3. Optimists: AC (top) and MLAC (bottom). The graph format is the same as Figure 2.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

then sums each action's payoff. Given that $\delta = 0.7$, the expected payoff for betting on heads is $0.75(-0.7 + 0.9817) + 0.25(-0.7 + 0.0826) \approx 0.0569$, while the expected payoff from betting on tails is $0.75(0.7 - 0.0826) + 0.25(0.7 - 0.9817) \approx 0.3926$. Consequently, while *Stan* and a player who knew the true bias would bet on heads, *IB-Optimist* would bet on tails.

AC-Optimist differs from *IB-Optimist* in that the minimum credence for heads is $\alpha = 0.75$ multiplied by their maximum credence for heads. This product is approximately 0.7363, and likewise for tails to obtain approximately 0.6881. The approximate expected payoffs are 0.2204 for heads and 0.5601 for tails. Therefore, despite their differences, *AC-Optimist*'s behaviour in this particular game is the same as *IB-Optimist*'s, with the same relative underperformance. In general, both players perform as well as *Stan* given a non-extreme bias, but do worse given an extreme bias.

Meanwhile, *MLAC-Optimist* has an interesting inversion of *AC-Optimist*'s performance. It closes the gap with *Stan* for extreme biases, but performs poorly as the coin bias approaches 0.5. This occurs because *MLAC-Optimist*'s update rule makes them shift quickly towards high confidence in the most frequently observed value. Given an extreme bias, this behaviour tends to be advantageous, since their belief intervals strongly tend to include the true bias. However, if the coin bias is 0.7 or 0.3, *MLAC-Optimist* tends to be overconfident in early samples. For instance, in the preceding example, *MLAC-Optimist* would use the heads interval, approximately $[0.7363, 0.9817]$, leading to overconfidence if the true bias is 0.7 or lower. Given a 0.5 bias, this overconfidence leads to particularly strong misalignment of belief intervals and coin biases, because *MLAC-Optimist*'s learning rule is effectively slanted towards extreme coin toss biases. Moreover, randomizing between the heads interval and the tails interval when the sample frequency is 0.5, rather than using some form of equivocal belief state, results in costly bold decisions until the observations are sufficiently numerous that *MLAC-Optimist*'s intervals are both close to 0.5. In sum, *MLAC-Optimist* does not eliminate the Ambiguity Dilemma for Alpha Cut: their overall performance is still below that of *Stan*.

Returning to Alpha Cut, an Alpha Cut player must consider multiple events in the Bernoulli task in order to make decisions. In a betting problem where they were applying Alpha Cut to their belief in just one event, they would learn faster. However, the Bernoulli task is already a simple problem; simplifying it would limit the interest of the performance results. Moreover, it seems that this dramatic dilution of their rule's convergence properties as the number of events relevant to a decision increases would also be present in decision problems with many relevant events, such as typical choice situations in business or policy. An implication is that the potentially dramatic effects of Alpha Cut on the learning rate in single event-based reasoning problems that have generally been studied so far can be reduced when an Alpha Cut agent considers a wider range of events.

Of course, with more convergent initial credal sets (for instance, a credal set ranging from $Beta(2, 1)$ to $Beta(1, 2)$) *AC-Optimist* would perform better in the Bernoulli task and analogous problems. Yet this performance improvement would be at the cost of a less adequate representation of ambiguity. The initial credal sets we considered already go beyond players' background knowledge of the physical probabilities. This is the essence of the Ambiguity Dilemma: better ambiguity

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

representation comes at a performance cost, better performance comes at an ambiguity representation cost.

To sum up, Alpha Cut does not escape the Ambiguity Dilemma, because it has very similar results to Imprecise Bayesianism in the Bernoulli task. *MLAC-Optimist* closes the gap with *Stan* given extreme biases, but with offsetting costs of poor performance with 0.3, 0.5 and 0.7 coin biases. Thus, the performance side of the Ambiguity Dilemma is not improved by these approaches. In the next section, we shall turn to the ambiguity representation side of the Ambiguity Dilemma.

4.3 Representation

Imprecise Bayesians can use the divergence of credal sets to represent ambiguity, at least in many situations (Walley 1991). Inertia limits the scope of this representation, because a vacuous credal set (in the context of the Bernoulli trials, a set that contains all beta distributions) is a maximally divergent credal set with respect to its probabilities for each event, yet using it to represent ambiguity prevents learning, apart from any learning that is required for deductive consistency. For instance, in the Bernoulli task problem, an Imprecise Bayesian player with a maximally divergent initial credal set (arguably what is the appropriate representation of the ambiguity in the Bernoulli task, given the absence of background knowledge in favour of any particular coin bias) is just as uncertain about the coin bias after observing 100 tosses, all landing heads, as they were before observing any tosses – a belief interval of $[0, 1]$ in each case. In contrast, Alpha Cut updaters can make free use of vacuous credal sets while still learning, as Seamus Bradley has illustrated in detail (Bradley 2022).

One advantage that Alpha Cut advocates themselves have not yet noted is that, unlike using a highly divergent but non-inert credal set to represent maximal ambiguity (we shall call this a "wide set" for short) the use of Alpha Cut to avoid inertia avoids the problem that even a wide set will have strong probabilities for or against some hypotheses, despite the absence of evidence. For example, unless probability distributions assigning 0 and probability distributions assigning 1 are included in the credal set (so the set is vacuous rather than merely wide) an imprecise probability agent in the Bernoulli task will be *certain* that "All the coin tosses land heads" is false over an infinite domain of coin tosses (Zabell 1996). Even if that bullet might be bitten, there is a following bullet that is tougher: the imprecise probability agent will be *certain*, a priori, that at least one coin toss will land heads in such an infinite sequence. Even for those who think that there are some a priori justified beliefs or that all beliefs are partly a priori would presumably stop short of regarding this sort of a priori belief as justified. Thus, *only* by using a vacuous initial credal set can an Imprecise Bayesian or Alpha Cut agent's belief intervals reflect the initial ambiguity across all the hypotheses, and of these two theories only Alpha Cut agents can learn in the Bernoulli task given vacuous credal sets.

Nonetheless, despite these strengths, we shall argue that Alpha Cut introduces a new assortment of problems for representing ambiguity via divergent credal sets. Given these problems, it is debatable whether Alpha Cut actually has net advantages for ambiguity representation.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

4.3.1 Local inertia

The points we make in the remainder of this section are independent of the Optimist decision rule. To emphasize this independence, we shall discuss a generic Alpha Cut player, *AC*, just as we earlier referred to a generic Imprecise Bayesian player as *IB*.

Imagine a version of the Bernoulli task where, unlike in our tests, the coin bias is 1. Thus, although players do not know it, they will only observe samples that uniformly consist of heads tosses. Suppose that *AC* has a vacuous credal set at the start of a test, so that their belief states are maximally ambiguous. Suppose they observe 200 tosses, all of which land heads. Intuitively, these observations are very strong evidence that the coin is very biased towards heads. For example, *Stan* will have a credence of approximately 0.9951 that the next coin will be heads given such observations. Part of Alpha Cut's *raison d'être* is to enable learning with imprecise probabilities in such situations. After their first observation of heads, *AC*'s belief interval becomes $[(1)(0.75), 1] = [0.75, 1]$. Given that the coin always lands heads, *AC-Optimist*'s upper probability distribution will always assign a probability of 1 to heads, while their lower probability distribution will always assign $(1)(\alpha) = 0.75$, and so their belief interval will be unchanged after the next 199 (or more) observations.[7]

There are multiple aspects to this problem. From the perspective of formal epistemology, the problem of local inertia is that an Alpha Cut agent fails to update despite apparently relevant evidence. A consequence is that *AC*'s belief intervals fail to track the apparent reduction in ambiguity, aside from the switch to a 0.75 lower limit after the initial observations. From the perspective of decision theory, the problem is that an Alpha Cut agent will be insufficiently sensitive to information that can be reliably used to guide decisions in the Bernoulli task if they become "stuck" in local inertia.

This local inertia problem only occurs with uniformly heads (or tails) samples. Given any heterogeneity, the belief interval converges asymptotically to the sample frequency. However, the problem is not specific to the Bernoulli task. For example, a multinomial decision problem, featuring Dirichlet priors rather than beta priors, would raise similar problems for Alpha Cut.

What if the initial credal set is wide but non-vacuous? This avoids the previously mentioned local inertia problem, albeit at the cost of losing vacuous credal sets' capacities for ambiguity representation. Unfortunately, a less serious but still paradoxical short-run version of the same problem occurs. Suppose that *AC*'s initial credal set is the same as in our tests (the convex set with $Beta(99, 1)$ and $Beta(1, 99)$ as its extrema) so that *AC*'s belief interval is $[0.01, 0.99]$. As in our example with a $[0, 1]$ initial interval, we assume that the coin bias is 1. After observing their initial sample of 4 uniformly heads tosses, the new interval is approximately $[0.75(0.9904), 0.9904] = [0.7428, 0.9904]$. So far, so good: a uniformly heads sample has pushed the interval towards confidence in heads.

Things do not remain good. Even after 293 total heads tosses, the lower limit has risen only very slightly, to approximately 0.7481. The upper limit also rises marginally to

[7]Throughout our discussion of local inertia, we focus here on *AC*'s belief state with respect to heads. Similar anomalous learning behaviour occurs with respect to the bias towards tails.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

approximately 0.9975. After 293 uniformly heads observations, generalized conditioning intervals become narrower than those from cutting, so *AC* uses generalized conditioning intervals, like *IB*. Thus, almost all of the 293 observations are virtually irrelevant to the belief interval. The result is very non-convergent learning. By contrast, *Stan* has posterior for heads of approximately 0.9966 given 293 heads tosses, which is very close to the true coin bias of 1. Intuitively, there is a large difference in the ambiguity concerning heads given 1 coin toss landing heads and given 293 tosses landing heads, but this large difference is not represented by *AC*'s belief intervals. The cause of the problem is that, in the Bernoulli task, *AC* revises their beliefs in early games by changes in the upper limit of their interval for an event, but their upper limit has little further that it can rise, even if they observe quite large uniform samples.

The problem of local inertia does not prove that Alpha Cut is unworkable, but it does show how Alpha Cut introduces new problems for ambiguity representation that are not present in Imprecise Bayesianism. Local inertia is particularly troubling for Alpha Cut when the initial credal set is vacuous. This is ironic, because learning with these sets is supposed to be a strength of Alpha Cut.

4.3.2 The unit sample problem

We call the following problem the "unit sample problem". Although unit samples do not occur in our tests, simple calculations enable them to be explored, so we can see how players change their belief states after observing just one coin toss.

One selling point of imprecise probabilities for ambiguity representation is to avoid strong beliefs based on minimal evidence. Consider *Stan*'s $Beta(1, 1)$ prior. While this prior implies a neutral credence for the outcomes of each of the 5000 individual tosses in a test (a 0.5 credence for heads and a 0.5 credence for tails) it implies credences that are strongly biased with respect to particular relative frequencies of tosses in a test. For instance, *Stan* is almost certain that the overall distribution of heads tosses in a test will be close to 2500/5000, rather than close to 0/5000 or 5000/5000. In contrast, *AC* (like *IB*) is initially very uncertain whether the frequency will be close to 50/5000 (i.e. 1% of 5000) or 4950/5000 (i.e. 99% of 5000) or somewhere in between.

Next, we suggest an intuition: as part of avoiding strong beliefs based on minimal evidence, you cannot conclude that a coin is unfair because you saw a single toss land heads. More generally, you should only reject a probability distribution on the basis of your total evidence if the total evidence is improbable (its prior is less than 0.5) given that probability distribution. *IB* trivially satisfies these intuitions in the Bernoulli task, since they never remove probability distributions from their credal set in this decision problem.[8] In the Bernoulli task, an Imprecise Bayesian will *not* reason as if they are thinking "I have observed this coin land once and it landed heads, so it's not a fair coin".[9]

In contrast, after just one heads observation, *AC*'s belief interval in heads jumps to approximately $[(0.75)(0.9901), 0.9901] = [0.7426, 0.9901]$. The jump is odd, but

[8]Note that *Stan* also does not excise any probability distributions in the Bernoulli task, but instead conditionalizes using their initial beta distribution.

[9]Imprecise Bayesian agents also satisfy the intuition in other decision problems, because (as we have defined them) they only excise a probability distribution from their credal set if their evidence has a probability of 0 given that distribution.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

we assume that supporters of Alpha Cut are already willing to endorse such an initial jump, at least given $\alpha = 0.75$. A deeper problem is *AC* excises probability distributions if the latter's priors for heads are less than 0.75 even if the prior is greater than or equal to 0.5. Hence, *AC* with $\alpha = 0.75$ reasons as if they think "I have observed one coin toss and it landed heads, so this coin is unfair" and even effectively reasons "I have observed one coin toss and it landed heads, so this coin is not just moderately (less than 0.75) biased towards heads."

For example, consider the beta distribution $Beta(70, 30)$. This implies a prior of $70/(70 + 30) = 0.7$ that the first toss in a test will land heads. Yet this beta distribution is excised from *AC*'s credal set if the first toss lands heads and $\alpha = 0.75$. Thus, even though heads is more probable than not according to such probability distributions, they are excised from the credal set by Alpha Cut.

The problem is not rejection of probability distributions as such. Even Bayesians must do that if the prior probability of their observations is zero. So do classical statisticians; we shall later explain how they avoid the unit sample problem. The problem is the rejection of a uniform distribution (and even more counterintuitive excisions) given a unit sample, which seems an inapposite reflection of the minimal information that such a sample provides.

How low does α have to be to avoid the unit sample problem? In the Bernoulli task, uniform distributions are excised from the set given a single observation if $\alpha > 0.505$.

The problem of ambiguity representation under these conditions is not specific to the Bernoulli task, nor to high values of α. For example, given a die instead of a coin (so that a uniform prior is a credence of 1/6) and Dirichlet priors instead of beta priors, an Alpha Cut player using $\alpha = 0.5$ will reject the uniform distribution based on a unit sample. One can see the generalization of the problem by imagining a lottery of unknown bias or fairness. By increasing the number of lottery tickets in the game, one can always ensure that an Alpha Cut agent will reject the uniform distribution. The result is in deep conflict with how some supporters of Alpha Cut have described its selection: a uniform distribution is not "wrong" (to use Seamus Bradley's phrase for what Alpha Cut eliminates (Bradley 2022: 17)) with respect to a unit sample, yet it is still cut given a wide range of prima facie suitable α values.

A defender of Alpha Cut might argue that a reinterpretation solves this problem. Rather than "excising" probability distributions (such as those assigning a uniform distribution) on the basis of "unreliability" or being "wrong", one is instead "selecting" probability distributions that *perform best in predicting* the relevant outcome. On this interpretation, the Alpha Cut agent's judgement is not that uniform distributions are "more wrong" or "unreliable" given a unit sample, but simply that their performance has not (yet) warranted including them in the updated credal set.

However, this interpretation clouds the relationship between credal sets and beliefs. The epistemological worry is that, according to many imprecise probabilists, an agent should not believe that the coin is biased based on a single observation, given the austere epistemic context of the Bernoulli task. Suppose, for the sake of argument, that on the proposed reinterpretation of Alpha Cut excisions, it is rational for them to disregard uniform distributions for the purpose of making bets in the Bernoulli task. How do the concomitant belief intervals relate to *AC*'s belief

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

state? If they are identical, then the epistemological problem is still present. If they are different, then why is *AC* not betting in accordance with their belief states? We cannot discuss the various responses that an Alpha Cut supporter might have, but at the very least, much more clarification and development is needed to make this response work.

Alternatively, a defender of Alpha Cut could consider modifying α to fit the number of possible outcomes in a particular decision problem. Provided that α is sufficiently low, uniform distributions are not excised based on a unit sample. Unfortunately, setting α in this contextual way would conflict with other desiderata for α, such as adjusting α to an agent's "taste for epistemic risk", including a taste for faster updating than generalized conditioning (Bradley 2022: 24). It would also remove the option of adjusting α based on the stakes, as has been suggested in some related views (Braithwaite 1953; Hill 2013; Bradley 2017; Hill 2019) and it would require slower convergence in some decision problems, including the Bernoulli task.

Neither the unit sample problem, nor the local inertia problem discussed in the previous subsection, occurs for Imprecise Bayesianism. Hence, Alpha Cut's escape from classic inertia comes at a cost in terms of ambiguity representation. While not quite a poisoned chalice, Alpha Cut's offer to an Imprecise Bayesian seeking an inertia-free update rule is a mixed bag.

4.4 Coherence

An Imprecise Bayesian agent's chosen betting odds, given any imprecise probability decision rule, are synchronically coherent. However, they may or may not be diachronically coherent depending on their decision rule. Diachronic coherence conflicts with some other desiderata that Imprecise Bayesians have held, such as allowing for ambiguity aversion (Bradley 2015, 2018; Bradley and Steele 2016; Neth 2023). Nonetheless, generalized conditioning is only diachronically incoherent when combined with some imprecise probability decision rules.

In contrast, Alpha Cut agents are at risk of diachronic incoherence given any of the prominent imprecise probability decision rules. Supporters of Alpha Cut updating are already aware of this fact. Seamus Bradley suggests three possible responses (Bradley 2022: 25) which we shall address one-by-one. Before continuing, we emphasize that throughout what follows in our discussion of Alpha Cut's diachronic properties, the topic is *extensive form* sequences of decisions, where the agent makes a choice at each decision moment in the sequence, rather than *normal form* sequences of decisions, where the agent makes a one-time choice whether to accept a package of decisions conditional on different events. This distinction is important because some aspects of choice behaviour, such as remembering past decisions, will be relevant in what follows, but only apply for extensive form sequences of decisions.

The first response we shall discuss is that if α is sufficiently low in a particular decision problem then an Alpha Cut agent can at least sometimes avoid diachronic incoherence. Bradley points out that there is a positive correlation between α and incoherence in decision problems investigated so far. This functional relationship is also theoretically plausible, because as α tends to 1, Alpha Cut differs more from Imprecise Bayesianism (which can be diachronically coherent) and tends more

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

towards maximum likelihood estimation (which is frequently diachronically incoherent if used to determine credences).

As Bradley notes, it is an open question whether it is always possible to avoid incoherence in a particular decision problem by choosing a sufficiently low value for α. However, the performance results in the Bernoulli task suggest a troubling interaction between keeping α low enough to obtain diachronic coherence and keeping α high enough to avoid exacerbating the Ambiguity Dilemma, since Alpha Cut players learn faster as α is higher. This is also a factor to consider if one explores Bradley's second response, which seems to be that one may be able to find trade-offs between coherence and the advantages of higher α values.

The third response suggested by Bradley requires the most discussion. He raises the possibility of rejecting the idea that, if an agent regards each step in a sequence of choices over time as acceptable, then they must regard the total sequence as acceptable. This sort of move has been explored extensively (Schick 1986; Elga 2010; Bradley 2015; Rinard 2015; Mahtani 2018). The idea is that, if the road to an Alpha Cut agent being diachronically incoherent is paved with attractive choices, just as the road to hell is paved with good intentions, then a rational decision-maker should look beyond each individual paving stone and consider the overall road. Consequently, Bradley proposes rejecting what has been called the "package principle". We shall call the position he favours "sequence holism".

We cannot cover all the intricacies in the debates on the package principle versus sequence holism. However, we shall note an issue that has not yet been addressed by sequence holists. It was first raised, in a different context, by Peter J. Hammond (Hammond 1988). Suppose that, at each decision moment in the sequence of some particular decision problem, an agent cares about their past choices, in order to construct sequences that they consider satisfactory. Sequence holists say that the agent should use some special decision rule to incorporate the past choices into their present decision-making. Yet the agent could also take past choices into account by having preferences about the *combinations* of past choices and present choices. If they do the latter, then Hammond has further arguments leading towards Standard Bayesianism. Perhaps these steps can be resisted, but any sequence holist should specify exactly which steps in Hammond's reasoning they reject.

A further problem for sequence holism concerns assumptions about memory. First, note that the diachronic coherence of Standard Bayesianism does not presuppose that they can recall their past decisions. Provided that a Standard Bayesian's credences stay the same except for any alterations required by conditionalization, they cannot be subject to a diachronic Dutch Book, even if they have forgotten their past decisions.

In contrast, if one is applying a sequence holist decision rule, then one must remember one's past decisions. Otherwise, the agent does not necessarily know how their present decision relates to their overall decisions in a sequence. Hence, in comparison to a Standard Bayesian agent, a sequence holist must make stronger assumptions about the reliability of an agent's memory in order to be sure of diachronic coherence. In the case of long-term decisions or those where the agent's memory is weakened (such as by alcohol consumption or jet lag) these assumptions may be highly unrealistic.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

A defender of sequence holist decision rules might think assuming recall of past choices is just one of many unrealistic assumptions that we often make in decision theory, such as stable preferences, perfect calculation abilities, and so on. However, our point is not that this assumption proves that sequence holist decision rules are fundamentally and universally flawed. Instead, it is just to show that diachronic coherence via these rules requires stronger assumptions than maximizing expected payoffs using Standard Bayesianism. These assumptions must be taken into account when evaluating whether sequence holism is an adequate defence of Alpha Cut against the charge of diachronic incoherence.

A more fundamental issue for decision making when using Alpha Cut concerns synchronic coherence. Both Imprecise Bayesianism (given some decision rules) and Standard Bayesianism are insensitive to how a decision problem is presented. For example, in the Bernoulli task, *IB*'s credal set when they are deciding whether to bet on heads (c_h) is identical to their credal set when they are deciding whether to bet on tails (c_t). The same is true of *Stan*'s credences.

In contrast, an Alpha Cut player's reasoning violates act-state independence – the assumption that whether states of the world occur is (stochastically) independent of the decision-maker's actions. Act-state independence can sometimes function as a mere modelling technique to increase tractability. However, in the Bernoulli task, it is a prerequisite of an accurate conception of the problem, because the coin toss results are entirely independent of players' choices. An agent who thinks that states (rather than just payoffs) in the tests can vary with their choices has a fundamental misconception about what they are doing and what is happening. In the decision problem as we have formulated it, this delusion does not harm *AC-Optimist*, but it would be possible to reformulate the decision problem such that *AC-Optimist* (or any Alpha Cut player) would either have to stop using Alpha Cut updating or be synchronically incoherent.

To illustrate how the decision problem's assumption of act-state independence is violated by the beliefs of Alpha Cut players, we provide an example where *AC-Optimist*'s reasoning in a game is contrasted with *Stan*'s reasoning in an equivalent game, before summarizing the key feature of Alpha Cut that causes the violation. In any test, *Stan*'s initial beta distribution is $Beta(1, 1)$ for coins landing heads and $Beta(1, 1)$ for coins landing tails. Suppose that, at the beginning of the test, *AC-Optimist*'s initial credal set is the convex set with extrema of $Beta(99, 1)$ and $Beta(1, 99)$, as in our tests. Meanwhile, after 4 observations, with 3 tosses landing heads and 1 toss landing tails, *AC-Optimist* has pre-cut belief intervals of approximately $[0.0385, 0.9808]$ for heads and approximately $[0.0192, 0.9615]$ for tails. *Stan* now has $Beta(4, 2)$, implying posteriors of approximately 0.6667 for heads and approximately 0.3333 for tails.

We now slightly alter the Bernoulli task. We imagine that players consider betting on heads and tails separately (but simultaneously) with the added option to put separate bets on both in the same game. Players choose to abstain if and only if they reject both offers.

Next, we stipulate the price in this example: 0.6. *Stan*'s expected payoffs in both the decision whether to bet on heads and the decision whether to bet on tails are approximately 0.0667 for heads and -0.0667 for tails. Hence, *Stan* chooses to bet on

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

heads in the first part of their decision and to not bet on tails in the second part of their decision.

Using Alpha Cut with $\alpha = 0.75$, the belief intervals are approximately $[0.7356, 0.9808]$ for heads and approximately $[0.7211, 0.9615]$ for tails. *AC-Optimist* regards the expectations both for betting heads and for betting tails as positive. Given that $\delta = 0.6$, their expected payoffs are $0.75(0.9808(0.4) + 0.0192(-0.6)) + 0.25(0.7356(0.4) + 0.2644(-0.6)) = 0.3195$ for betting on heads and $0.75(0.0385(-0.4) + 0.9615(0.6)) + 0.25(0.2789(-0.4) + 0.7211(0.6)) = 0.5014$ for betting on tails. Therefore, *AC-Optimist* buys both the heads ticket and the tails ticket.

They have effectively paid $0.6 + 0.6 = 1.2$ as their overall stake. While they have secured a payoff of 1, this prize will be less than their stake of 1.2, and concomitantly they are guaranteed a sure loss of 0.2. Thus, *AC-Optimist*'s choice behaviour in the modified Bernoulli task is synchronically incoherent.

The *Optimist* rule plays no essential role in this example.[10] Instead, the key feature in *AC-Optimist* that leads to this behaviour is the violation of act-state independence: when *AC-Optimist* is thinking about betting on heads, the Alpha Cut rule raises their expected payoffs relative to an equivalent *IB* player. This boost is proportional to α. However, this boost is also applied when *AC-Optimist* considers whether to bet on tails. To avoid such sequences of decisions resulting from their decision rule, an Alpha Cut player must recall past decisions, as we discussed earlier.

Therefore, coherence must be regarded as a tangled nest of problems for Alpha Cut, even if no thorn is fatal in itself for Alpha Cut's appeal. Diachronic incoherence problems are avoidable, but only insofar as it is reasonable to assume that past decisions in a sequence can be remembered by the Alpha Cut agent.

Suppose that a supporter of Alpha Cut is willing to accept either diachronic incoherence or sequence holism. In the next section, we shall detail how there is already an update rule that (1) uses imprecise probabilities, (2) leads to a similar situation with respect to diachronic coherence, and (3) avoids many of the problems of Alpha Cut, without introducing novel problems, at least in the context of the Bernoulli task.

5. Calibration

In this section, we shall compare Calibration with Alpha Cut. The key figures in the Calibration literature are Henry E. Kyburg and Jon Williamson. Kyburg developed an imprecise probability theory, Evidential Probability, based on rules for transforming information about physical probabilities into measures of evidential support (Kyburg and Teng 2001). He thought that these measures of evidential support should determine interval-valued imprecise credences (Kyburg 2006). In contrast, Williamson uses these intervals alongside the Maximum Entropy Principle to constrain the choice of a probability distribution for credences (Wheeler and Williamson 2011; Williamson 2013). Williamson's theory is one version of Objective Bayesianism.

[10]Hurwicz criterion rules are diachronically incoherent, but the example is one of synchronic incoherence.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

5.1 Definition

By "Calibration" (the term is Williamson's) we shall mean the use of Kyburg's rules for determining interval-valued probabilities, which are employed by both Kyburg and Williamson in their wider formal epistemologies. The applications of Calibration that we shall discuss are fairly simple and do not require the full formal machinery that Kyburg and Williamson have developed. Thus, in this subsection, we shall only outline the general philosophical details. We provide a full formal specification of Calibration, in the context of the Bernoulli task, in Appendix 6.7.

Suppose that our only evidence regarding whether some individual thing is a G is that it is an F and that r% of F are G. A quite common claim by epistemologists, including Kyburg, is that our evidential support for the hypothesis that it is a G is at least sometimes measurable by r%, *ceteris paribus*. Inferences from (a) statistics regarding a population and the premise that an individual (or set of individuals) is a member of that population to (b) hypotheses about that individual have variously been called "direct inference", "the statistical syllogism" or "the proportional syllogism" (Carnap 1962; Seidenfeld 1979; Levi 1980; Stove 1986; Franklin 2001). Since our knowledge of the relevant statistic is often imprecise, r can also take values such as intervals, both when our statistical information is interval-valued (e.g. that the proportion is somewhere in $[0.7, 0.9]$) and when an interval is an approximation of a qualitative natural language expression (e.g. $[0.99, 1]$ for "almost all"). For example, if all you know about the colour of a particular swan's plumage is that about $[0.6, 0.7]$ of swans are white, then that is how much your evidence supports the swan being white.

Of course, our statistical evidence often provides conflicting information. If you also knew that $[0.999, 1]$ Australian swans are non-white (black or black-necked swans) and that the bird is an Australian swan, then it is intuitively rational to use the statistics about Australian swans instead of swans in general. This is an example of a "defeater" statistic.

Similarly, given statistics about the *proportion* of red balls in an urn and the proportion of *selections* of red balls from that urn, it makes sense to dismiss the former information if neither statistic for these proportions is a subinterval of the other. In general, in Kyburg's system, information about a joint physical probability distribution takes precedence over marginal information when they conflict.

Finally, suppose we have exhausted the extent to which these criteria (which Kyburg makes formally precise and mechanical) can exclude statistical information. Kyburg's final rule generates an interval-valued probability, which, roughly speaking, is the narrowest cover of those intervals that (a) are not subintervals of each other and (b) are each proper subintervals of all other surviving intervals. For example, if the surviving intervals are $[0, 1]$, $[0.1, 0.9]$, $[0.3, 0.5]$ and $[0.6, 0.7]$, then the Calibration interval is $[0.3, 0.7]$.

Calibration updating resembles Bayesian statistical reasoning when the background information about physical probabilities is highly informative, but classical statistical reasoning when the background information about physical probabilities is meagre (Kyburg and Teng 2001). In the austere epistemic context of the Bernoulli task, Calibration involves the use of confidence intervals to generate

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Calibration intervals and thus uses a tool from classical statistics. However, unlike ordinary classical statistics, these intervals are interpreted as imprecise single-case probabilities for coins landing heads and for coins landing tails. Williamson makes the further step of using such intervals as part of determining Bayesian credences. We shall provide further details of the specific implementation of Calibration in the Bernoulli task after briefly discussing the general coherence properties of Calibration updating.[11]

5.2 Coherence

Agents updating via Calibration can be synchronically coherent (Wheeler and Williamson 2011). Firstly, they can use certain imprecise probability decision rules with this property (Kyburg 1990: Ch. 14). Secondly, they can use what Williamson has called "empirically based subjectivism" where an agent chooses some arbitrary credence function whose values fall within the agent's Calibration intervals for each event in the function's domain (Williamson 2007). Thirdly, they can use Williamson's version of Objective Bayesianism, where the choice of credence function is almost entirely constrained by the Maximum Entropy Principle (Williamson 2010). Any of these approaches to decision-making will result in synchronic coherence.

In contrast, Calibration updating is diachronically incoherent, unless some special moves such as sequence holism are made. The reasons are the same as for Alpha Cut: any update rule that is not equivalent to conditionalization (as in Standard Bayesianism) or generalized conditioning using some decision rules (as in Imprecise Bayesianism) is vulnerable to a diachronic Dutch Book (Pettigrew 2020). While Calibration corresponds to conditionalization or generalized conditioning in some special cases, there are also infinitely many possible situations where they diverge.

However, Calibration fundamentally does no worse than Alpha Cut with respect to the desideratum of diachronic coherence. Moreover, unlike Alpha Cut, a Calibration agent does not require recall of previous decisions to avoid synchronic Dutch Books, although sequence holism is also apparently an option for Calibration. Overall, Calibration seems somewhat better than Alpha Cut with respect to synchronic and diachronic coherence, but more investigation is needed to

[11]An anonymous referee raises the point that Calibration derives imprecise probabilities from beliefs about statistical probabilities and hence only provides non-vacuous intervals (intervals narrower than $[0, 1]$) if agents have non-vacuous beliefs about statistical parameters. This restriction on informative epistemic probabilities could be seen as a disadvantage relative to Standard Bayesianism or Imprecise Bayesianism, if one thinks that there are cases where non-vacuous belief intervals or credences are justified yet no non-vacuous statistical probabilities can be rationally accepted. Alternatively, some epistemologists, such as evidentialists, might see this restriction as an advantage, because it limits precision in belief states to cases of precision in one's evidence (Peden 2024). Since inertia is not a feature of Calibration, vacuous intervals in situations of extremely exiguous evidence are not fatal to subsequent learning (Peden 2024). The epistemological debates involved are important to the overall evaluation of Calibration, but beyond the scope of this article, where we confine our attention to theories' ambiguity representation capacities, coherence properties, and performance in the Bernoulli task.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

determine whether Calibration is compatible with sequence holism and similar imprecise probability strategies for avoiding diachronic incoherence.

5.3 Performance

Calibration's performance on the Bernoulli task has been previously studied (Radzvilas *et al.* 2023). In this subsection, we just compare the summary results with those for Imprecise Bayesianism, then for Standard Bayesianism, and finally for Alpha Cut.

Calibration players in the Bernoulli task have a parameter γ, which is their significance level for choosing which coin bias estimates to reject and which to tentatively accept. In the Bernoulli task, they perform best with a very high significance level of 0.5. While Calibration supporters would almost never use such a high significance level in science, it is consistent with their theory to use it for a decision problem like the Bernoulli task. Thus, we shall discuss only *Calibration* players' results for $\gamma = 0.5$ and all of our discussions below assume this significance level.

Starting with a maximally wide estimate of the coin bias (a vacuous confidence interval, ranging from 0 to 1, apparently representing the strong ambiguity at the start of the Bernoulli task) a *Calibration-Optimist* player and players with equivalent choices can match *Stan*'s performance. Interestingly, no other theory has yet produced this result. The reason is that, within the context of the Bernoulli task, *Calibration-Optimist* behaves very similarly to *Stan*, with a high significance level having similar effects to a flat prior and the Optimist decision rule closing much of the remaining gap, leaving only statistically insignificant (estimated at a 0.05 significance level) differences in their choice behaviour. Other Calibration players do not do as well, except some that make identical choices to *Calibration-Optimist* (Radzvilas *et al.* 2023).

Calibration players perform better in the Bernoulli task than an *IB* player whenever the decision rule is held constant. For example, *Calibration-Optimist* performs better than the *IB-Optimist* player (Radzvilas *et al.* 2023). Thus, *Calibration* offers a wider set of tools for representing ambiguity (since it does not feature inertia) yet it also escapes the performance dimension of the Ambiguity Dilemma.

Comparing *Calibration-Optimist* and *AC-Optimist*, only the former can match *Stan* in the Bernoulli task. (The results for *Calibration-Optimist* are approximately identical to the results for *Stan* in Figure 2.) In the case of *MLAC-Optimist*, this player closes the gap with *Calibration-Optimist* given an extreme coin bias, but performs worse with 0.3, 0.7 or 0.5 biases. Perhaps there are some decision problems where Alpha Cut does better than Calibration in terms of short-run performance, but none have yet been discovered.

One important distinction between (1) the performance results for Imprecise Bayesian players and Alpha Cut players explored so far and (2) the performance results for *Calibration* players is that the latter begin with maximally imprecise intervals of $[0, 1]$, whereas the former begin with non-vacuous credal sets. Since no player initially has even approximate background knowledge of the coin biases, it seems that the $[0, 1]$ interval is the more fitting representation of the ambiguity in

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

players' initial evidence. At least using the Optimist decision rule and $\gamma = 0.5$, the Ambiguity Dilemma is resolved entirely: *Calibration-Optimist* performs approximately as well as any other player in the Bernoulli task, even when their initial interval is the completely vacuous $[0, 1]$.[12] Thus, Calibration offers imprecise probabilists a stronger theory than Alpha Cut with respect to performance, even when starting from an apparently more faithful representation of the ambiguity in the Bernoulli task.

5.4 Representation

Like Alpha Cut, a vacuous initial interval is insufficient for inertia in Calibration (Kyburg and Teng 2001; Peden 2024). However, they differ with respect to the local inertia problem. Suppose that players only observe uniform samples (all heads or all tails) in the entirety of a test or the first part of a test. Given a vacuous initial credal set, *AC* (an Alpha Cut agent with $\alpha = 0.75$ and the same initial credal set as the Imprecise Bayesian players in the Bernoulli task) becomes stuck with a $[0.75, 1]$ interval toward heads. Given a non-vacuous credal set, such as that used in our Bernoulli task tests, *AC* becomes highly insensitive to evidence in the early part of the test.

What happens with *Calibration* players given uniform samples? In brief, they have *faster* learning rather than slower learning when samples are uniform.

For instance, given 292 heads results out of 292 tosses, a *Calibration* player with $\gamma = 0.5$ has an interval of approximately $[0.9953, 1]$, which is very similar to *Stan*'s posterior of approximately 0.9966. This confidence interval is actually narrower than the confidence interval given a sample report of 146/292 tosses landing heads (a 50-50 sample) which is approximately $[0.4786, 0.5214]$. By comparison, *AC*'s intervals at $\alpha = 0.75$ for heads are approximately $[0.7481, 0.9975]$ given 292/292 heads tosses and $[0.4688, 0.625]$ given 146/292 heads tosses. In general, there is no local inertia problem for *Calibration* players, because given the background information of the Bernoulli task, confidence interval estimation revises quickly with uniform samples. This faster convergence is due to the lower error probabilities when samples are uniform.[13] Therefore, Calibration intervals seem to represent the ambiguity given large and uniform samples more intuitively than Alpha Cut intervals.

The unit sample problem also does not arise for Calibration players. The problem for Alpha Cut was that if α exceeded the probability of an event given a uniform distribution, then a unit sample would be sufficient to excise a uniform distribution from the Alpha Cut agent's belief interval. Perhaps worse, even if the event's prior according to other probability distributions was over 0.5, these distributions could still be excised from the belief interval.

[12]That is not to say that there are no reasons why someone might object to *Calibration-Optimist*. For instance, the Optimist rule is diachronically incoherent.

[13]Error probabilities are the risks of a mistake in an inference, given the assumptions of a test. In the context of the Bernoulli task, where the inference is a confidence interval estimation, the error probabilities for *Calibration* players are given by the significance level γ.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

In contrast, Calibration avoids the unit sample problem for fundamental reasons. In the Bernoulli task and similar problems, a Calibration agent only excises a physical probability distribution from their interval-valued estimate if the likelihood of their observations given that distribution is less than $1 - \gamma$. Even with the highest possible γ value in the Bernoulli task, 0.5, a uniform distribution is not rejected given a unit sample, since the likelihood of a unit sample's toss outcome will always be 0.5 given a uniform distribution.[14] In general, if a physical probability distribution is outside the confidence interval, then the data must have a likelihood of less than 0.5 given that distribution. A further consequence is that there is no analogue to Alpha Cut's potential excision of probability distributions that assign priors to the evidence that are greater than 0.5 but less than α.

To illustrate, the confidence interval given solely the observation of a single heads toss is $[0.25, 1]$ if $\gamma = 0.5$. For more familiar significance levels, the confidence intervals are $[0.025, 1]$ if $\gamma = 0.05$ and $[0.005, 1]$ if $\gamma = 0.01$. Which of these is the best representation of ambiguity is debatable and perhaps there is no objective answer – the point is merely that the unit sample problem does not occur.

What about if the problem is not binomial? Here, Calibration players using confidence interval estimation can employ the Bonferroni correction developed by Olive Jean Dunn (Dunn 1961). (There are other methods, but we shall only discuss this approach.) Assume a problem identical to the Bernoulli task, except with q decision-relevant states. For instance, in tossing a six-sided die instead of a coin, $q = 6$ instead of $q = 2$. The Bonferroni correction effectively requires dividing γ by q to determine the significance level. This procedure means that, no matter q's value, a uniform distribution that assigns $1/q$ to each possible outcome will not be rejected by a unit sample. *A fortiori*, distributions assigning higher likelihoods than the uniform distribution's likelihood are also not rejected. To illustrate, given only the observation of a six-sided die roll landing on a 3, the confidence interval for the proportion of 3s among the die tosses, given an original significance level of $\gamma = 0.5$, adjusted by the Bonferroni correction, is approximately $[0.0833, 1]$, which is consistent with the uniform distribution probability of $1/6 \approx 0.1667$.

Thus, the unit sample problem does not occur for Calibration. A consequence of the unit sample problem is that, while Alpha Cut must adjust α in a particular decision problem to avoid the unit sample problem, Calibration only has to adjust γ by a technique like the Bonferroni correction, which is already part of the classical statistics methodology that they use in problems like the Bernoulli task.

Considering this section on Calibration as a whole, *if* one is willing to pay Alpha Cut's price in diachronic decision-making, then Alpha Cut seems to have no advantages over Calibration. Alternatively, if one wants to stick to Standard Bayesianism or Imprecise Bayesianism because of diachronic coherence or some other desideratum, then one can consistently reject both Alpha Cut and Calibration. Our only proviso on the latter point is that Calibration performs better than Imprecise Bayesianism with respect to the Ambiguity Dilemma and also with respect to inertia. Yet there are also differences between Calibration and Imprecise Bayesianism that arguably count in favour of the latter's representation of ambiguity

[14]If $\gamma > 0.5$, then Calibration reasoning becomes inconsistent in the Bernoulli task, because both an estimate and a contrary estimate can pass the significance level.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

(Seidenfeld 1978, 2007) and Imprecise Bayesianism's diachronic coherence properties seem more promising than those of Calibration (Peden 2024). Consequently, we evaluate Calibration favourably only in comparison to Alpha Cut, and only regarding the criteria of ambiguity, coherence, and performance, as conceived in this article.

6. Conclusion

Our examination of Alpha Cut has found that, while it has advantages over traditional Imprecise Bayesianism in the special case of inertia, it also brings a significant number of new paradoxes.[15] Additionally, anyone willing to bite the bullet on these paradoxes for Alpha Cut can find a more robust alternative in the Calibration family of formal epistemologies.

Of the theories that we have considered, Calibration seems to be in the strongest position with respect to ambiguity representation. It avoids inertia even with vacuous initial belief intervals. Uniquely among non-Standard Bayesian theories investigated so far, Calibration can also avoid the Ambiguity Dilemma.

Calibration does less well than Imprecise Bayesianism or Standard Bayesianism with respect to coherence. While Calibration agents can be synchronically coherent, Calibration is diachronically incoherent, barring some unexplored approach such as combining it with sequence holism. Furthermore, we have explained how sequence holism makes stronger assumptions about an agent's faculty of memory than some alternatives, such as maximizing expected payoffs using Standard Bayesianism.

Surveying the literature on the Bernoulli task and related issues as a whole, it is interesting that no single theory can improve on Standard Bayesianism with respect to ambiguity representation without also incurring costs in coherence, performance or both. Standard Bayesianism, with suitable priors, can perform extremely well in the Bernoulli task. Its coherence strengths are well-known. Insofar as it has a weakness on the criteria we discuss, it is with respect to ambiguity representation, at least in the eyes of many formal epistemologists and decision theorists.

A pessimistic diagnosis of our discussions and results would be to conjecture a fundamental tension between two or more of the desiderata of ambiguity representation, coherent decision-making and performance in the Bernoulli task.[16] While we have no proof of such an inescapable problem, we have no disproof either. It is not clear how any of the theories we discuss can satisfy all three desiderata without fundamental changes. However, as soon as one considers what such fundamental changes might look like, one steps into frontier debates in formal epistemology and decision theory, such as the requirements of rationality in a sequence of decisions (Bradley and Steele 2014a; Pettigrew 2020; Neth 2023) and the proper relationship between evidence and belief (Sturgeon 2008; Vallinder 2018; Peden 2024). These formidable conceptual challenges suggest that, if there is a fundamental tension, formal epistemology and decision theory are not yet close to being able to prove its existence.

[15]Comparable paradoxes have been documented in other areas of imprecise probability theory (Gong and Meng 2021).

[16]We thank an anonymous referee for suggesting a version of this explanation.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Additionally, there are viable explanations of our results that are more optimistic. It is possible that some form of hybridization or pluralism is necessary to develop a consistently better theory. For example, at least based on the issues we have discussed, Calibration seems to have very strong properties for ambiguity representation, while Standard Bayesianism has the most attractive overall properties for coherence. This contrast suggests the possibility of combining these very different theories.

For instance, Calibrationist might use Standard Bayesian reasoning as a proxy for decision-making. Essentially, the Calibrationist agent would construct a Standard Bayesian agent (presumably one with beliefs and preferences that are similar to the Calibrationist agent's) and act "as if" they were this Standard Bayesian agent. Even the Ellsberg Paradox preferences might be accommodatable into such a theory, if Richard Bradley is correct about their compatibility with Standard Bayesianism (Bradley 2016) which is significant given the Ellsberg Paradox's influence on many imprecise probabilists.

Alternatively, Standard Bayesians might use Calibration to represent evidence (perhaps conceived in terms of reliable error probabilities) and related concepts like ambiguity, but view these as not fully determining an agent's credences. These Calibration intervals may be used to guide, though not determine, the Standard Bayesian's choice of prior. For instance, *Stan*'s posteriors are often close to *Calibration* players' intervals, especially intervals estimated using high significance levels.

These two alternative hybridizations are similar in most respects, with the principal difference being what is regarded as the belief states – the Calibration intervals or the Standard Bayesian probabilities. Certain common elements, such as similarities between the Calibration intervals and the Standard Bayesian credences, may be quantifiable. In general, it seems possible in principle to combine Calibration and Standard Bayesianism in various ways, but no proposal has been investigated in any detail yet. There would be important questions to answer, such as how ambiguity representation and Standard Bayesian credences (under their normal interpretation or as proxy credences for a Calibration agent) could diverge in a rational agent.

Therefore, although we have focused on problems for Alpha Cut, we do not think that pessimism is warranted. Particular aspects of the problems we discuss seem to be surmountable by particular theories; the persistent challenge has only been developing a *unified* theory that does everything that various economists, philosophers and other theorists have wanted. Moreover, our results do not directly imply problems for some similar theories to Alpha Cut, such as those explored by Brian Hill and Richard Bradley (Hill 2013, 2019; Bradley 2017). Many possibilities are unexplored in decision theory and formal epistemology. These subjects thus seem to still be on the road to developing a theory with consistently attractive properties with respect to ambiguity, coherence and performance.

Acknowledgements. The work described in this article was supported by German Research Foundation project SP 279/21-1 (project no. 420094936), and the University of Modena and Reggio Emilia, Italy. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this article. We are also grateful to Thomas Augustin, Michalis Christou, Gert de Cooman, Robert Frühstückl,

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Jakob Gschwandtner, Alexander Linsbichler, Julian Reiss and Jónatan Sólon Magnússon for discussions about this article.

References

Ahn D., S. Choi, D. Gale and S. Kariv 2014. Estimating ambiguity aversion in a portfolio choice experiment: Estimating ambiguity aversion. *Quantitative Economics* **5**(2), 195–223.

Al-Najjar N. I. and J. Weinstein 2009. The ambiguity aversion literature: a critical assessment. *Economics and Philosophy* **25**(3), 249–284.

Bacchus F., H. E. Kyburg and M. Thalos 1990. Against conditionalization. *Synthese* **85**(3), 475–506.

Benétreau-Dupin Y. 2015. The Bayesian who knew too much. *Synthese* **192**(5), 1527–1542.

Bradley R. 2016. Ellsberg's paradox and the value of chances. *Economics and Philosophy* **32**(2), 231–248.

Bradley R. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.

Bradley R. 2024. Catastrophe insurance decision making when the science is uncertain. *Economics and Philosophy* **41**(1), 161–177.

Bradley S. 2015. How to choose among choice functions. In *ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, ed. T. Augustin, S. Doria, E. Miranda and E. Quaeghebeur, 57–66. Aracne Editrice.

Bradley S. 2018. A counterexample to three imprecise decision theories. *Theoria* **85**(1), 18–30.

Bradley S. 2019. Imprecise probabilities. In *The Stanford Encyclopedia of Philosophy*, Spring 2019 edition, ed. E. N. Zalta. Metaphysics Research Lab, Stanford University.

Bradley S. 2022. Learning by ignoring the most wrong. *KRITERION – Journal of Philosophy* **36**(1), 9–31.

Bradley S. and K. Steele 2014a. Should subjective probabilities be sharp? *Episteme* **11**(3), 277–289.

Bradley S. and K. Steele 2014b. Uncertainty, learning, and the "problem" of dilation. *Erkenntnis* **79**(6), 1287–1303.

Bradley S. and K. Steele 2016. Can free evidence be bad? Value of information for the imprecise probabilist. *Philosophy of Science* **83**(1), 1–28.

Braithwaite R.B. 1953. *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge: Cambridge University Press.

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

Cattaneo M.E.G.V. 2008. Fuzzy probabilities based on the likelihood function. In *Soft Methods for Handling Variability and Imprecision*, ed. D. Dubois, M. A. Lubiano, H. Prade, M. A. Gils, P. Grzegorzewski and O. Hryniewicz, 43–50. Berlin: Springer.

Cattaneo M.E.G.V. 2014. A continuous updating rule for imprecise probabilities. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, ed. A. Laurent, O. Strauss, B. Bouchon-Meunier and R. R. Yager, 426–435. Berlin: Springer.

Cozman F.G. 2012. Sets of probability distributions, independence, and convexity. *Synthese* **186**(2), 577–600.

Davidson B. and R. Pargetter 1987. Guilt beyond reasonable doubt. *Australasian Journal of Philosophy* **65**(2), 182–187.

De Bock J. and G. de Cooman 2015. Conditioning, updating and lower probability zero. *International Journal of Approximate Reasoning* **67**, 1–36.

Dunn O.J. 1961. Multiple comparisons among means. *Journal of the American Statistical Association* **56**(293), 52–64.

Elga A. 2010. Subjective probabilities should be sharp. *Philosopher's Imprint* **10**(5), 1–11.

Ellsberg D. 1961. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* **75**(4), 643–669.

Feduzi A. 2010. On Keynes's conception of the weight of evidence. *Journal of Economic Behavior & Organization* **76**(2), 338–351.

Fox C.R. and A. Tversky 1995. Ambiguity aversion and comparative ignorance. *Quarterly Journal of Economics* **110**(3), 585–603.

Franklin J. 2001. Resurrecting logical probability. *Erkenntnis* **55**(2), 277–305.

Gärdenfors P. and N.-E. Sahlin 1982. Unreliable probabilities, risk taking, and decision making. *Synthese* **53**(3), 361–386.

Gilboa I. and D. Schmeidler 1993. Updating ambiguous beliefs. *Journal of Economic Theory* **59**(1), 33–49.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Gong R. and X.-L. Meng 2021. Judicious judgment meets unsettling updating: dilation, sure loss and Simpson's paradox. *Statistical Science* **36**(2), 169–190.

Goodsell Z. and J.M. Nebel 2024. Symmetry, invariance, and imprecise probability. *Mind* **134**(535), 758–773.

Greaves H. and D. Wallace 2006. Justifying conditionalization: conditionalization maximizes expected epistemic utility. *Mind* **115**(459), 607–632.

Gustafsson J.E. 2022. *Money-Pump Arguments*. Cambridge: Cambridge University Press.

Halevy Y. 2007. Ellsberg revisited: an experimental study. *Econometrica* **75**(2), 503–536.

Hammond P.J. 1988. Consequentialist foundations for expected utility. *Theory and Decision* **25**(1), 25–78.

Hill B. 2013. Confidence and decision. *Games and Economic Behavior* **82**, 675–692.

Hill B. 2019. A non-Bayesian theory of state-dependent utility. *Econometrica* **87**(4), 1341–1366.

Hurwicz L. 1951. The Generalised Bayes-Minimax principle: a criterion for decision-making under uncertainty. *Cowles Commission Discussion Paper: Statistics* **335**, 1–7.

Isaacs Y., A. Hájek and J. Hawthorne 2021. Non-measurability, imprecise credences, and imprecise chances. *Mind* **131**(523), 894–918.

Jia R., E. Furlong, S. Gao, L.R. Santos and I. Levy 2020. Learning about the Ellsberg Paradox reduces, but does not abolish, ambiguity aversion. *PLoS ONE* **15**(3), e0228782.

Joyce J.M. 2005. How probabilities reflect evidence. *Philosophical Perspectives* **19**(1), 153–178.

Kasser J. 2015. Two conceptions of weight of evidence in Peirce's illustrations of the logic of science. *Erkenntnis* **81**(3), 629–648.

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

Kuipers T.A. 1976. Inductive probability and the paradox of ideal evidence. *Philosophica* **17**(1), 197–205.

Kyburg H.E. 1968. Bets and beliefs. *American Philosophical Quarterly* **5**(1), 54–63.

Kyburg H.E. 1974. *The Logical Foundations of Statistical Inference*. Dordrecht: Springer.

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

Kyburg H.E. 2006. Belief, evidence, and conditioning. *Philosophy of Science* **73**(1), 42–65.

Kyburg H.E. and C.M. Teng 1999. Choosing among interpretations of probability. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, ed. K.B. Laskey and H. Prade, 359–365. San Francisco: Morgan Kaufmann.

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

Lassiter D. 2019. Representing credal imprecision: from sets of measures to hierarchical Bayesian models. *Philosophical Studies* **177**(6), 1463–1485.

Levi I. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, MA: MIT Press.

Lyon A. 2017. Vague credence. *Synthese* **194**(10), 3931–3954.

Mahtani A. 2018. Imprecise probabilities and unstable betting behaviour. *Noûs* **52**(1), 69–87.

Milne P. 1991. A dilemma for subjective Bayesians – and how to resolve it. *Philosophical Studies* **62**(3), 307–314.

Moral S. 2019. Learning with imprecise probabilities as model selection and averaging. *International Journal of Approximate Reasoning* **109**, 111–124.

Nance D.A. 2016. *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*. Cambridge: Cambridge University Press.

Neth S. 2023. Rational aversion to information. *British Journal for the Philosophy of Science*. doi: 10.1086/727772.

Norton J.D. 2011. Challenges to Bayesian confirmation theory. In *Philosophy of Statistics*, volume **7** of *Handbook of the Philosophy of Science*, ed. P.S. Bandyopadhyay and M.R. Forster, 391–439. New York: Elsevier.

Norton J.D. 2021. *The Material Theory of Induction*. Calgary: University of Calgary Press.

Omar A. and T. Augustin 2019. Estimation of classification probabilities in small domains accounting for nonresponse relying on imprecise probability. *International Journal of Approximate Reasoning* **115**, 134–143.

Pearl J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann.

Peden W. 2018. Imprecise probability and the measurement of Keynes's "Weight of Arguments.". *Journal of Applied Logics – IFCoLog Journal of Logics and their Applications* **5**(3), 677–708.

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Peden W. 2024. Evidentialism, inertia, and imprecise probability. *British Journal for the Philosophy of Science* **75**(4), 797–819.

Pettigrew R. 2020. What is conditionalization, and why should we do it? *Philosophical Studies* **177**(11), 3427–3463.

Piatti A., M. Zaffalon, F. Trojani and M. Hutter 2009. Limits of learning about a categorical latent variable under prior near-ignorance. *International Journal of Approximate Reasoning* **50**(4), 597–611.

Popper K.R. 1958. A third note on degree of corroboration or confirmation. *The British Journal for the Philosophy of Science* **8**(32), 294–302.

Radzvilas M., W. Peden and F. De Pretis 2021. A battle in the statistics wars: a simulation-based comparison of Bayesian, Frequentist and Williamsonian methodologies. *Synthese* **199**(5–6), 13689–13748.

Radzvilas M., W. Peden and F. De Pretis 2023. Making decisions with evidential probability and objective Bayesian calibration inductive logics. *International Journal of Approximate Reasoning* **162**, 109030.

Radzvilas M., W. Peden and F. De Pretis 2024a. The ambiguity dilemma for imprecise Bayesians. *The British Journal for the Philosophy of Science* https://doi.org/10.1086/729618

Radzvilas M., W. Peden and F. De Pretis 2024b. Ambiguous decisions in Bayesianism and imprecise probability. *BJPS Short Reads*. https://www.thebsps.org/short-reads/bayesianism-radzvilas-et-al/

Radzvilas M., W. Peden, D. Tortoli and F. De Pretis 2024. A comparison of imprecise Bayesianism and Dempster–Shafer theory for automated decisions under ambiguity. *Journal of Logic and Computation* https://doi.org/10.1093/logcom/exae069

Ramsey F.P. 1988. Truth and probability. In *Decision, Probability and Utility*, ed. P. Gärdenfors and N. Sahlin, 19–47. Cambridge: Cambridge University Press.

Reiss J. 2014. What's wrong with our theories of evidence? *Theoria: An International Journal for Theory, History and Foundations of Science* **29**(2), 283–306.

Rinard S. 2013. Against radical credal imprecision. *Thought: A Journal of Philosophy* **2**(2), 157–165.

Rinard S. 2015. A decision theory for imprecise probabilities. *Philosopher's Imprint* **15**(7), 1–16.

Rowbottom D.P. 2007. The insufficiency of the Dutch book argument. *Studia Logica* **87**(1), 65–71.

Runde J. 1990. Keynesian uncertainty and the weight of arguments. *Economics and Philosophy* **6**(2), 275–292.

Schick F. 1986. Dutch bookies and money pumps. *Journal of Philosophy* **83**(2), 112–119.

Seabold S. and J. Perktold 2010. Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference*, ed. S. van der Walt and J. Millman, 92–96.

Seidenfeld T. 1978. Direct inference and inverse inference. *Journal of Philosophy* **75**(12), 709–730.

Seidenfeld T. 1979. *Philosophical Problems of Statistical Inference: Learning from R.A. Fisher*. Theory and Decision Library. Dordrecht: Springer.

Seidenfeld T. 2007. Forbidden fruit: when epistemic probability may not take a bite of the Bayesian apple. In *Probability and Inference. Essays in Honour of Henry E. Kyburg Jr.*, **volume 2** of *Texts in Philosophy*, ed. W. Harper and G. Wheeler, 267–279. London: College Publications.

Stove D.C. 1986. *The Rationality of Induction*. Oxford: Clarendon Press.

Sturgeon S. 2008. Reason and the grain of belief. *Noûs* **42**(1), 139–165.

Troffaes M.C. and G. de Cooman 2014. *Lower Previsions*. Chichester: Wiley.

Vallinder A. 2018. Imprecise Bayesianism and global belief inertia. *The British Journal for the Philosophy of Science* **69**(4), 1205–1230.

Vineberg S. 2022. Dutch book arguments. In *The Stanford Encyclopedia of Philosophy*, Fall 2022 edition, ed. E.N. Zalta and U. Nodelman. Metaphysics Research Lab, Stanford University.

Walley P. 1991. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

Wheeler G. 2021. Comment: moving beyond sets of probabilities. *Statistical Science* **36**(2), 201–204.

Wheeler G. and J. Williamson 2011. Evidential probability and objective Bayesian epistemology. In *Philosophy of Statistics*, **volume 7** of *Handbook of the Philosophy of Science*, ed. P.S. Bandyopadhyay and M.R. Forster, 307–331. New York: Elsevier.

Williamson J. 2007. Motivating objective Bayesianism: from empirical constraints to objective probabilities. In *Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr.*, **volume 2** of *Texts in Philosophy*, ed. W. Harper and G. Wheeler, 155–183. London: College Publications.

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

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

Williamson J. 2011. Objective Bayesianism, Bayesian conditionalisation and voluntarism. *Synthese* **178**(1), 67–85.

Williamson J. 2013. Why frequentists and Bayesians need each other. *Erkenntnis* **78**(2), 293–318.

Wilson N. 2001. Modified upper and lower probabilities based on imprecise likelihoods. In *ISIPTA '01, Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, Ithaca, NY, USA*, ed. G. de Cooman, T. Fine and T. Seidenfeld, 370–378. Shaker.

Zabell S.L. 1996. Confirming universal generalizations. *Erkenntnis* **45**(2–3), 267–283.

Technical Appendix

6.1 States and observation histories

The set of decision-relevant states is $\Omega := \{\omega_h, \omega_t\}$ where ω_h is the state where the coin lands heads and ω_t is the state where it lands tails. A typical element of Ω will be denoted ω_i. The set of observation histories is $S := \mathbf{S} \cup \overline{s}$ with a typical element s, where $\mathbf{S} := \Omega^T$ is the set of non-empty histories that can be generated by $T \geq 1$ coin tosses and $\overline{s} = \emptyset$ is the "no observation" history. The counting function $\kappa : \mathbf{S} \to \mathrm{Z}_{\geq 0}$ is such that, for any history $s \in \mathbf{S}$, $\kappa(s) := n(t \in T : (s)_t = \omega_h)$, where $n(\cdot)$ denotes the cardinality of the set. For any history $s \in \mathbf{S}$, $n(s) \in [0, T]$.

6.2 Choices and payoffs

The set of each player's possible choices is $C := \{c_h, c_t, c_a\}$, where c_h is the choice to bet on ω_h, c_t is the choice to bet on ω_t, and c_a is the choice to abstain from betting. A typical element of C will be denoted c_j. Players' payoffs are represented by a von Neumann–Morgenstern utility function $u : C \times \Omega \to [-1, 1]$, such that $u(c_h, \omega_h) = 1 - \delta$, $u(c_h, \omega_t) = -\delta$, $u(c_t, \omega_t) = \delta$, $u(c_t, \omega_h) = \delta - 1$, $u(c_a, \omega_h) = u(c_a, \omega_t) = 0$, where $\delta \sim U[0, 1]$ is the randomly selected price in a particular game.

6.3 Stan

The beliefs of the Standard Bayesian agent, *Stan*, can be represented with a model $M_{SB} := \{\Omega, S, \Theta, p, f\}$, where Ω is the set of states, S is the set of observation histories, $\Theta := \{x \in \mathbb{R} : x \in [0, 1]\}$ is the set of possible coin biases towards ω_h with a typical element θ, $p : S \to \Delta^+(\Theta)$ is the credence function representing *Stan*'s beliefs about coin biases, and $f : \{p\} \times S \to \Delta^+(\Omega)$ is the aggregate credence function representing *Stan*'s beliefs about decision-relevant states.

Stan knows that each coin toss is a Bernoulli trial (a binomial event in an exchangeable sequence) and that the coins tosses' probabilities have a binomial distribution. Thus, *Stan*'s beliefs can be conveniently modelled with a beta distribution $Beta(a, b)$ as a probability density function $p(\overline{s})$ where, for each $\theta \in \Theta$,

$$p(\theta|\overline{s}) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}, \quad \text{where} \quad B(a,b) = \int_0^1 \theta'^{a-1}(1-\theta')^{b-1} d\theta' \quad \text{is the Beta function.} \tag{1}$$

After T observations, *Stan*'s evidence is a non-empty history $s \in \mathbf{S}$. This agent updates their credence in each coin bias $\theta \in \Theta$ by revising the prior credence

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

$p(\theta|\bar{s}) \in (0,1)$ via Bayes' rule:

$$p(\theta|s) = \frac{p(\theta|\bar{s})p(s|\theta)}{p(s)}. \tag{2}$$

The Bayes' rule can be reformulated using a counting function κ for each history $s \in \mathbf{S}$ and every coin bias $\theta \in \Theta$ as

$$p(\theta|s) = p(\theta|\kappa(s), T) = \frac{p(\theta|\bar{s})p(\kappa(s), T|\theta)}{p(\kappa(s), T)} = \frac{\theta^{a+\kappa(s)-1}(1-\theta)^{b+T-\kappa(s)-1}}{B(a+\kappa(s), b+T-\kappa(s))}. \tag{3}$$

The aggregate belief function f is such that, given $\bar{s}$, the prior beliefs in ω_h and ω_t can be defined as

$$f(\omega_h|p,\bar{s}) = \int_0^1 \theta p(\theta|\bar{s})d\theta = \frac{a}{a+b}, \tag{4}$$

$$f(\omega_t|p,\bar{s}) = 1 - f(\omega_h|p,\bar{s}) = \frac{b}{a+b}. \tag{5}$$

The posterior beliefs in ω_h and ω_t given any history $s \in \mathbf{S}$ can be defined as

$$f(\omega_h|p,s) = \int_0^1 \theta p(\theta|\kappa(s), T)d\theta = \frac{a+\kappa(s)}{a+b+T}, \tag{6}$$

$$f(\omega_t|p,s) = 1 - f(\omega_h|p,s) = \frac{b+T-\kappa(s)}{a+b+T}. \tag{7}$$

Stan's prior beliefs are represented by a beta distribution $Beta(1,1)$. *Stan* always chooses an action $c_j \in C$, such that, given any history $s \in S$,

$$c_j \in \arg\max_{c_k \in C}\left(E_u[c_k|p,s]\right), \quad \text{where} \quad E_u[c_k|p,s] = \sum_{\omega_i \in \Omega} f(\omega_i|p,s)u(c_k,\omega_i). \tag{8}$$

6.4 Imprecise Bayesian

An Imprecise Bayesian agent's beliefs in the Bernoulli task can be represented with a belief model $M_{IB} := \left\{\Omega, S, \Theta, P, \hat{f}\right\}$, where Ω is the set of states, S is the set of observation histories, P is the credal set, and $\hat{f} : P \times S \to \Delta^+(\Omega)$ is the aggregate belief function. The credal set P is a set of functions where each function $p \in P$ is the credence function, as defined in Appendix 6.3. We assume that the set P is convex, which means that, for any pair $p, p' \in P$ and any parameter $\lambda \in [0,1]$, there exists a credence function $p'' \in P$, such that $\lambda p(\theta|\bar{s}) + (1-\lambda)p'(\theta|\bar{s}) = p''(\theta|\bar{s})$ for each $\theta \in \Theta$.

After observing any history $s \in \mathbf{S}$, the Imprecise Bayesian agent updates each prior assigned by each credence function $p \in P$ according to Bayes' rule, as defined in Section 6.3. The aggregate belief function $\hat{f}$ assigns, to each credence function-observation history combination $(p,s) \in P \times S$, an aggregate belief $\hat{f}(p,s) \in \Delta^+(\Omega)$ on Ω, where the marginal beliefs about ω_h and ω_t can be defined as

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

$$\hat{f}(\omega_h|p,s) = \frac{a+\kappa(s)}{a+b+T}, \tag{9}$$

$$\hat{f}(\omega_t|p,s) = \frac{b+T-\kappa(s)}{a+b+T}. \tag{10}$$

Since P is convex, the Imprecise Bayesian agent's aggregate beliefs about each $\omega_i \in \Omega$ given any $s \in S$ can be defined as an interval $\Phi_{\omega_i|P,s} := \left[\phi^{min}_{\omega_i|P,s}, \phi^{max}_{\omega_i|P,s}\right]$, where $\phi^{min}_{\omega_i|P,s} := \min_{p\in P}\left(\hat{f}(\omega_i|p,s)\right)$ is the minimum and $\phi^{max}_{\omega_i|P,s} := \max_{p\in P}\left(\hat{f}(\omega_i|p,s)\right)$ is the maximum aggregate belief in ω_i given P and s.

An Imprecise Bayesian agent's belief system about events ω_h and ω_t can be defined as a set of intervals $\Phi_{P,s} := \left\{\Phi_{\omega_h|P,s}, \Phi_{\omega_t|P,s}\right\}$, where $\Phi_{\omega_i|P,s}$ is a typical element of $\Phi_{P,s}$.

6.5 Alpha Cut

At each decision moment, the Alpha Cut agent rules out beliefs about each event $\omega_i \in \Omega$ that fall below a dynamic (i.e. changing with changing beliefs) belief threshold that is defined by the agent's parameter $\alpha \in (0,1)$ and their highest degree of belief in ω_i that the agent holds given the credal set P and the observed history s. A set of the agent's beliefs about any event $\omega_i \in \Omega$ given any convex credal set P and any history $s \in S$ can be defined as

$$\Phi_{\omega_i|\alpha,P,s} := \begin{cases} \Phi_{\omega_i|P,s} & \text{if} \quad \phi^{min}_{\omega_i|P,s} \geq \alpha\phi^{max}_{\omega_i|P,s}, \\ \left[\alpha\phi^{max}_{\omega_i|P,s}, \phi^{max}_{\omega_i|P,s}\right] & \text{otherwise}. \end{cases} \tag{11}$$

An Alpha Cut belief system about events ω_h and ω_t can be defined as a set of intervals $\Phi_{\alpha,P,s} := \left\{\Phi_{\omega_h|\alpha,P,s}, \Phi_{\omega_t|\alpha,P,s}\right\}$, where $\Phi_{\omega_i|\alpha,P,s}$ is a typical element of $\Phi_{\alpha,P,s}$.

6.6 Maximum Likelihood Alpha Cut

Given an Alpha Cut $\Phi_{\alpha,P,s}$, a Maximum Likelihood Alpha Cut player selects a unique interval using the observed history s. In the context of the Bernoulli trial, if the relative frequency of heads in observed tosses is greater than 0.5, then the player uses the interval for ω_h. If this relative frequency is less than 0.5, then the player uses the interval for ω_t. If the relative frequency is 0.5, then the player makes a random selection (that is, chooses according to a uniform probability distribution) between using the interval for ω_h or using the interval for ω_t.

6.7 Calibration

A Calibration player's beliefs in the Bernoulli task can be represented with a belief model $M_{Cal} := \left\{\Omega, S, \Gamma, f^*\right\}$, where Ω is the set of states, S is the set of observation histories, Γ is the set of considered significance levels with a typical element γ, and $f^* : \Omega \times \Gamma \times S \to \mathcal{P}([0,1])$ is a function which assigns, to every state-significance level-observation history combination $(\omega_i, \gamma, s) \in \Omega \times \Gamma \times S$, a Clopper–Pearson

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

interval $f^*(\omega_i, \gamma, s) := \left[\phi^{min}_{\omega_i,\gamma,s}, \phi^{\max}_{\omega_i,\gamma,s}\right]$. The upper and lower bounds of this interval can be derived using the regularized incomplete beta function $I_x(a,b) := \frac{B(x;a,b)}{B(a,b)}$. Given ω_h and any significance level $\gamma \in \Gamma$,

$$\phi^{min}_{\omega_h,\gamma,s} := I^{-1}_{\frac{\gamma}{2}}(\kappa(s), T - \kappa(s) + 1), \tag{12}$$

$$\phi^{max}_{\omega_h,\gamma,s} := I^{-1}_{1-\frac{\gamma}{2}}(\kappa(s) + 1, T - \kappa(s)), \tag{13}$$

and, given ω_t and any significance level $\gamma \in \Gamma$,

$$\phi^{min}_{\omega_t,\gamma,s} := I^{-1}_{\frac{\gamma}{2}}(T - \kappa(s), \kappa(s) + 1), \tag{14}$$

$$\phi^{max}_{\omega_t,\gamma,s} := I^{-1}_{1-\frac{\gamma}{2}}(T - \kappa(s) + 1, \kappa(s)). \tag{15}$$

A Calibration player's belief system can be represented as a set of intervals $\Phi_{\gamma,s} := \left\{\Phi_{\omega_h,\gamma,s}, \Phi_{\omega_t,\gamma,s}\right\}$, where $\Phi_{\omega_i,\gamma,s} := \left[\phi^{min}_{\omega_i,\gamma,s}, \phi^{max}_{\omega_i,\gamma,s}\right]$ is a typical element of $\Phi_{\gamma,s}$.

6.8 The Optimist Decision rule

The Optimist rule can be simultaneously defined for the Imprecise Bayesian, Alpha Cut, Calibration and Maximum Likelihood Alpha Cut players that use it. Given any belief system $\Phi_* := \left\{\Phi_{\omega_h*}, \Phi_{\omega_t*}\right\} \in \left\{\Phi_{P,s}, \Phi_{\alpha,P,s}, \Phi_{\gamma,s}\right\}$, the maximum expected payoffs of actions c_h and c_t can be defined as

$$E^{max}_u[c_h|\Phi_*] := u(c_h,\omega_h)\phi^{max}_{\omega_h*} + u(c_h,\omega_t)\left(1 - \phi^{max}_{\omega_h*}\right), \quad \text{where} \quad \phi^{max}_{\omega_h*} \in \Phi_{\omega_h*} \quad \text{and} \quad \Phi_{\omega_h*} \in \Phi_*, \tag{16}$$

$$E^{max}_u[c_t|\Phi_*] := u(c_t,\omega_h)\left(1 - \phi^{max}_{\omega_t*}\right) + u(c_t,\omega_t)\phi^{max}_{\omega_t*}, \quad \text{where} \quad \phi^{max}_{\omega_t*} \in \Phi_{\omega_t*} \quad \text{and} \quad \Phi_{\omega_t*} \in \Phi_*, \tag{17}$$

while the minimum expected payoffs of actions c_h and c_t can be defined as

$$E^{min}_u[c_h|\Phi_*] := u(c_h,\omega_h)\phi^{min}_{\omega_h*} + u(c_h,\omega_t)\left(1 - \phi^{min}_{\omega_h*}\right), \quad \text{where} \quad \phi^{min}_{\omega_h*} \in \Phi_{\omega_h*} \quad \text{and} \quad \Phi_{\omega_h*} \in \Phi_*, \tag{18}$$

$$E^{min}_u[c_t|\Phi_*] := u(c_t,\omega_h)\left(1 - \phi^{min}_{\omega_t*}\right) + u(c_t,\omega_t)\phi^{min}_{\omega_t*}, \quad \text{where} \quad \phi^{min}_{\omega_t*} \in \Phi_{\omega_t*} \quad \text{and} \quad \Phi_{\omega_t*} \in \Phi_*, \tag{19}$$

Choosing c_a yields a payoff 0 in every state of the world, and so its maximum and minimum expected payoffs given any belief system $\Phi_* \in \left\{\Phi_{P,s}, \Phi_{\alpha,P,s}, \Phi_{\gamma,s}\right\}$ can be defined as

$$E^{max}_u[c_a|\Phi_*] = E^{min}_u[c_a|\Phi_*] = 0. \tag{20}$$

We can now define the Optimist rule using the preceding terms. It is a particular implementation of the Hurwicz criterion (Hurwicz 1951). The Optimist rule requires assigning a weight of 3/4 to each action's maximum expected payoffs and 1/4 to the action's minimum expected payoffs. Given any belief system

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press

$\Phi_* \in \{\Phi_{P,s}, \Phi_{\alpha,P,s}, \Phi_{\gamma,s}\}$, a player using the Optimist rule chooses an action $c_j \in C$, such that

$$c_j \in \arg\max_{c_k \in C}\left(\frac{3}{4}E_u^{max}[c_k|\Phi_*] + \frac{1}{4}E_u^{min}[c_k|\Phi_*]\right). \tag{21}$$

William Peden is a Postdoctoral Researcher in the Institute for Philosophy and Scientific Method at Johannes Kepler University, Linz. He is researching applications of imprecise probabilities in confirmation theory and the philosophy of statistics. Email: william.peden@jku.at

Mantas Radzvilas is a Senior Research Fellow in the Department of Philosophy at the University of Konstanz. His current research focuses on statistical learning and decision-making under uncertainty, as well as on foundational issues of game theory.

Daniele Tortoli is a Research and Teaching Assistant in the Department of Communication and Economics at the University of Modena and Reggio Emilia. Trained in computer science, his research interests include artificial intelligence and quantitative marketing. Email: daniele.tortoli@unimore.it

Francesco De Pretis is an Adjunct Professor and Research Fellow in the Department of Communication and Economics at the University of Modena and Reggio Emilia, and a Visiting Scholar at the School of Public Health, Indiana University Bloomington. An applied mathematician with a focus on risk and uncertainty, his research has appeared in journals such as the *Annals of Operations Research*, *Computational Economics*, the *International Journal of Approximate Reasoning* and *Risk Analysis*. Email: francesco.depretis@unimore.it

Cite this article: Peden W, Radzvilas M, Tortoli D, and De Pretis F (2026). Ambiguity, Coherence and Performance. *Economics and Philosophy* **42**, 420–456. https://doi.org/10.1017/S026626712510059X

https://doi.org/10.1017/S026626712510059X Published online by Cambridge University Press