A comparison of imprecise Bayesianism and Dempster–Shafer theory for automated decisions under ambiguity

MANTAS RADZVILAS, *Department of Philosophy, University of Konstanz, 78464 Konstanz, Germany.*

WILLIAM PEDEN, *Institute for Philosophy and Scientific Method, Johannes Kepler University, 4040 Linz, Austria.*
E-mail: william.peden@jku.at

DANIELE TORTOLI, *Department of Communication and Economics, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy.*

FRANCESCO DE PRETIS, *Department of Communication and Economics, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy. School of Public Health, Indiana University, Bloomington, IN 47405, USA.*

Abstract

Ambiguity occurs insofar as a reasoner lacks information about the relevant physical probabilities. There are objections to the application of standard Bayesian inductive logic and decision theory in contexts of significant ambiguity. A variety of alternative frameworks for reasoning under ambiguity have been proposed. Two of the most prominent are Imprecise Bayesianism and Dempster–Shafer theory. We compare these inductive logics with respect to the Ambiguity Dilemma, which is a problem that has been raised for Imprecise Bayesianism. We develop an agent-based model comparison that isolates the difference between the two inductive logics in their updating methods. We find that Dempster–Shafer theory does not avoid the Ambiguity Dilemma. We discuss the implications of this result.

Keywords: belief functions, decisions under severe uncertainty, Dempster–Shafer theory, formal epistemology, Imprecise Bayesianism, Imprecise probability

1 Introduction

Inductive logics are formal systems for making and evaluating non-deductive inferences. One challenge in inductive logic is representing ambiguity, which occurs insofar as the relevant physical probabilities are unknown. Imprecise Bayesians, who model belief states via sets of probability functions called "credal sets", have argued that their approach is better for representing ambiguity than Standard Bayesianism, which uses a single probability function to model beliefs [3, 8, 16, 33, 34, 77]. The claim is usually that the divergence among the probability functions in the credal set can represent the level of ambiguity, because we have less reason to exclude probability functions from the credal set insofar as we know less about the relevant physical probabilities.

Vol. 00, No. 00,
https://doi.org/10.1093/logcom/exae069

In previous research, we have used agent-based modelling to show how, in a broad category of decision problems, Imprecise Bayesian inductive logics can face what we have called the “Ambiguity Dilemma” [59]. In brief, the problem is that, in decision problems based around Bernoulli trials with an initially unknown bias, increased divergence in credal sets results in slower updating. In turn, this slower updating results in worse decision-making performance. “Washing out” convergence does not occur fast enough to stop a detectable difference emerging between (a) Imprecise Bayesian agents with divergent credal sets and (b) a Standard Bayesian using a flat prior. Moreover, no decision rule has been found that can avoid this problem for Imprecise Bayesians; some make it worse. Hence, there is a dilemma: in the type of decision problem used so far, to what extent should we prefer the representational tools of Imprecise Bayesianism or the potential decision-making advantages of some Standard Bayesian probability distributions?

Thus far, the Ambiguity Dilemma has only been investigated for Imprecise Bayesianism, which updates via a version of conditionalization. Yet that is not the only way that non-additive belief representations can be updated. Dempster–Shafer theory is a broad family of approaches to uncertainty based on Arthur P. Dempster’s rule of combination and subsequent research by Glenn Shafer. Dempster–Shafer theory has been expanded and refined to become a prominent paradigm in inductive logic, knowledge representation, and formal epistemology.

Dempster–Shafer theory has some prima facie grounds for being promising as a means of avoiding the Ambiguity Dilemma, given the role of slower updating in Imprecise Bayesian players’ relative underperformance in previous studies. It can be proven that, given the same evidence and initial beliefs, Dempster–Shafer theory always places bounds on one’s beliefs that are at least as tight as those of Imprecise Bayesianism, and sometimes tighter [38, pp. 286–287]. However, this theorem raises the question as to whether these tighter bounds result in significantly better performances.

Consequently, our research question was simple: does *specifically* adopting the update approach of Dempster–Shafer theory improve on Imprecise Bayesianism with respect to the Ambiguity Dilemma? We continue to build on methods originally developed by Henry E. Kyburg and Choh Man Teng [40, 57–59]. In this approach to comparing inductive logics, the reasoning systems are coded as “players” of a game based on binomial trials, in an agent-based model. While not all decision problems are analogous to this situation, there is a wide range of situations that can be modelled (at least as an idealization) as featuring binomial trials or other types of exchangeable events. Thus, from a precisely specified agent-based model, we are able to unveil substantial information about the inductive logics’ comparative performances. In addition to identifying these differences, the controlled environment of these tests enabled us to analyse their causes.

To isolate the particular effects of applying Dempster’s rule of combination rather than Imprecise Bayesian updating, we controlled for the choice of decision rule and the initial belief state. Thus, we used the same set of decision rules and the same initial (interval-valued) belief state for both the Dempster–Shafer players and the Imprecise Bayesian players in our tests.

In Section 2, we provide the background to our study. In Section 3, we explain our agent-based model’s decision problem and our tests for evaluating performances. In Sections 4 and 5, we respectively detail the Imprecise Bayesianism and Dempster–Shafer theory update rules in our tests. In Section 6, we explain how these update rules were paired with decision rules to create players for our tests. In Section 7, we compare the results. In Section 8, we conclude that switching to Dempster–Shafer updating does not, in itself, avoid the Ambiguity Dilemma. However, Dempster–Shafer theory is a very flexible approach, so we hope that our results stimulate further investigation of ways to avoid or mitigate the Ambiguity Dilemma using this paradigm of inductive logic.

2 Standard Bayesianism, Imprecise Bayesianism and Dempster–Shafer theory

We shall briefly contextualize these theories and our research question. We provide more detailed explanations of each theory in Sections 3, 4, and 5.

Among quantitative inductive logics, "Standard" Bayesianism is currently the most popular approach [27, 62, 73]. In this approach, a reasoner's uncertainties are represented via additive probabilities, usually interpreted in terms of (actual or ideal) degrees of belief. The probability functions assigning these degrees of belief are "credence functions". "Credences" is another term for degrees of belief. Additionally, Standard Bayesianism requires that the credences are updated via conditionalization, following Bayes' rule.

There is relatively little controversy about the applicability of Standard Bayesian reasoning to situations where we know or can assume the relevant physical probabilities. Aside from a few objections, such as objections to the very idea of "degrees of belief," even critics of Standard Bayesianism would use Bayes' rule in classic gambling problems [72]. To take a textbook example, given background knowledge of the relative frequencies of red balls in each of a group of urns, and a random selection (behind a screen) of a red ball from one of the urns, even most critics of Standard Bayesianism would use Bayes' rule if they were betting about which urn was the source of the red ball.

This consensus dissipates once a greater degree of *ambiguity* enters into the problems under analysis. In the most general sense, ambiguity occurs insofar as a reasoner lacks evidence regarding the relevant hypotheses and possible actions. The most prominent type of ambiguity in the literature occurs insofar as the reasoner lacks information about the relevant physical probabilities. For instance, if you are betting about independent and identically distributed (i.i.d.) coin tosses with unknown bias, then Standard Bayesianism requires that you have equally precise credences before and after observing a large sample of the coin tosses. If the posterior matches the prior, then the credence in the hypothesis is unchanged, despite the addition of evidence. This feature, sometimes called the "Paradox of Ideal Evidence," has been the grounds of many criticisms of Standard Bayesianism [33, 56, 60]. More generally, it has been argued that Standard Bayesianism involves "spurious precision" because credences of great exactitude, complexity, and strength can be assigned with minimal information [3, 37, 51].

A related criticism is the idea that initial beliefs in ambiguity should be neutral (or at least as neutral as possible) between different hypotheses, yet representing neutral belief in Standard Bayesianism is infamously difficult and arguably impossible. The traditional approach is to use a symmetry principle, such as the Principle of Indifference, the Maximum Entropy Principle, or some other rule for assigning "neutral" probabilities [13, 30, 42, 79, 80]. These symmetry principles define conditions under which hypotheses or events should be assigned equal probabilities. For example, the Principle of Indifference requires assigning a uniform distribution to the fundamental states of the domain.

The problem is that, while there is a natural sense in which a uniform distribution across some partition is neutral with respect to that partition, it cannot be neutral with respect to all partitions of the domain [17, 46, p. 73]. Most famously, symmetry principles typically imply probabilities of zero for all universal hypotheses over infinite domains [12, 82, 84]. Even outside of this extreme case, neutrality for some sets of hypotheses or events can imply strongly non-neutral probabilities for other sets, due to the requirements of additivity and coherence. Consider the implications of assigning equal probabilities to heads and tails for each of a sequence of coin tosses. For unions and intersections of these events, this assignment will require strong probabilities (values close to 0 or 1) for some unions, strong probabilities against some intersections, and so on. The desire for better representations of neutrality is one reason why some have been attracted to Dempster–Shafer theory [69, pp. 22–25 and pp. 207–208]; [14].

Thus, ambiguity-related criticisms have been one reason for the development of alternative inductive logics that are putatively superior for representing ambiguity. One of the most prominent is Imprecise Bayesianism, in which credal sets are used to model belief states. Many Imprecise Bayesians represent greater ambiguity via greater divergence in a credal set for a hypothesis's credence assignment. Moreover, insofar as there are special cases where Standard Bayesianism is the appropriate representation of the ambiguity in an agent's evidence (perhaps when there is no ambiguity) Imprecise Bayesianism can mimic Standard Bayesianism via a completely non-divergent set. Hence, Imprecise Bayesianism offers a strictly more flexible approach for representing ambiguity.

The Ambiguity Dilemma shows that, in a broad class of situations, these representations can have costly effects for decision-making performance. In particular, in the type of decision problem used so far, greater divergence results in (a) slower convergence of beliefs to the sample frequency, in the presence of reliable sample data and (b) for some decision rules, a failure to utilize reliable information in sample data when making decisions [59].

However, Imprecise Bayesianism is not the only approach to representing ambiguity in a formalism that is strictly more flexible than Standard Bayesianism. Dempster–Shafer theory is another prominent approach among non-additive representations of beliefs and evidential support relations. In Dempster–Shafer theory, greater ambiguity can be represented by a greater difference between "belief" and "plausibility" as defined in Section 4 below. Thus far, the Ambiguity Dilemma has only been studied in relation to Imprecise Bayesianism. This raises the question: can Dempster–Shafer theory do better?

We stress that our study examines the question whether *specifically* adopting updating via Dempster–Shafer theory makes a difference to the Ambiguity Dilemma. Like Imprecise Bayesianism, Dempster–Shafer theory is an extremely flexible approach that allows for many representations of initial beliefs. Additionally, it would be impossible to cover all the different situations where the evidence seems ambiguous. For these reasons, we employed a previously utilised way to represent the initial ambiguity in a particular decision problem where the Ambiguity Dilemma occurs for Imprecise Bayesianism. We then tested whether using Dempster–Shafer theory, using the same approach that most Imprecise Bayesians would use for this decision problem, has the effect of mitigating, aggravating, or not affecting the Ambiguity Dilemma.

3 Decision problem and tests

In this section, we first describe the decision problem that we used as the basis of our tests, before detailing the tests that we used for comparisons. We selected a problem in the literature that (a) features pertinent ambiguity and (b) examines short-run performances.

The latter desideratum is important, because in a wide range of problems there will be long-run convergence between the decisions made by Standard Bayesians, Imprecise Bayesians and Dempster–Shafer reasoners. Hence, our focus on short-run performances is suitable for identifying any differences between Dempster–Shafer theory and Imprecise Bayesianism with respect to the Ambiguity Dilemma.

3.1 Decision problem

The agent-based model is based on a non-interactive "game," consisting of observations of simulated coin tosses. However, unlike real coin tosses, the players of this game lack almost any background knowledge of the stochastic properties of these coin tosses. In particular, we assume only that the players know that the coin tosses are Bernoulli trials, so that they are (a) binomial with tosses landing either heads or tails and (b) i.i.d. and therefore exchangeable.

Players make their decisions myopically and separately. The latter implies that players cannot change their decision rules on the grounds that some other player is performing better. The former implies that their decision-making is purely concerned with their payoffs (and conformity to their rules) for a single game and without the need for strategic reasoning. For this reason, we also assume that players always have sufficient resources to bet. Finally, we stress that players are aiming to maximize payoffs within the prescriptions of their rules, rather than trying to maximize some function of our assessment criteria.[1]

At predetermined points in the sequence, players can choose whether to bet on one or other result of the last coin toss in a game. Each game consists of 5 tosses, with an opportunity to bet on the fifth toss. The players accumulate their observations of coin tosses over the games. Thus, for their first choice in the first game, they have four observed coin tosses; in the second game, they have nine (the five from the first game, plus four in the second game) and so on as the games progress. In our main comparisons, there were 1000 games, each consisting of 5 tosses.

In detail, for each toss, players observe either ω_h, the state that the coin landed heads or ω_t, the state that the coin landed tails. We define the set of these possible states, $\Omega := \{\omega_h, \omega_t\}$, with a typical element ω_i, so that Ω contains every possible outcome of an observation. A sequence of tosses is called a "history'. We define the set of possible histories that can occur in a finite number $T \geq 1$ of coin tosses as $\mathcal{S} := \Omega^T$. The set of all possible histories is $S := \left\{\mathcal{S} \cup \{\bar{s}\}\right\}$ with a typical element s. Thus, each history $s \in \mathcal{S}$ is a sequence $s := (s_1, \ldots, s_T)$ where, for each $t \in [1, T]$, $s_t \in \Omega$. When describing the players' update rules, it will be useful to refer to their evidence prior to making any observations, so we also define the "no observation" history as $\bar{s} := \varnothing$.

The players make statistical inferences using their observed sample frequencies. To formalize these sample frequencies for players' update rules, we use a counting function in terms of ω_h. This function is defined as $\kappa : \mathcal{S} \to \mathbb{Z}_{\geq 0}$ such that, for every history $s \in \mathcal{S}$, $\kappa(s) = n\left(\left\{t \in T : (s)_t = \omega_h\right\}\right)$, where $n(\cdot)$ is the set's cardinality.

At the predetermined points in the sequences of tosses (the end of each game) players have a choice among a set of actions $C := \{c_h, c_t, c_a\}$ with a typical element c_j, where c_h represents betting on heads, c_t represents betting on tails, and c_a represents abstention from betting. Players' payoffs from games are represented with a von Neumann-Morgenstern utility function $u : C \times \Omega \to [-1, 1]$, that assigns, to each possible action-state combination $\left(c_j, \omega_i\right) \in C \times \Omega$, a cardinal utility payoff $u\left(c_j, \omega_i\right) \in [-1, 1]$. For brevity, we shall refer to just "payoffs".

Players know the payoff structure of the games, which is described in Figure 1. This figure details the action payoffs that are associated with every possible outcome in a particular game. A player wins a payoff of $(1 - \delta)$ if they choose c_h when a game's final toss results in ω_h. They win δ if they choose c_t when a game's final toss results in ω_t. In contrast, if their chosen bet's outcome does not occur, then they lose $-\delta$ if they choose c_h but the state is ω_t. Finally, they lose $(\delta - 1)$ if they choose c_t and game's final toss is ω_h.

Note how the payoffs are defined in terms of δ, except if a player chooses to abstain from betting in a particular game via action c_a to receive a guaranteed result of 0. The values of δ were restricted such that $\delta \in [0, 1]$. The values were generated randomly in our tests. They were the same for all players, as we describe in more detail in Subsection 3.3.

[1] We thank Teddy Seidenfeld for encouraging reflection on these points.

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

FIGURE 1. Player Payoff Matrix.

3.2 The Standard Bayesian

For the previously described decision problem, the Ambiguity Dilemma involves comparative performance relative to a Standard Bayesian player with a flat prior [57]. In the context of the Ambiguity Dilemma, we call this player "*Stan*". *Stan* performs well in this decision problem and sets a benchmark.

Stan's details are very important, since they will form the basis of our exposition of the Imprecise Bayesian and the Dempster–Shafer players. We define *Stan* via an epistemic model $\mathcal{B} := \{\Omega, \Theta, S, \kappa, p\}$, where $p : S \to \Delta^+(\Theta)$ is a credence function that assigns a strictly positive probability distribution $p(s) \in \Delta^+(\Theta)$ on Θ to every history $s \in S$. As before, Ω is the set of states, $\Theta := \{x \in \mathbb{R} : x \in [0,1]\}$ is the set of coin biases[2] towards ω_h with a typical element θ, S is the set of possible observation histories, and κ is the counting function for ω_h. The parameter $p(\theta \mid s) \in (0,1)$ is the marginal probability of a coin bias $\theta \in \Theta$ relative to a history s.

Stan's credences are beta distribution priors, matching both earlier studies and common practice in Bayesian statistics for binomial decision problems. Beta distribution priors have the useful feature that they are conjugate priors for the binomial likelihood function, meaning that the posterior distribution is also a beta distribution. This greatly simplifies both the computation and analysis of updating using beta distributions, which is one reason for their popularity when developing Bayesian models of binomial phenomena. A beta distribution is characterised by a beta function $B(a,b)$, where $a > 0$ and $b > 0$ are hyperparameters (specifically, shape parameters) specifying the particular beta distribution. In our model, a defines the degree of bias in the beta distribution towards heads; b is the corresponding bias towards tails.

Stan has a flat prior $B(1,1)$. This prior is initially equivocal between heads and tails, but rapidly converges towards the sample frequency. Not only are flat priors popular in Bayesian statistics for this type of decision problem: these properties are also shared by other popular approaches, such as the Jeffreys prior of $B(0.5, 0.5)$. The Jeffreys prior is a very different distribution of the probability mass, but it is also equivocal and converges rapidly.

Stan's evidence consists of observing a sequence of coin tosses. *Stan* updates their credence in each $\theta \in \Theta$ by revising p using Bayes' rule:

$$p(\theta \mid s) = \frac{p(\theta \mid \bar{s})\, p(s \mid \theta)}{p(s)}, \text{ where } p(\theta \mid \bar{s}) > 0 \text{ denotes a prior probability of } \theta. \tag{1}$$

Since *Stan* knows that each toss is a Bernoulli trial and the game generates a binomial distribution, the Bayes' rule can be reformulated using a counting function κ for each history $s \in \mathcal{S}$ and every

[2] For conciseness, we describe the biases of the Bernoulli trials as "coin biases" even though they are the long-run stochastic tendencies of the whole event, the tossing of the coin.

coin bias $\theta \in \Theta$ as

$$p(\theta \mid \kappa(s), T) = \frac{p(\theta \mid \bar{s})\, p(\kappa(s), T \mid \theta)}{p(\kappa(s), T)}, \tag{2}$$

where $p(\kappa(s), T \mid \theta) = \begin{pmatrix} T \\ \kappa(s) \end{pmatrix} \theta^{\kappa(s)} (1-\theta)^{T-\kappa(s)}$. Thus, the posterior probability distribution $p(\kappa(s), T) \in \Delta^{+}(\Theta)$ and the prior probability distribution $p(\bar{s}) \in \Delta^{+}(\Theta)$ are both beta distributions.

The prior credence in any bias $\theta \in \Theta$ can be described as

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

Thus, the Bayes' rule for each bias and history is

$$\begin{aligned} p(\theta \mid \kappa(s), T) &= \frac{\begin{pmatrix} T \\ \kappa(s) \end{pmatrix} \theta^{a+\kappa(s)-1} (1-\theta)^{b+T-\kappa(s)-1} \Big/ B(a,b)}{\int_0^1 \left(\begin{pmatrix} T \\ \kappa(s) \end{pmatrix} \theta'^{a+\kappa(s)-1} (1-\theta')^{b+T-\kappa(s)-1} \Big/ B(a,b) \right) d\theta'} \\ &= \frac{\theta^{a+\kappa(s)-1}(1-\theta)^{b+T-\kappa(s)-1}}{B(a+\kappa(s), b+T-\kappa(s))}. \end{aligned} \tag{4}$$

Consequently, *Stan*'s posterior is another beta distribution characterised by $a + \kappa(s)$ and $b + T - \kappa(s)$.

We define an aggregate belief function $f : \{p\} \times S \to \Delta^{+}(\Omega)$, such that, given $\bar{s}$, the prior belief in ω_h is

$$f(\omega_h \mid p, \bar{s}) = \int_0^1 \theta p(\theta)\, d\theta = \frac{a}{a+b}, \tag{5}$$

while we define the prior belief in ω_t as

$$f(\omega_t \mid p, \bar{s}) = 1 - f(\omega_h \mid p, \bar{s}) = \frac{b}{a+b}. \tag{6}$$

Stan's posterior belief in ω_h given $s \in \mathcal{S}$ is

$$f(\omega_h \mid p, s) = \int_0^1 \theta p(\theta \mid \kappa(s), T)\, d\theta = \frac{a + \kappa(s)}{(a+\kappa(s)) + (b + T - \kappa(s))}, \tag{7}$$

while their posterior belief in tails given that history s is

$$f(\omega_t \mid p, s) = 1 - f(\omega_h \mid p, s) = \frac{b + T - \kappa(s)}{a+b+T}. \tag{8}$$

We define *Stan*'s expectation-based choice using a model $\mathcal{D} := \{C, S, \pi, p\}$, where C is the set of possible actions, S is the set of possible histories, u is the payoff function, and p is *Stan*'s credence function. For any history $s \in S$, the expected payoff from some action $c_j \in C$ is

$$\mathbb{E}_u\left[c_j \mid p, s\right] = u\left(c_j, \omega_h\right) f(\omega_h \mid p, s) + u\left(c_j, \omega_t\right) f(\omega_t \mid p, s). \tag{9}$$

c_j \ ω_i	ω_h	ω_t	Expected Payoff
c_h	$(0.2)\,(0.45) = 0.09$	$(0.8)(-0.55) = -0.44$	-0.35
c_t	$(0.2)\,(-0.45) = -0.09$	$(0.8)(0.55) = 0.44$	0.35
c_a	$(0.2)(0) = 0$	$(0.8)(0) = 0$	0

FIGURE 2. Player Payoffs with a 0.2 probability for heads.

c_j \ ω_i	ω_h	ω_t	Expected Payoff
c_h	$(0.984)\,(0.45) = 0.4428$	$(0.016)(-0.55) = -0.0088$	0.434
c_t	$(0.984)\,(-0.45) = -0.4428$	$(0.016)(0.55) = 0.0088$	-0.434
c_a	$(0.984)(0) = 0$	$(0.016)(0) = 0$	0

FIGURE 3. Player Payoffs with a 0.984 probability for heads.

It follows that *Stan* always has a unique expected payoff associated with each action, so they can use payoff maximization to make their decisions: *Stan* always chooses an action $c_j \in C$, such that

$$c_j \in \arg\max_{c_k \in C} \left(\mathbb{E}_u \left[c_k \mid p, s\right]\right). \tag{10}$$

In our tests, *Stan* had a unique expected payoff-maximizing action in every game.

3.3 Tests

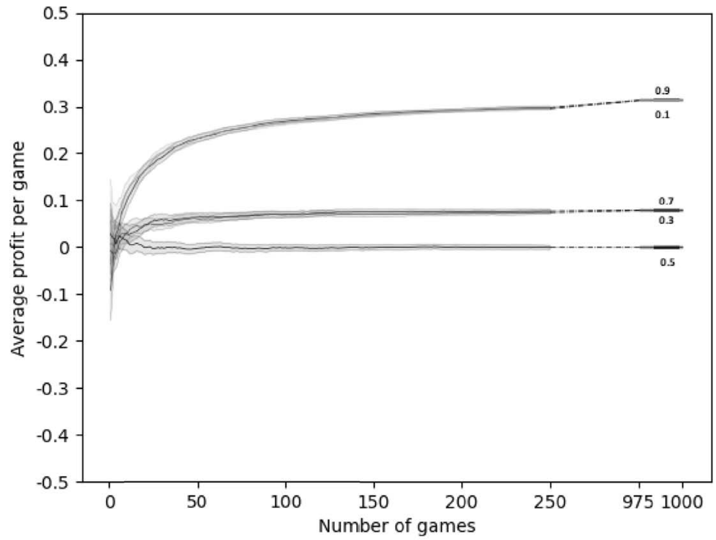
Our comparisons of players' performances were all based around a series of "tests." These consisted of 1000 games, each with randomly generated tosses. Players were unaware of the coin toss sequences at the beginning of each test, so they had to inductively extrapolate coin biases using their observations. Players did not retain information from test to test. All players faced the same tosses. We investigated 5 different coin biases, defined in terms of heads, with values of 0.1, 0.3, 0.5, 0.7 and 0.9. For example, a 0.9 bias means that the long-run probability of generating heads tosses in that setting is 0.9. We investigated both 0.3 and 0.7, as well as 0.1 and 0.9, to check for asymmetries in players' rules. Our tests did not reveal any significant player performance asymmetries across these pairs of biases.

We randomly generated sequences of 1000 ticket prices for each test. Since each player faced the same set of randomly generated tosses and ticket prices, we could regard the tests as controlled virtual experiments. Since the tests were independent and randomised, we could use them for confidence interval estimates of measures of players' performances. We ran 1000 tests per bias, so that our confidence interval estimates of players' performances were both informative and possessed a low risk of random error. All of these confidence intervals were estimated at the 0.95 level.[3]

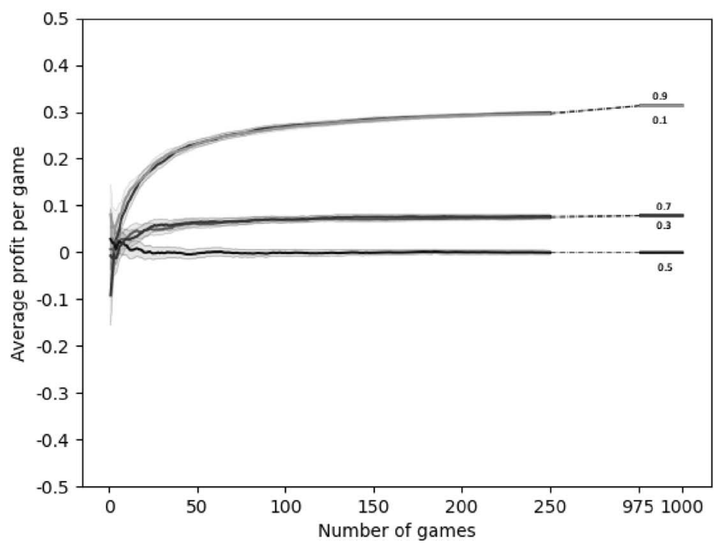
3.4 Comparison methods

We used four types of comparison methods. First, we constructed graphs of players' average payoffs over the tests; this method has the advantage of showing the evolution of players' performances, to diagnose the causes of differences. Second, we examined specific aspects of players' performances, such as the rates at which they lost money in absolute terms and their performance in the extremely

[3] To ensure compatibility with earlier research, our tests had the same technical specifications — see [58]. This included the use of the `statsmodel` econometric and statistical library [65].

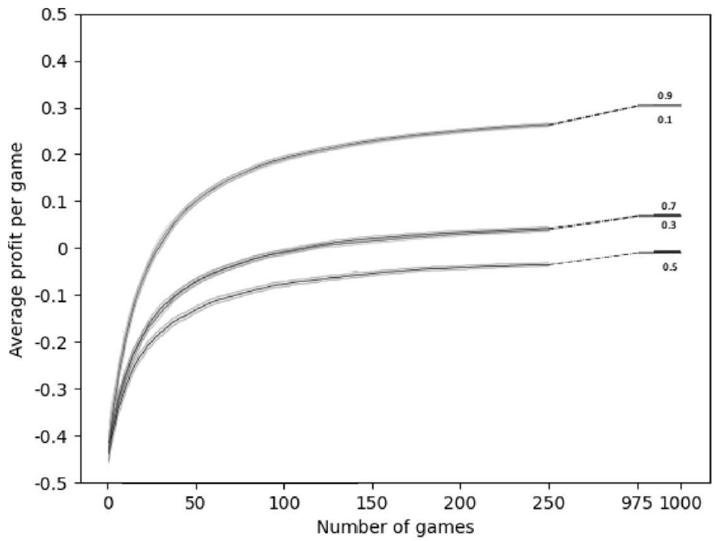


(a) DS-Optimist

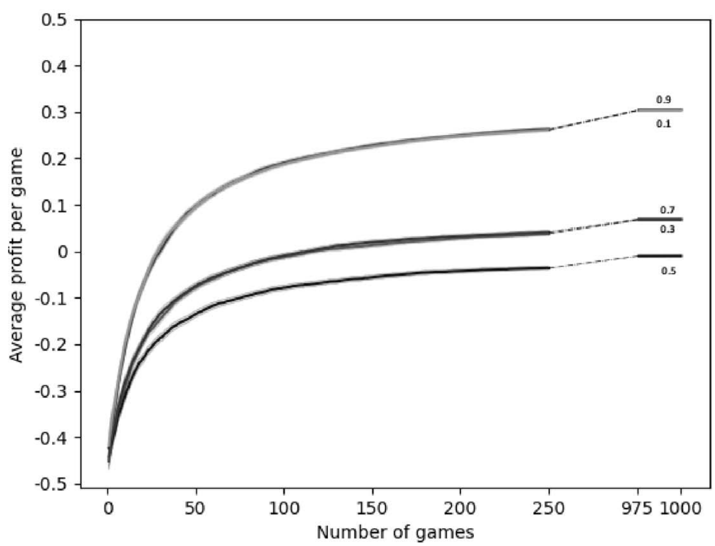


(b) IB-Optimist

FIGURE 4. Optimists: Dempster–Shafer players vs. Imprecise Bayesians. Lines represent the average profit per bet (Y-axis) against the number of bets (X-axis) for DS-Optimist players (top graph) and Imprecise Bayesians (bottom graph). The confidence intervals around them are calculated at the 0.95 level. Numbers adjacent to lines mark the coin bias.



(a) DS-Maximin



(b) IB-Maximin

FIGURE 5. Maximin: Dempster–Shafer players vs. Imprecise Bayesians. Lines represent the average profit per bet (Y-axis) against the number of bets (X-axis) for DS-Optimist players (top graph) and Imprecise Bayesians (bottom graph). The confidence intervals around them are calculated at the 0.95 level. Numbers adjacent to lines mark the coin bias.

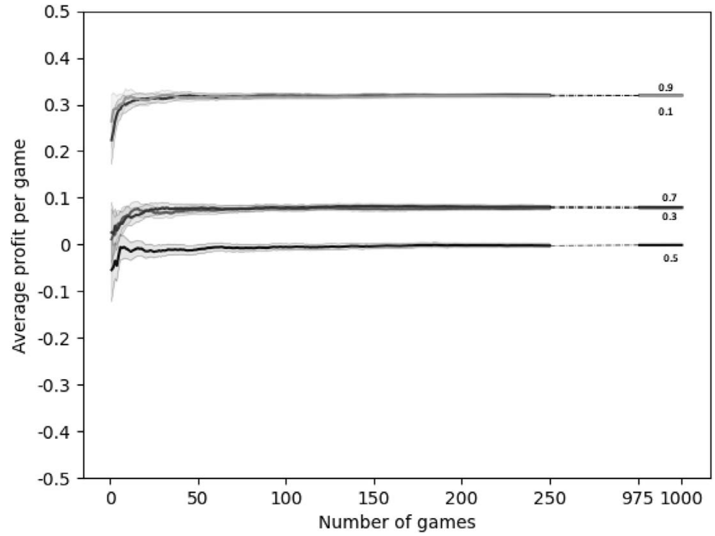


FIGURE 6. Stan. Average profit per bet (Y-axis) against the number of bets (X-axis). Solid lines are the averages. The confidence intervals around them are calculated at the 0.95 level.

short run, because short-run aggregate payoff maximization is not the only criterion that could be used when assessing decision-making. Third, we used *Stan* as a benchmark, to evaluate comparative performance with respect to the Ambiguity Dilemma.

3.4.1 Graphical comparisons We constructed graphs to show both the average profit rates over the tests, as well as their evolutionary path for players. We found very regular results for all the player-decision rule combinations that we investigated. Hence, we do not provide graphs for all their performances. Instead, we show the results for (a) *Stan* (b) the best and worst performing Imprecise Bayesian players, and (c) the best and worst Dempster–Shafer players. These results are reported in Figures 4, 5 and 6.

The graphs show player performances (in terms of payoffs) at each moment in a test, averaged over the 1000 tests that we performed. The zero point in each graph is the average of all players' performances for a 0.5 bias, to provide a common and symmetric scale across all of the graphs.

As a supplement to the graphs, we provide information about the average frequencies that players made particular types of decision in the problem, in Tables 1 to 4.

3.4.2 Caution Check The graphs do not show the rates at which players make net losses at points in the tests. However, these rates can be important in real-world contexts. Hence, the rates of net losses are a further metric to examine. Following our earlier research, we call this metric the "Caution Check" [59]. We report the Caution Check results in Figures 8 and 9.

One might think that avoiding net losses is a relatively weak test for cautious reasoning, but, in fact, it is sufficiently strong to show differences in players' performances. Stronger tests of successful loss avoidance would involve more controversial assumptions, such as choosing what constitutes "unacceptable" net losses. In contrast, frequencies of net losses provide a more clearly appropriate way to assess whether players are being successfully cautious.

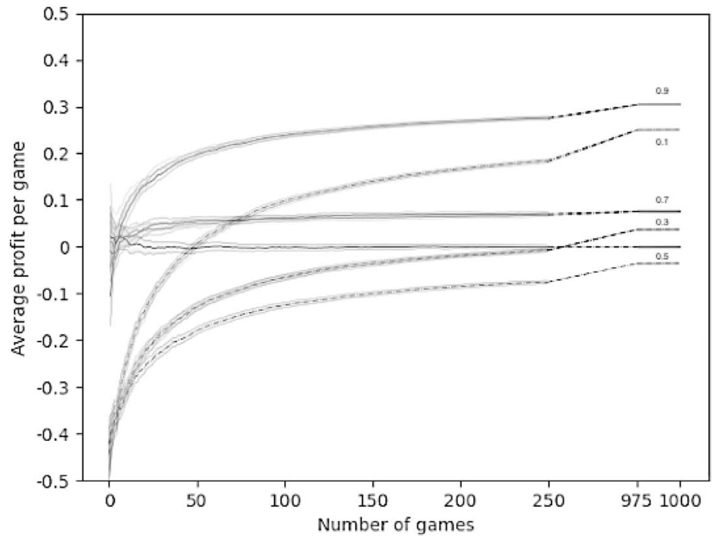


FIGURE 7. DS-Dominance+Ultra-Optimist vs. DS-Dominance. Average profit per bet (Y-axis) against the number of bets (X-axis). DS-Dominance+Ultra-Optimist is represented by the solid lines. DS-Dominance is represented by dashed lines. The confidence intervals around them are calculated at the 0.95 level. Numbers adjacent to lines mark the coin bias.

TABLE 1 Possible player conditions.

Condition	Description
K_1	Directly choosing h.
K_2	Directly choosing t.
K_3	Directly choosing a.
K_4	Randomizing between h and t.
K_5	Randomizing between h and a.
K_6	Randomizing between t and a.
K_7	Randomizing between h, t and a.

3.5 Ultra-Minimum Evidence Comparisons

We examined players performances in extremely short periods of 5, 10, 25 and 50 games, because differences in loss avoidance might only manifest in the extremely short-run. We call this assessment the "Ultra-Minimum Evidence Comparison" [59] because it focuses on the early period where players have the smallest quantity of coin toss observations.

We found that, as the tests proceeded, there was an approximately linear convergence towards the patterns in our main comparisons. Hence, we only show the results for the average performances over the first five games, because this period featured the largest differences from our main comparisons. We report these results in Figures 10 and 11.

TABLE 2 Stan and MaxEnt players. Mean values and related standard errors multiplied by $z = 1.96$ (5% significance level) of the number of times a given condition was met by the pure action players. Statistics based on 1000 tests, each comprising 1000 games. Conditions' legend provided in Table 1.

Coin Bias	Condition	Stan	IB-MaxEnt	DS-MaxEnt
0.1	K_1	102.507 ± 0.689	166.650 ± 0.765	166.509 ± 0.765
	K_2	897.493 ± 0.689	833.350 ± 0.765	833.491 ± 0.765
	K_3	0	0	0
0.3	K_1	300.704 ± 1.033	347.032 ± 1.026	346.833 ± 1.027
	K_2	699.296 ± 1.033	652.968 ± 1.026	653.167 ± 1.027
	K_3	0	0	0
0.5	K_1	500.205 ± 1.115	499.724 ± 0.933	499.725 ± 0.933
	K_2	499.795 ± 1.115	500.276 ± 0.933	500.275 ± 0.933
	K_3	0	0	0
0.7	K_1	699.271 ± 1.066	653.374 ± 1.028	653.545 ± 1.028
	K_2	300.729 ± 1.066	346.626 ± 1.028	346.455 ± 1.028
	K_3	0	0	0
0.9	K_1	898.902 ± 0.705	834.783 ± 0.771	834.944 ± 0.770
	K_2	101.098 ± 0.705	165.217 ± 0.771	165.056 ± 0.770
	K_3	0	0	0

3.6 Standard Bayesian benchmark

In the Ambiguity Dilemma, *Stan* serves as a benchmark that shows what degree of performance is possible. However, in this study, we are not making a general decision-theoretic or formal epistemological comparison of *Stan* and the other approaches. There are too many differences in objectives, philosophical assumptions, and formal differences that are beyond our scope [1, 7, 8, 66, 69, 83]. Instead, comparisons with the performance of *Stan* show the degree of the Ambiguity Dilemma: to what extent does an inductive logic's tools for representing ambiguity cause less successful performances than *Stan's*?

Note that *Stan* is not the only possible Standard Bayesian player that could be implemented for this decision problem. Firstly, while many Bayesian statisticians would use a flat prior for our study's decision problem, this prior is not required by Standard Bayesianism as such. Secondly, there are other priors that would presumably behave similarly to *Stan*, such as a Jeffreys prior [31]. Thirdly, beta distributions are not the only permissible way to reason about Bernoulli trials according to Standard Bayesian inductive logic, but they are popular due to their convenience, responsiveness to evidence, and equivocation across the most basic partition of the sample space — in our decision problem, tosses landing heads or tails.

4 Imprecise Bayesianism

In the next two sections, we begin by describing the representation of beliefs and the update rule for each type of player. Once they are defined, we explain the decision rules that we used to generate their choices given their beliefs in the agent-based model.

As their name suggests, Imprecise Bayesians have similarities to Standard Bayesians. The key difference is that the former's beliefs are modelled as a set of different credence functions, rather

TABLE 3 Imprecise Bayesians. Mean values and related standard errors multiplied by $z = 1.96$ (5% significance level) of the number of times a given condition was met by the mixed action players. Statistics based on 1000 tests, each one comprising 1000 games. Conditions' legend provided in Table 1.

Coin Bias	Condition	Maximin	Optimist	Pessimist	Dominance	ORO	Regret
0.1	K_1	94.158 ± 0.659	132.609 ± 0.725	113.533 ± 0.690	94.158 ± 0.659	103.670 ± 0.671	119.838 ± 0.698
	K_2	829.048 ± 0.768	867.391 ± 0.725	848.288 ± 0.752	829.048 ± 0.768	838.509 ± 0.762	854.689 ± 0.746
	K_3	76.794 ± 0.475	0	38.179 ± 0.356	0	57.821 ± 0.440	25.473 ± 0.298
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	76.794 ± 0.475	0	0
0.3	K_1	277.507 ± 0.991	315.684 ± 1.016	296.567 ± 0.997	277.507 ± 0.991	286.919 ± 0.994	302.961 ± 1.001
	K_2	645.705 ± 1.044	684.316 ± 1.016	664.980 ± 1.021	645.705 ± 1.044	655.242 ± 1.034	671.384 ± 1.024
	K_3	76.788 ± 0.476	0	38.453 ± 0.343	0	57.839 ± 0.442	25.655 ± 0.288
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	76.788 ± 0.476	0	0
0.5	K_1	461.724 ± 1.063	500.228 ± 1.071	480.815 ± 1.077	461.724 ± 1.063	471.271 ± 1.072	487.216 ± 1.082
	K_2	461.501 ± 1.077	499.772 ± 1.071	480.537 ± 1.072	461.501 ± 1.077	470.879 ± 1.082	486.952 ± 1.073
	K_3	76.775 ± 0.478	0	38.648 ± 0.364	0	57.850 ± 0.452	25.832 ± 0.310
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	76.775 ± 0.478	0	0
0.7	K_1	645.984 ± 1.038	684.311 ± 1.039	665.105 ± 1.033	645.984 ± 1.038	655.546 ± 1.038	671.550 ± 1.033
	K_2	277.304 ± 1.007	315.689 ± 1.039	296.603 ± 1.026	277.304 ± 1.007	286.896 ± 1.015	303.035 ± 1.032
	K_3	76.712 ± 0.474	0	38.292 ± 0.343	0	57.558 ± 0.444	25.415 ± 0.287
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	76.712 ± 0.474	0	0
0.9	K_1	830.407 ± 0.770	868.600 ± 0.731	849.513 ± 0.751	830.407 ± 0.770	839.886 ± 0.773	855.920 ± 0.750
	K_2	92.912 ± 0.661	131.400 ± 0.731	112.169 ± 0.702	92.912 ± 0.661	102.395 ± 0.683	118.673 ± 0.714
	K_3	76.681 ± 0.470	0	38.318 ± 0.353	0	57.719 ± 0.438	25.407 ± 0.295
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	76.681 ± 0.470	0	0

TABLE 4 Dempster–Shafer theory players. Mean values and related standard errors multiplied by $z = 1.96$ (5% significance level) of the number of times a given condition was met by the mixed action players. Statistics based on 1000 tests, each one comprising 1000 games. Conditions' legend provided in Table 1.

Coin Bias	Condition	Maximin	Optimist	Pessimist	Dominance	ORO	Regret
0.1	K_1	94.882 ± 0.660	132.862 ± 0.726	113.996 ± 0.689	94.882 ± 0.660	104.086 ± 0.674	120.270 ± 0.697
	K_2	829.282 ± 0.768	867.138 ± 0.726	848.294 ± 0.753	829.282 ± 0.768	838.487 ± 0.763	854.621 ± 0.747
	K_3	75.836 ± 0.476	0	37.710 ± 0.352	0	57.427 ± 0.439	25.109 ± 0.298
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	75.836 ± 0.475	0	0
0.3	K_1	278.082 ± 0.990	315.808 ± 1.015	296.902 ± 0.996	278.082 ± 0.990	287.259 ± 0.994	303.224 ± 1.003
	K_2	646.096 ± 1.044	684.192 ± 1.015	665.097 ± 1.020	646.096 ± 1.044	655.320 ± 1.035	671.426 ± 1.023
	K_3	75.822 ± 0.475	0	38.001 ± 0.341	0	57.421 ± 0.440	25.350 ± 0.287
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	75.822 ± 0.475	0	0
0.5	K_1	462.229 ± 1.064	500.230 ± 1.071	481.060 ± 1.077	462.229 ± 1.064	471.512 ± 1.071	487.387 ± 1.082
	K_2	461.927 ± 1.076	499.770 ± 1.071	480.763 ± 1.071	461.927 ± 1.076	471.071 ± 1.082	487.112 ± 1.072
	K_3	75.844 ± 0.477	0	38.177 ± 0.360	0	57.417 ± 0.452	25.501 ± 0.306
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	75.844 ± 0.476	0	0
0.7	K_1	646.313 ± 1.039	684.195 ± 1.039	665.233 ± 1.033	646.313 ± 1.039	655.626 ± 1.038	671.594 ± 1.032
	K_2	277.877 ± 1.007	315.805 ± 1.039	296.939 ± 1.025	277.877 ± 1.007	287.187 ± 1.013	303.310 ± 1.032
	K_3	75.810 ± 0.473	0	37.828 ± 0.341	0	57.187 ± 0.442	25.096 ± 0.285
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	75.810 ± 0.473	0	0
0.9	K_1	830.676 ± 0.768	868.365 ± 0.731	849.524 ± 0.751	830.676 ± 0.768	839.854 ± 0.772	855.836 ± 0.749
	K_2	93.647 ± 0.663	131.635 ± 0.731	112.634 ± 0.704	93.647 ± 0.663	102.828 ± 0.682	119.057 ± 0.715
	K_3	75.677 ± 0.468	0	37.842 ± 0.354	0	57.318 ± 0.434	25.107 ± 0.294
	K_4	0	0	0	0	0	0
	K_5	0	0	0	0	0	0
	K_6	0	0	0	0	0	0
	K_7	0	0	0	75.677 ± 0.468	0	0

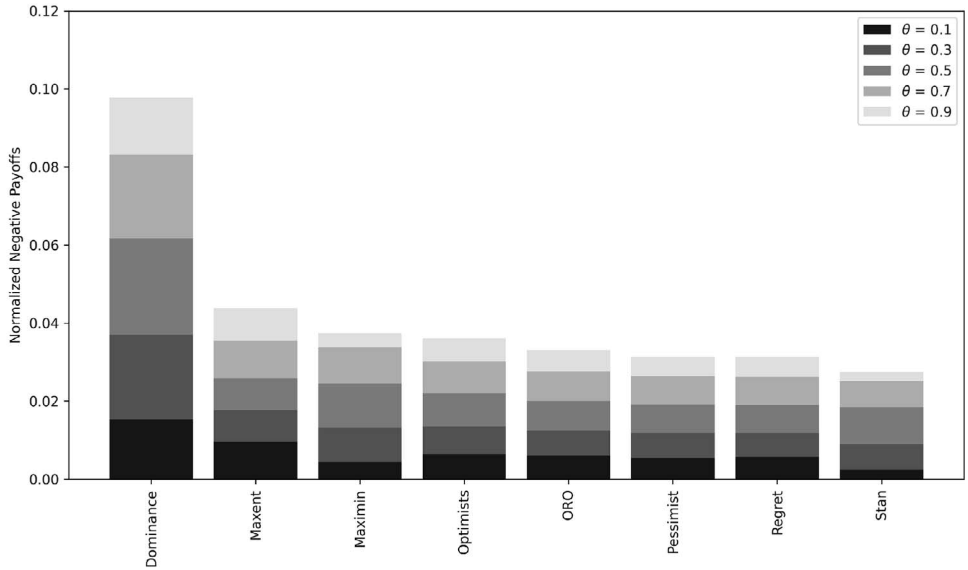


FIGURE 8. Caution Check: Dempster–Shafer players. A stacked bar chart reporting the frequency of net losses in the cumulative payoff series for all tests and all coin biases. Bars are scaled in the unit interval and ordered from tallest (worst—*IB-Dominance*) to smallest (best—*Stan*).

than as a single credence function. We shall refer to these sets of credence functions as "credal sets." One argument that has been presented for Imprecise Bayesianism is that it offers a richer formalism for modelling beliefs in conditions of ambiguity [3, 7, 33, 44, 77].

We formally define a 'credal set' as a convex set P. Not all Imprecise Bayesians make a convexity requirement, but it is a widespread assumption. It usually simplifies the modelling of an agent's beliefs [44, 68]. The convexity requirement implies that if the credal set contains a credence function that assigns a probability of (for example) 0.25 to an event and another credence function that assigns a probability of 0.75 to that event, then the credal set also contains credence functions assigning each intermediary real value to that event. Each credence function in P corresponds to a beta distribution.

We shall define a generic player template, *IB*, who becomes a full player when combined with a decision rule. *IB*'s credal set is defined via a function $\phi : P \times S \to \Delta^{+}(\Omega)$ that assigns, to each credence function-history combination $(p, s) \in P \times S$, a probability distribution $\phi(p, s) \in \Delta^{+}(\Omega)$. *IB* updates a credal set by conditionalizing each credence function in P on their new evidence.

For any set P, any history $s \in S$ and any state $\omega_i \in \Omega$, we can define *IB*'s minimum and maximum aggregate beliefs in ω_i as $\varphi^{min}_{\omega_i|s} := \min_{p \in P}(\phi(\omega_i \mid p, s))$ and $\varphi^{max}_{\omega_i|s} := \max_{p \in P}(\phi(\omega_i \mid p, s))$. Since the set P is assumed to be convex, *IB*'s set of aggregate beliefs about ω_h and ω_t can be represented by a function $\mu_P : \Omega \times S \to \mathcal{P}([0,1])$ that assigns, to each state-observation history combination $(\omega_i, s) \in \Omega \times S$, an aggregate belief interval $\mu_P(\omega_i, s) := \left[\varphi^{min}_{\omega_i|s}, \ldots, \varphi^{max}_{\omega_i|s}\right]$ with a typical element $\varphi_{\omega_i|s} \in (0, 1)$. The intervals are such that, for any $\varphi_{\omega_i|s} \in \mu_P(\omega_i, s)$, it is the case that $(1 - \varphi(\omega_i, s)) \in \mu_P(\omega_{-i}, s)$.

In sum, the *IB* generic player has a credal set of beta credence functions, in contrast to *Stan*. Like *Stan*, they update by conditionalization, except they must update a set of probability

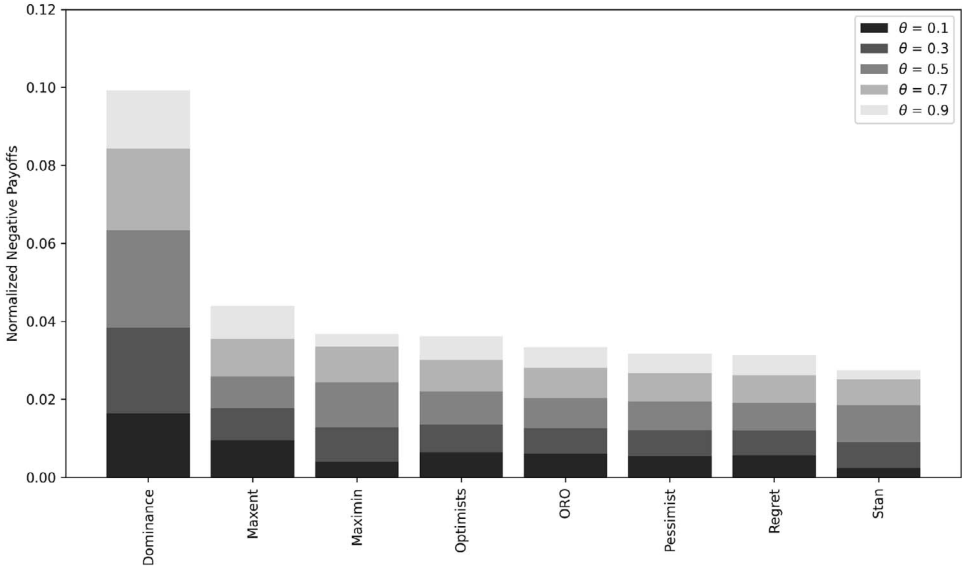


FIGURE 9. Caution Check: Imprecise Bayesians. As Figure 8, but for Imprecise Bayesian players.

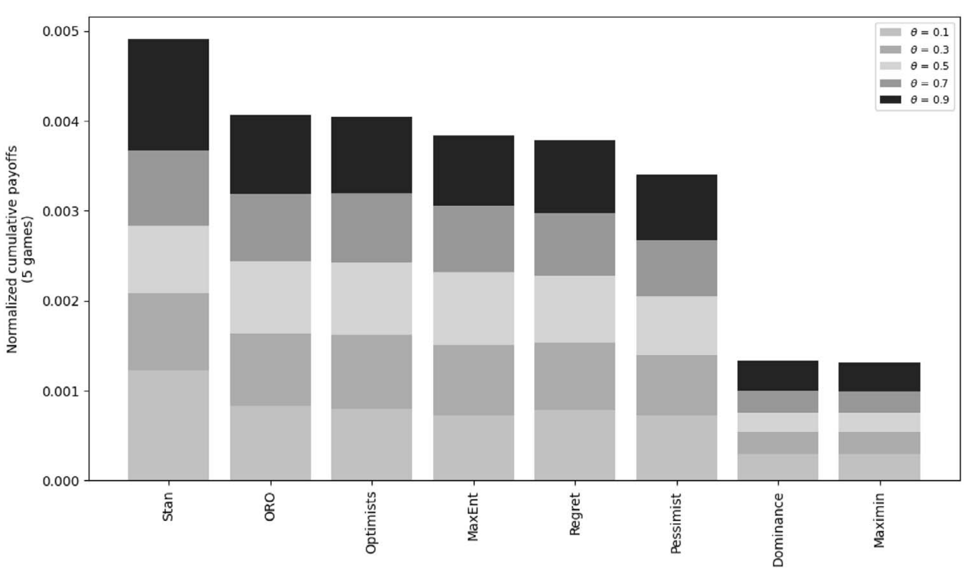


FIGURE 10. Ultra-Minimum Evidence Comparison—five games: Dempster–Shafer players. A stacked bar chart for normalized cumulative payoffs for all players and all coin biases after five games. Bars are scaled in unit interval and ordered from tallest (best—*Stan*) to smallest (worst—*Dominance*).

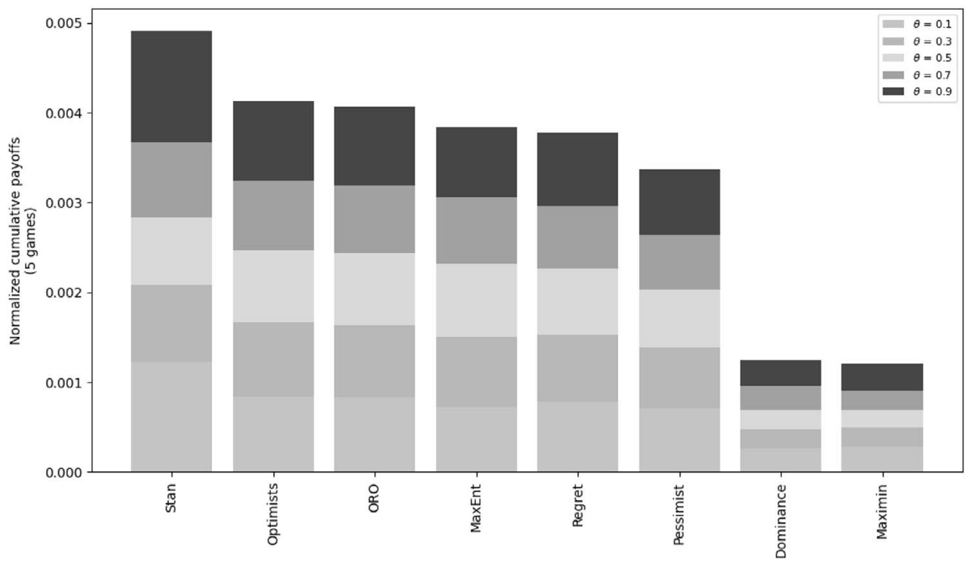


FIGURE 11. Ultra-Minimum Evidence Comparison—five games: Imprecise Bayesians. As Figure 10, but for Imprecise Bayesian players.

distributions rather than a unique distribution. The Imprecise Bayesian players in our model are essentially pairings of this generic Imprecise Bayesian player with decision rules, as we explain in Section 6.

5 Dempster–Shafer theory

Dempster–Shafer theory was developed by Shafer, who expanded and refined earlier ideas from Dempster [19, 69]. The theory has subsequently been investigated using a variety of update methods, interpretations, and other modifications [20, 52]. While Dempster–Shafer theory can be represented by convex sets of credence functions [38] and Dempster initially described his rule of combination in terms of "imprecise probabilities" [19], Shafer was clear that he did not interpret the theory as bounds on sets of admissible Bayesian credences [69, p. ix]. Shafer aimed at a general theory of reasoning under uncertainty, which would contain Standard Bayesianism as a special case, which might sometimes be justified by epistemic or pragmatic reasons [69, p. vii].

Shafer had several aims. For example, he wanted a way of distinguishing between (1) a lack of belief in a hypothesis and (2) a belief that the hypothesis is false [69, pp. 22–23]. Shafer also wanted to avoid how evidence for a conjunction's parts are combined in Bayesianism [69, p. 28].

Dempster–Shafer theory has attracted considerable interest among artificial intelligence researchers, logicians, computer scientists, mathematicians, and others researching uncertain inference. However, compared to Standard or Imprecise Bayesian approaches, formal epistemologists have rarely discussed it, although interest has increased in recent years [8, 24, 28, 36, 47, 48, 61, 70, 71]. Thus, exploring the differences it makes for the Ambiguity Dilemma is an appealing and focused topic for research. In this article, we isolate the difference made by just adopting Dempster–Shafer updating.

In what follows, we begin by explaining the synchronic belief states of the Dempster–Shafer players. We then explain their Dempster conditioning rule. We do so in terms of a generic Dempster–Shafer player template *DS*, who is later paired with particular decision rules to generate Dempster–Shafer players.

Like their Bayesian counterparts, *DS* knows that each coin toss can yield only one of the two outcomes — either ω_h or ω_t. So, as in the Bayesian models, the set of possible decision-relevant states is $\Omega := \{\omega_h, \omega_t\}$. *DS*'s frame of discernment can then be defined as $\mathcal{W} := \mathcal{P}(\Omega) = \{\varnothing, \{\omega_h\}, \{\omega_t\}, \Omega\}$ with a typical element W. *DS*'s initial belief assignment can be represented with a mass function $m : \mathcal{W} \to [0,1]$ that assigns, to each element $W \in \mathcal{W}$, a belief mass $m(W) \in [0,1]$ in such a way that $m(\varnothing) = 0$ and $\sum_{W \subseteq \Omega} m(W) = 1$. Each assignment $m(W)$ represents the amount of initial evidential support allocated to the subset W of Ω, and not to any proper subset of W. In our model, we shall also assume that $m(\{\omega_h\}) > 0$ and $m(\{\omega_t\}) > 0$, in order to represent a situation whether *DS* believes that both ω_h and ω_t are possible outcomes of a coin toss. In addition, we shall assume that $m(\Omega) := 1 - (m(\{\omega_h\}) + m(\{\omega_t\})) > 0$ in order to represent *DS* as initially uncertain about the bias of the coin.

Given the function m, the initial belief function can be defined as $Bel_{\bar{s}} : \mathcal{W} \to [0,1]$, such that, for any $W \in \mathcal{W}$, $Bel_{\bar{s}}(W) := \sum_{V \subseteq W} m(V)$. A belief function $Bel_{\bar{s}}$ is a completely monotone capacity, in the sense that $Bel_{\bar{s}}(\varnothing) = 0$, $Bel_{\bar{s}}(\Omega) = 1$ and $Bel_{\bar{s}}\left(\bigcup_{q=1}^{y \geq 2} V_q\right) \geq \sum_{U \neq \varnothing, U \subseteq \{1,\dots,y\}} (-1)^{n(U)+1} Bel_{\bar{s}}\left(\bigcap_{q \in U} V_q\right)$. The relationship between the belief function $Bel_{\bar{s}}$ and the mass function m is such that, for each $W \in \mathcal{W}$, $m(W) = \sum_{\varnothing \neq V \subseteq W} (-1)^{n(W)-n(V)} Bel_{\bar{s}}(V)$. Notice how, given the outlined assumptions about the mass function m, the assignment of values of function $Bel_{\bar{s}}$ to each $W \in \mathcal{W}$ can be defined as $Bel_{\bar{s}}(W) = m(W)$ for each $W \in \mathcal{W}$.

The initial plausibility function can be defined as $Pl_{\bar{s}} : \mathcal{W} \to [0,1]$ that assigns, to each $W \in \mathcal{W}$, a plausibility value $Pl_{\bar{s}}(W) := \sum_{V \cap W \neq \varnothing} m(V) = 1 - Bel_{\bar{s}}\left(\bar{W}\right)$, where $\bar{W}$ is the complement of W. Thus, given the established relationship between functions $Bel_{\bar{s}}$ and m, function $Pl_{\bar{s}}$ is such that $Pl_{\bar{s}}(\{\omega_h\}) = 1 - m(\{\omega_t\})$ and $Pl_{\bar{s}}(\{\omega_t\}) = 1 - m(\{\omega_h\})$.

A plausibility function provides an upper bound for the evidential support of a hypothesis, while a belief function provides a lower bound. Thus, Dempster–Shafer measures of evidential support can be interpreted as interval-valued. The presence (absence) of larger (smaller) quantities of evidence can be modelled via the width of such intervals, insofar as one grants the assumption that (with respect to a particular hypothesis) more concordance in the total evidence is a proxy (or equivalent to) more evidence.[4]

In order to isolate the particular effect of Dempster–Shafer updating on the Ambiguity Dilemma, we assume that *DS*'s function m is such that $m(\{\omega_h\}) = \frac{1}{100}$, $m(\{\omega_t\}) = \frac{1}{100}$ and $m(\Omega) = \frac{98}{100}$. That is, we assume that *DS*'s belief function $Bel_{\bar{s}}$ and plausibility function $Pl_{\bar{s}}$ generate the same intervals of beliefs on ω_h and ω_t as *IB*'s function μ_P given a history $\bar{s}$ and a convex credal set P ranging from $B(1,99)$ to $B(99,1)$. This isolates the update process as the relevant difference between the two types of player in our agent-based model.

To represent Dempster conditioning, we need to define a belief function Bel_e corresponding to the mass function m_e assigning mass 1 to evidence e. This mass function will define the particular variation of the *DS* player template that we shall use for comparisons against the Imprecise Bayesian players. In our model, each player learns solely from observation histories in S. Thus, the evidence-based mass function will represent a special case of the general *DS* player template who, given any

[4]The "weight of evidence" literature has extensively explored this assumption [11, 18, 22, 23, 35, 49, 55, 63, 69].

history $s \in S$, assigns a mass $m_s(W) \in [0,1]$ to each $W \in \mathcal{W}$ in a similar way as an Imprecise Bayesian player derives the maximum and minimum probability of each state by using the counting function κ to update the extreme beta distributions $B(1,99)$ and $B(99,1)$. That is, given any history $s \in \mathcal{S}$, the mass function m_s is such that

$$m_s(\{\omega_h\}) := \frac{1+\kappa(s)}{100+T}; \tag{11}$$

$$m_s(\{\omega_t\}) := \frac{1+T-\kappa(s)}{100+T}; \tag{12}$$

$$m_s(\Omega) := 1 - (m_s(\{\omega_h\}) + m_s(\{\omega_t\})). \tag{13}$$

As in the case of the initial belief function $Bel_{\bar{s}}$, the evidence-based belief function can be defined as a map $Bel_s : \mathcal{W} \to [0,1]$, such that $Bel_s(W) = m_s(W)$ for each $W \in \mathcal{W}$.

To represent *DS*'s final beliefs derived via Dempster's combination rule, we define a belief function $Bel : \mathcal{W} \to [0,1]$, such that, for any pair $W \in \mathcal{W}$,

$$Bel(W) := Bel_{\bar{s}}(W) \oplus Bel_s(W) = \frac{\sum_{V,Y:V\cap Y=W} m(V)\, m_s(Y)}{1-\sum_{V,Y:V\cap Y=\varnothing} m(V)\, m_s(Y)}. \tag{14}$$

Using the final belief function *Bel*, we can define the final plausibility function as $Pl : \mathcal{W} \to [0,1]$, such that, for any $W \in \mathcal{W}$, $Pl(W) := 1 - Bel(\bar{W})$.

Via Equation (14), we can derive final beliefs in each $W \in \mathcal{W}$ from the Dempster conditioning rule:

$$Bel(\varnothing) = 0. \tag{15}$$

$$Bel(\{\omega_h\}) := \frac{m(\{\omega_h\})\, m_s(\{\omega_h\}) + m(\Omega)\, m_s(\{\omega_h\}) + m(\{\omega_h\})\, m_s(\Omega)}{1-(m(\{\omega_h\})\, m_s(\{\omega_t\}) + m(\{\omega_t\})\, m_s(\{\omega_h\}))}, \tag{16}$$

$$Bel(\{\omega_t\}) := \frac{m(\{\omega_t\})\, m_s(\{\omega_t\}) + m(\Omega)\, m_s(\{\omega_t\}) + m(\{\omega_t\})\, m_s(\Omega)}{1-(m(\{\omega_h\})\, m_s(\{\omega_t\}) + m(\{\omega_t\})\, m_s(\{\omega_h\}))}. \tag{17}$$

$$Bel(\Omega) := \frac{m(\Omega)\, m_s(\Omega)}{1-(m(\{\omega_h\})\, m_s(\{\omega_t\}) + m(\{\omega_t\})\, m_s(\{\omega_h\}))}. \tag{18}$$

A Dempster–Shafer player's final belief interval about each $\omega_i \in \Omega$ can be represented with a function $\rho : \mathcal{W} \times S \to \mathcal{P}([0,1])$ that assigns, to each combination $(W,s) \in \mathcal{W} \times S$, an interval $\rho(W,s) := \left[Bel(W), 1 - Bel(\bar{W})\right]$. Since a Dempster–Shafer player will be betting on events ω_h and ω_t, the decision-relevant intervals are $\rho(\{\omega_h\}, s) := [Bel(\{\omega_h\}), 1 - Bel(\{\omega_t\})]$, $\rho(\{\omega_t\}, s) := [Bel(\{\omega_t\}), 1 - Bel(\{\omega_h\})]$. A useful feature of this interval-based setup is that, for any value $\varrho \in \rho(\{\omega_i\}, s)$, $(1-\varrho) \in \rho(\{\omega_{-i}\}, s)$.

6 Decision rules for *IB* and *DS* Players

Both Imprecise Bayesian and Dempster–Shafer players make decisions with belief intervals, so they require special decision rules that can select among the available actions given interval-valued beliefs. For any history $s \in S$, we can define a set of decision-relevant intervals about the state ω_h as $\Psi_s := \{\mu_P(\omega_h, s), \rho(\{\omega_h\}, s)\}$ with a typical interval $\Psi_s := \left[\psi^{min}, \ldots, \psi^{max}\right]$. Given any interval

$\mathit{\Psi}_s \in \Psi_s$ and any value $\psi \in \mathit{\Psi}_s$, the expected payoff associated with any action $c_j \in C$ can be defined as

$$\mathbb{E}_u\left[c_j \mid \psi\right] := \psi u\left(c_j, \omega_h\right) + (1-\psi)\, u\left(c_j, \omega_t\right). \tag{19}$$

Given any interval $\mathit{\Psi}_s \in \Psi_s$, the minimum payoff expectation from any action $c_j \in C$ can be defined as

$$\mathbb{E}_u^{min}\left[c_j \mid \mathit{\Psi}_s\right] := \min_{\psi \in \mathit{\Psi}_s}\left(\mathbb{E}_u\left[c_j \mid \psi\right]\right), \tag{20}$$

while the maximum payoff expectation can be defined as

$$\mathbb{E}_u^{max}\left[c_j \mid \mathit{\Psi}_s\right] := \max_{\psi \in \mathit{\Psi}_s}\left(\mathbb{E}_u\left[c_j \mid \psi\right]\right). \tag{21}$$

For each decision rule, there are two players who use it: the Imprecise Bayesian player using that rule (e.g. *IB-Dominance*) and the corresponding Dempster–Shafer player using that rule (e.g. *DS-Dominance*).

We excluded some decision rules that we investigated in similar research, since their performances are not notably different from players whom we include. We do not claim that these decision rules are an exhaustive list of possible approaches for making choices using interval-valued beliefs. There are ongoing programmes of developing decision theories for both Imprecise Bayesianism and Dempster–Shafer theory [8, 20]. We have aimed at including popular and historically significant decision rules. We have also avoided decision rules that are still in the early stages of their development and exploration, because we do not want to misinterpret their application to our study's decision problem. We encourage proponents of alternative rules to explore their application in the context of this decision problem, as well as other decision problems that might be used to compare these inductive logics.

As a running example, we shall show how an Imprecise Bayesian player using each rule would make a decision in a particular game. In this example, $\delta = 0.55$. The players have each made 25 observations, 24 of which are heads. The lower bound for ω_h is $(1 + 24)/125 = 0.2$, and the upper bound (plausibility in Dempster–Shafer theory, highest credence in Imprecise Bayesianism) is $(99 + 24)/125 = 0.984$. We provide the expected payoffs given the lower and upper expectations for ω_h in Figures 2 and 3 respectively.

The expected payoff for each action in the figures is the sum for its row. Notice how the expected payoff for c_t is higher than for c_h in Figure 2, but the reverse is true in Figure 3. Therefore, this is an example where the imprecision of beliefs is highly relevant to players' expected payoffs.

6.1 Dominance

The Dominance rule essentially says that any choice is rational unless one believes that its expected payoff is less than that of some alternative action. In the context of our model, a choice c_j is "dominated" if its maximum expected payoff is strictly less than the minimum expected payoff of one or more alternatives.

We define the set of non-dominated actions given any $\mathit{\Psi}_s \in \Psi_s$ as

$$D_{\mathit{\Psi}_s} := \left\{c_j \in C : \mathbb{E}_u^{max}\left[c_j \mid \mathit{\Psi}_s\right] \geq \mathbb{E}_u^{min}\left[c_k \mid \mathit{\Psi}_s\right],\ \text{for every } c_k \in C\right\}. \tag{22}$$

In some cases, the set $D_{\mathit{\Psi}_s}$ is not a singleton, which means that players using the Dominance rule must choose among multiple permissible actions. To resolve this selection problem, we use the uniform randomization tie-breaker, meaning that Dominance players randomly choose an action c_j from a set D_{Ψ_s}. The other players defined below all had a unique permissible action in each game and hence did not need a tie-breaker procedure.

In our running example, no action has a lowest expected payoff that is above the highest expected payoff for any other action. Therefore, a Dominance player would randomize among all three actions.

6.2 Maximin

The next rule is an adaptation of maximin decision theory to interval-valued beliefs [4]. The concept behind this rule is that it is rational to maximize one's actions' minimum expected payoffs [76]. Note that, unlike the classic Maximin rule, the Maximin rule in our tests is dynamic: players following this rule use their updated interval-valued expectations, so the Maximin rule extends beyond situations of complete ambiguity.

Maximin players choose an action $c_j \in C$, such that

$$c_j \in \arg\max_{c_k \in C} \left(\mathbb{E}_u^{min} \left[c_k \mid \Psi_s \right] \right). \tag{23}$$

In our running example, c_a has the highest minimum expected payoff of any possible action, since $0 > -0.35 > -0.434$. Therefore, a Maximin player would abstain.

6.3 Regret

The concept behind the Regret rule (often called the "Minimax Regret" rule) is that it is rational to minimize expected opportunity costs, defined in terms of expected payoffs. Alternatively, its rationale could be put this way: one should minimize regrets, in the technical sense of losses given unavoidable risks, relative to one's information [64]. Regret players calculate a regret number based on their belief intervals and chooses an action with a minimal regret number.

The set of regret-minimizing choices can be defined with a regret function $\zeta_{\Psi_s} : C \to \mathbb{R}$ that assigns, to each action $c_j \in C$, a regret value

$$\zeta_{\Psi_s}\left(c_j\right) := \max_{\psi \in \Psi_s} \left(\max_{c_k \in C} \left(\mathbb{E}_u \left[c_k \mid \psi \right] \right) - \mathbb{E}_u \left[c_j \mid \psi \right] \right). \tag{24}$$

Regret players choose an action $c_j \in C$, such that

$$c_j \in \arg\min_{c_k \in C} \left(\zeta_{\Psi_s}\left(c_k\right) \right). \tag{25}$$

In our running example, the regret values for c_h, c_t and c_a are respectively 0.7, 0.868, and 0.434. Therefore, a Regret player would choose c_a.

6.4 Optimist

The next rule is defined in terms of the Hurwicz criterion approach. There is a constant parameter $\alpha \in [0, 1]$, representing the degree to which a player is cautious [29]. As α is closer to 1, the player puts more weight on lower expected payoffs. As α is closer to 0, the player puts more weight on higher expected payoffs.

The players using the Hurwicz criterion in our model have an α value of 0.25, thus assigning a weight of 0.25 to actions' minimum expected payoffs and 0.75 to their maximum expected payoffs. Because of the distribution of weights that puts more emphasis on expected outcomes of actions, this rule is called the Optimist rule. A player using this rule chooses an action $c_j \in C$, such that

$$c_j \in \arg\max_{c_k \in C} \left(\frac{1}{4} \mathbb{E}_u^{min} \left[c_k \mid \Psi_s \right] + \frac{3}{4} \mathbb{E}_u^{max} \left[c_k \mid \Psi_s \right] \right). \tag{26}$$

In our running example, the weighted expected payoff of c_h is 0.238. The weighted expected payoff of c_t is 0.154. Finally, the weighted expected payoff of c_a is 0. Therefore, Optimist players will choose c_h.

6.5 *Pessimist*

This player type is very similar to the Optimists. The only difference is that they are more cautious: $\alpha = 0.75$ rather than $\alpha = 0.25$. A player using this rule is otherwise identical to an Optimist player.

In our running example, the weighted expected payoff of c_h is -0.154. The weighted expected payoff of c_t is -0.238. Finally, the weighted expected payoff of c_a is 0. Therefore, Pessimist players will choose c_a.

6.6 *Opportunity Risk Optimization*

This decision rule is an adaptation of earlier suggestions by Daniel Ellsberg [21, p. 664] into a rule for imprecise probabilities [58]. The concept is that a rational agent tries to optimize a balance between using their expected payoffs and minimizing their risks of losses, but discounts the former expected payoffs by the extent to which their evidence is ambiguous. The Opportunity Risk Optimization (ORO) rule is based on a method for weighting an action c_j by its average expected payoff, given a point-estimation method and the action's minimum expected payoff. The weighting of the first factor increases with their sample sizes, as formalised below.

The average expected payoff of any action $c_j \in C$, given any $\Psi_s \in \Psi_s$ can be defined as

$$\mathbb{E}_u^{avg} \left[c_j \mid \Psi_s \right] := \psi^{avg} u \left(c_j, \omega_h \right) + \left(1 - \psi^{avg} \right) u \left(c_j, \omega_t \right), \text{ where } \psi^{avg} := \frac{\psi^{max} + \psi^{min}}{2}. \tag{27}$$

We then define the ORO rule as choosing an action satisfying the following requirement:

$$c_j \in \arg\max_{c_k \in C} \left(\varpi \left(\Psi_s \right) \mathbb{E}_u^{min} \left[c_k \mid \Psi_s \right] + \left(1 - \varpi \left(\Psi_s \right) \right) \mathbb{E}_u^{avg} \left[c_k \mid \Psi_s \right] \right),$$

$$\text{where } \varpi \left(\Psi_s \right) := \frac{\psi^{max} - \psi^{min}}{\frac{99}{100} - \frac{1}{100}} = \frac{50 \left(\psi^{max} - \psi^{min} \right)}{49}. \tag{28}$$

Like the Optimist rule, this rule essentially requires maximizing an auxiliary quantity, generated from the expected payoff intervals.

In our running example, the minimum expected payoffs for c_h, c_t and c_a are respectively –0.35, –0.434, and 0. The average expected payoffs are respectively 0.042, -0.042 and 0. Finally, $\varpi (\Psi_s) = 0.8$. Substituting into Equation (28), we find values of –0.2716, –0.3556, and 0 respectively. Therefore, ORO players would choose c_a.

6.7 *MaxEnt*

The MaxEnt rule is an adaptation of the Maximum Entropy Principle [30, 78, 81] to the context of interval-valued beliefs. The idea is that one should make decisions using the minimally informative probability distribution that is compatible with one's belief intervals. Note that this probability distribution is interpreted as a tool for decision-making, but not necessarily as credences. For example, an Imprecise Bayesian using the MaxEnt rule would regard the probabilities as auxiliary quantities for decision-making, rather than as Standard Bayesian credences.

An entropy-maximizing probability function for a given belief interval is defined as $\eta : \Theta \to \Delta^{+}(\Theta)$ which assigns a uniform probability distribution on the set Θ and a belief function $\pi : \Psi_s \to \mathcal{P}([0,1])$ that assigns, to any interval $\varPsi_s \in \Psi_s$, a set of entropy-maximizing beliefs $\pi(\varPsi_s) \subseteq \varPsi_s$, where each belief $\psi \in \pi(\varPsi_s)$ is such that

$$\psi \in \arg\min_{\psi' \in \varPsi_s} \left(\left| \psi' - \int_0^1 \theta \eta(\theta)\, d\theta \right| \right). \tag{29}$$

Since $\int_0^1 \theta\eta(\theta)\, d\theta = \frac{1}{2}$, Equation (21) can be simplified to

$$\psi \in \arg\min_{\psi' \in \varPsi_s} \left(\left| \psi' - \frac{1}{2} \right| \right). \tag{30}$$

The MaxEnt rule requires choosing an action $c_j \in C$, such that

$$c_j \in \arg\max_{c_k \in C} \left(\mathbb{E}_u \left[c_k \mid \psi \right] \right) \text{ with some } \psi \in \pi(\varPsi_s). \tag{31}$$

In our running example, there is an entropy-maximizing function assigning probabilities of 0.5 to both heads and tails. For this function, the expected payoffs for c_h, c_t and c_a are respectively -0.05, 0.05, and 0. Therefore, a MaxEnt player would choose c_t. Note the contrast with all other players so far, who variously randomise, abstain, or choose c_h. This shows how different decision rules can meaningfully affect players' choices.

7 Results

7.1 Main comparisons

The main overall result from our tests was that the Imprecise Bayesian and Dempster–Shafer players performed approximately identically. Therefore, there was no setting (decision rule and coin bias) such that a Dempster–Shafer player's updating rule made a positive difference to performance. Consequently, adopting this version of Dempster–Shafer updating does not avoid the Ambiguity Dilemma.

We shall proceed in three steps. First, we shall use our running example to explain how the beliefs of Imprecise Bayesian and Dempster–Shafer players can diverge. Second, we shall explain the result in general terms. Third, we explain the extent to which different decision rules affected the results.

In our running example, there are 25 coin toss observations, 24 of which landed heads and 1 of which landed tails. For Dempster–Shafer players, the initial belief assignment can be represented with a mass function $m_{\bar{s}}$, such that $m_{\bar{s}}(\{\omega_h\}) = 0.01$, $m_{\bar{s}}(\{\omega_t\}) = 0.01$, and $m_{\bar{s}}(\varOmega) = 0.98$.

The beliefs that Dempster–Shafer players hold in the running example can be represented with an evidence-based belief function m_s, such that $m_s(\{\omega_h\}) = 0.2$, $m_s(\{\omega_t\}) = 0.016$ and $m_s(\varOmega) = 0.784$.

Both mass functions assign positive mass to $\varOmega$, which means that, once the Dempster combination rule is applied, a Dempster–Shafer player's final beliefs will assign a positive value to $\varOmega$. This can be checked by applying the rule in order to derive the values for the Dempster–Shafer player's final belief and plausibility functions:

1. $Bel(\{\omega_h\}) \approx 0.206286$, $Pl(\{\omega_h\}) \approx 0.976269$,
2. $Bel(\{\omega_t\}) \approx 0.023731$, $Pl(\{\omega_t\}) \approx 0.793714$.

In contrast, at the same point in the tests, an Imprecise Bayesian player will have credences in their set of 0.2 to 0.984 for ω_h and from 0.016 to 0.8 for ω_t. This result is interesting, because it shows how Dempster–Shafer theory and Imprecise Bayesianism can diverge, given analogous initial beliefs and the same data, even in a decision problem based around the relatively simple case of choosing how and whether to bet given a sequence of binomial trials.

Notice how the Dempster–Shafer players have narrower intervals in the running example. This feature occurs throughout our tests. In previous research, we found that the slow convergence of Imprecise Bayesian players caused their underperformance relative to *Stan*. However, while the gap between belief and plausibility closed slightly faster than the corresponding values for Imprecise Bayesians, Dempster–Shafer players do not perform detectably better and do not avoid the Ambiguity Dilemma.

This result is surprising, because updating via Dempster's rule is fundamentally different from Imprecise Bayesianism. The problem is that, while Dempster's rule enables faster convergence towards the cumulative sample frequency, it still can only represent ambiguity via an initial belief state that requires relatively large samples to revise. Conversely, as we discuss below, *Stan* updates more rapidly, at the cost of not having the representational tools possessed by Dempster–Shafer and Imprecise Bayesian players.

For both types of player, *IB-Optimist* and *DS-Optimist* were the best performers. Similarly, for both types of player, *IB-Maximin* and *DS-Maximin* were the worst performers. The graphs of their performances are shown in Figures 4 and 5. The relatively poor performance of Maximin players was detectable with all biases. The reason was that abstaining often has the maximum minimum expected payoff in early games, when intervals are wide, because (depending on the ticket price) it can be the case that both betting on heads and betting on tails have negative expected payoffs. In contrast, abstaining has a guaranteed payoff of 0. However, abstaining is a costly action (in terms of opportunity costs) when there is useful sample data, because it forgoes potential profits that some other players can reliably attain.

Players using the Dominance rule performed very similarly to players using the Maximin rule. This occurred because the Dominance rule often requires randomizing among c_h, c_t and c_a if the belief interval is wide, since none of these actions is dominated for a given ticket price unless it is superior given any point in the belief interval. In these randomizations, there is a 1/3 chance of choosing abstaining, and thereby missing potential profits. Such situations occur more often insofar as convergence is slower.

The performances given the other decision rules were in between these two extremes of the Optimist and the Maximin rules. The ordering of decision rule performances was the same for both types of player.

Players using the ORO, Pessimist and Regret rules performed slightly worse than those using the Optimist decision rule, except when the coin bias was 0.5. The performance gap was greatest when the coin bias was 0.1 or 0.9. Under those conditions, samples of Bernoulli trials provide more reliable information, but these rules make less use of this information than the rules for *IB-Optimist* and *DS-Optimist*. There was no statistically significant difference in the performance of the ORO, Pessimist and Regret rules. Meanwhile, MaxEnt players' performance was very sensitive to the coin bias: when the bias was 0.5, they did as well as any other player; when the coin bias was 0.1 or 0.9, they did poorly; when the coin bias was 0.3 or 0.7, they also lagged behind many other players. These problems for MaxEnt players were caused by a slow convergence towards the true bias. However, regardless of the decision rule, Dempster–Shafer players performed approximately identically to the corresponding Imprecise Bayesian player.

7.2 Ultra-minimum evidence

Our Ultra-Minimum Evidence comparisons in Figures 10 and 11 provide a crucial piece of information, by focusing in precisely on early games. They show that even in the very short run, there is not a significant advantage from using Dempster–Shafer updating, nor disadvantage. This is because, while the Dempster–Shafer belief intervals narrow quicker than the corresponding Imprecise Bayesian values, the difference is too marginal to create a substantial effect.

7.3 Caution Check

For our Caution Check test, the performances among the two types of player were almost identical—see Figure 8 and 9. Firstly, the rankings of the performances of decision rules were the same for both types of player. The Regret and Pessimist players had the lowest frequencies of net losses among the Imprecise Bayesian and Dempster–Shafer players. Both were close to *Stan*. The ORO and Optimist players were not far behind. MaxEnt and Maximin then followed. The Dominance players were the worst performing players on the Caution Check test, by a very large margin.

It may seem surprising that being more cautious about using information from early samples would increase the rate of net losses. However, while the early sample frequencies are more likely to diverge from the true coin bias, they do provide usable information for making decisions. Meanwhile, avoiding *net* losses in this decision problem will generally require making sufficient profits as well as avoiding losses in games, because some losses are inevitable (except by forgoing profits very frequently) due to the stochastic unpredictability of the coin tosses.

Why did Dominance players perform so badly on the Caution Check? Recall that the Dominance rule with a randomizing tiebreaker requires randomizing among c_h, c_t and c_a whenever none of these actions is dominated for the given belief interval and ticket price. For instance, suppose that the coin bias is 0.9 (towards heads). Until a Dominance player's belief interval has narrowed sufficiently, they have a 1/3 chance of selecting c_t. Hence, given the 0.9 coin bias, a Dominance player randomizing in this way has a 1/3 chance of a loss. Consequently, Dominance players not only sometimes miss profit opportunities by abstaining, but also systematically have a greater chance of making losses by betting on the wrong outcome in the early games.

The Dempster–Shafer players received no detectable advantage from their updating method in the Caution Check. This reinforces our general point that the Ambiguity Dilemma is robust against the use of Dempster's rule to update, ceteris paribus.

7.4 Standard Bayesian

As noted before, we are not providing a holistic comparison of Standard Bayesianism and Dempster–Shafer theory. There are many alleged advantages and disadvantages of both approaches, but these are beyond our scope. However, for this decision problem, we can definitely state that the Dempster–Shafer players do not improve on the performance of the Imprecise Bayesians with respect to comparative performance against *Stan*.

Interestingly, the reasons for this performance gap are fundamentally the same for both Dempster–Shafer players and Imprecise Bayesian players. These approaches involve slower convergence, but without a compensating gain in terms of the Caution Check. Moreover, this slower convergence is caused by the use of the tools by which these approaches can represent ambiguity in this decision problem.

Stan's flat prior means that their credences converge especially rapidly towards the sample frequency. At the same time, since their initial prior is equivocal between heads and tails occurring in the first game, they have a cautious element that causes their prior to be robust against different coin biases.[5] This result was also corroborated by the graphs for *Stan*, who quickly attained the long-run rate of profit.

For the Caution Check, we found that *Stan*'s performance could be approximately matched by some decision rules with either Imprecise Bayesian or Dempster–Shafer updating. However, it is notable that *Stan*'s profit making success in this problem does not come at the cost of greater frequencies of net losses.

Stan's performance also reinforces the importance of even the marginal differences in the Caution Check among Dominance and Maximin players, as well as the more general differences in favour of the Imprecise Bayesian players. In a very large proportion of decision problems, comparative performance is crucial to success. The differences can be marginal in size but very substantial in importance. Think of the science team that achieves priority in a discovery, the business that achieves the best comparative profits or losses against comparable investments, the military that achieves victory in a battle; even if these comparative successes are marginal, they can be the difference between fame or obscurity; fortune or bankruptcy; survival or death. Therefore, even when the gains from one learning method over another are numerically marginal, they can be practically significant.

8 Discussion

Since our results are stable and reliable, there are several issues to discuss. We shall separately discuss, in turn, (1) issues directly concerning the Ambiguity Dilemma, (2) issues relating to the representation of neutral belief states, (3) issues concerning different rules for making decisions given the use of Dempster–Shafer theory, and (4) significant limitations of our study.

8.1 The Ambiguity Dilemma

Both types of player start out with comparable representations of ambiguity, but with different results. The combination rule offers an interesting alternative to Imprecise Bayesian updating, but not one that avoids the Ambiguity Dilemma.

Of course, the Dempster–Shafer theory formalism allows for players with more convergent initial belief states. The same is true of Imprecise Bayesianism. Yet this choice would make their initial beliefs less representative of the ambiguity. Already, the degree of convergence in the credal set and belief functions goes beyond the initial information provided about the coin bias. Further restrictions to create a narrower range of beta distributions would aggravate this lack of ambiguity representation. Conversely, a broader initial range would aggravate the comparative decision-making performance problems relative to *Stan*.

Recall that the Ambiguity Dilemma is a trade-off, not a simple flaw, for Imprecise Bayesians and (as we have now shown) Dempster–Shafer theory. There are advantages to these approaches in terms of a rich and flexible method of representing evidence and knowledge. Choosing between these

[5] *Stan* does better compared to other players given “extreme” biases of 0.1 or 0.9, because the average variance for these biases is lower. Similarly, to a lesser extent, they do better with a 0.3 or 0.7 bias than a 0.5 bias.

advantages and the disadvantages of relatively diminished performance is presumably a contextual matter. For applications in formal epistemology, inductive logic, and the philosophy of science, one might choose to opt for the tools that Imprecise Bayesianism and Dempster–Shafer theory offer for representing ambiguity. Conversely, given decision problems that are relevantly comparable to that studied in this article, in applications such as financial machine learning or applied artificial intelligence in imaging software, one might prefer the decision-making performance of Standard Bayesian approaches such as *Stan*. In intermediate contexts, there is a vast range of possible alternatives, such as Dempster–Shafer reasoning with a belief function that is closer to *Stan*'s credence function or Imprecise Bayesianism with a more convergent credal set.

Previously, the Ambiguity Dilemma has been discussed just in relation to Imprecise Bayesianism. However, due to Dempster–Shafer theory's adaptability, it would be interesting to explore how it might have hitherto unforeseen benefits. What we have shown is that just adopting Dempster's combination rule is insufficient to avoid the Ambiguity Dilemma.

8.2 Neutrality and performance

In this subsection, we discuss a particular issue concerning the Ambiguity Dilemma and its scope. As mentioned in Section 2, one influential argument for Dempster–Shafer theory is that a method for representing neutral belief states seems to be a desideratum for an inductive logic, yet the Standard Bayesian tools for this task (such as the Principle of Indifference) are infamously contested.

In our decision problem, it is possible to determine a flat prior that is "neutral" with respect to the events of practical interest. Thus, *Stan*'s initial belief state is equivocal with respect to the coin tosses landing heads or tails. Yet this neutrality is not always possible. Consider the following modification to our decision problem: players are offered the opportunity to bet on the coin toss at the end of a game *and* the chance to guess the overall relative frequency of heads in the 5000 tosses in the test. *Stan* would have a strong initial belief that the relative frequency of heads will be close to 2500/5000. Furthermore, regarding decision-making performance, if there was a coin bias that was not 0.5, then *Stan* would tend to make early losses insofar as they could bet on their initial belief in the overall relative frequency. In contrast, erstwhile poor performers like the Maximin players would benefit because of their great tendency to abstain until their samples became sufficiently large for reliable estimates.

Hence, the Ambiguity Dilemma is more complex than simply formal epistemological advantage versus decision-theoretic advantage. Nonetheless, these considerations do not affect our result, which is that *for this specific decision problem*, Dempster–Shafer updating does not avoid the Ambiguity Dilemma.

Moreover, note that the Dempster–Shafer players (and Imprecise Bayesian players) also do not start with a strictly neutral belief state. The sets of beta distributions that they use for their initial belief intervals also imply strong beliefs for some hypotheses. Therefore, in some decision problems, the same challenges mentioned for *Stan* will occur for them. For instance, the beta distributions ranging from $B(1, 99)$ to $B(99, 1)$ impose strong constraints on an Imprecise Bayesian's beliefs about some higher-order events consisting of coin tosses. These credence functions all assign some outcomes (such as 5000 out of 5000 heads) credences close to zero. In general, one could engineer tests with coin biases and betting options such that the Imprecise Bayesian and Dempster–Shafer players would also suffer from being overly confident in their initial beliefs. Therefore, when assessing decision-making performances using agent-based modelling, one should explore test settings that are fair, plausible and connected to well-explored decision problems, rather than gerrymandered tests to "catch out" particular types of player.

Meanwhile, pure agnosticism—being neutral with respect to all beta distributions—would result in "inertia," where players fail to update [34, 67, 74, 77, 79]. In this decision problem, inertia would be very costly in terms of the missed potential profits. A formal system that represents neutrality, while avoiding inertia and other problems, remains a continuing quest for inductive logic [3, 50]. However, we stress that our results do not challenge the progress that Dempster–Shafer theory has made in representing ambiguity.

8.3 Rules and performance

Suppose that, due to one of the many arguments for Dempster–Shafer theory, someone favours this approach to updating in this type of decision problem. Given that choice, our results show the advantages of using a decision rule that is less likely to leave the choice of actions undetermined. For instance, such a reasoner would be better off if they followed *Optimist-DS* rather than *Dominance-DS* and *Maximin-DS*.

Tiebreaking is a conceptually controversial operation. Should one randomise? Or use a supplementary rule? Decision theorists have explored a variety of tiebreaker approaches in the context of imprecise probabilities [45, Chapters 6 and 7]. In situations of ambiguity, the choice of tiebreaker rule can be significant. Thus, a decision rule that avoids tiebreaking can also avoid some thorny conceptual issues, as well as the additional computational time employing a tiebreaker procedure, such as randomization.

To illustrate how tiebreaking makes a difference, we provide Figure 7. *DS-Dominance* is the player described in Subsection 6.1. The player we call *DS-Dominance+Ultra-Optimist* uses the Hurwicz criterion with $\alpha = 0$, instead of randomizing, given multiple undominated actions in a game. This α value means that, in a sense, *DS-Dominance+Ultra-Optimist* is even more "optimistic" than Optimist. Note how *DS-Dominance+Ultra-Optimist* does much better in this particular decision problem. On the other hand, there is nothing in the Dominance rule as such that entails either tiebreaking procedure [39, Chapter 14].

8.4 Limitations

We have strived to maintain comparability with other studies that use this decision problem to compare inductive logics [40, 57–59]. Our study inherits some of the limitations of these studies. Most notably, our decision problem is based around a series of Bernoulli trials. Obviously, not all decision problems concern such events, and rapid convergence is not always an advantage.

However, our results do *not* merely show that convergence is advantageous when it is advantageous so *Stan* benefits from convergence when it is beneficial. What our results principally show is that the slower convergence rates of Dempster–Shafer and Imprecise Bayesian players matters, while Dempster–Shafer reasoning does not significantly improve in itself on Imprecise Bayesian performances.

In addition, there are many decision problems where it is a reasonable to assume that the trials are exchangeable. It is also notable that the sorts of distributions that would harm the relative performance of *Stan* are not straightforward to identify, because the beliefs of the other players we studied also converge towards the sample frequency. Departures from the exchangeability assumption would have to be reflected in the Standard Bayesian player's choice of prior, to reproduce the type of ambiguity that we study here. Given a non-exchangeable problem, many Standard Bayesian statisticians would argue for different types of prior. Problems with player uncertainty about the structural features of the model (such as the outcome space or the stability of the stochastic parameters) would require very different analyses; these are interesting potential areas of strength for Imprecise Bayesianism and Dempster–Shafer theory, but topics for another day.

In this article, our focus has been on how Dempster–Shafer theory and Imprecise Bayesianism compare with respect to the Ambiguity Dilemma. Hence, we have used a range of decision rules primarily as a control method, in order to ensure that our main inductive logic comparison is not an artefact of any particular decision rule. A consequence of this focus is that we cannot rule out that other decision rules might have better results when paired with Dempster–Shafer (or Imprecise Bayesian) learning.

An additional limitation is that our results only show that there are *some* priors with which a Standard Bayesian can outperform Dempster–Shafer players. Yet these priors are not required by Standard Bayesian inductive logic. Many Bayesians are subjectivists and think that the choice of prior is largely unconstrained. Previous research has shown that biased priors can seriously affect the short run performance of Standard Bayesian players in this decision problem, even though "washing out" convergence results hold in the long run [57]. Since there are indefinitely many Standard Bayesian players who would be outperformed by the Dempster–Shafer players, our results do not indicate a general superiority of the former approach, even for the specific criteria we have used in this article.

Furthermore, our study has examined decision-theoretic performance, but Dempster–Shafer theory might be interpreted more as a method for combining and quantifying evidence. For instance, Judea Pearl has argued that Dempster–Shafer theory formalizes "provability" [52–54]. On the other hand, the Ambiguity Dilemma occurs regardless of the decision rule, and it is due to differences in how inductive logics extract reliable information from sample data. Thus, while the Ambiguity Dilemma as such does not hold with respect to the application of Dempster–Shafer theory to aims such as "provability" formalization, it does raise some inductive logical issues that Dempster–Shafer theorists might want to examine further.

In this article, we have focused on statistical decisions, in that players' decisions are informed by estimates of the probabilities of a stochastic process — the coin tosses. Imprecise Bayesianism and Dempster–Shafer theory have also been investigated in contexts such as policymaking and medical decisions, where causal reasoning is important [6, 10, 75]. While the type of decision-making we consider here does not require causal knowledge, many decisions based on causal reasoning involve stochastic knowledge (such as estimating causal effects [25, 26, 32]) and thus our article provides a useful first step towards comparisons in causal reasoning decision problems. There are also types of reasoning, such as hypothetical and counterfactual reasoning, where the players could be compared using our methods but which typically involve formalisms and controversies that are well beyond this article's scope [5, 43].

In sum, our results have the limitations that come from isolating a particular factor (Dempster's combination rule or Imprecise Bayesian conditioning) and investigating performances in one type of decision problem, but also the corresponding advantages of precision and identifying the relevant factors. One must be wary when extrapolating the Ambiguity Dilemma from this article's Dempster–Shafer players' performances to other decision problems or to alternative Dempster–Shafer players. On the other hand, we were able to isolate the role of the update method in players' performances. Consequently, we provide a solid basis for further research that will either warrant or reject extrapolations to alternative decision problems.

9 Conclusion

We have shown that simply adopting Dempster's combination rule does not achieve a better trade-off in the Ambiguity Dilemma. For this reason, we have compared *one* possible way that Dempster–Shafer theory could be applied to the decision problem. For example, there is nothing in Dempster–

Shafer theory *as such* that excludes learning by a method that is effectively equivalent to *Stan*. Yet note that such approaches would also be forgoing the ambiguity representation tools of the Dempster–Shafer players in our study. Therefore, they would not escape the Ambiguity Dilemma.

Dempster–Shafer theory and Imprecise Bayesianism are the most prominent approaches to reasoning using interval-valued beliefs and imprecise probabilities, but they are not alone [2, 9, 15, 35, 41]. Further examination of the short-run decision-making performance of these alternatives will enable our results to be placed in a broader context.

Perhaps there is some version of Dempster–Shafer theory that offers a trade-off in the Ambiguity Dilemma that is superior to Imprecise Bayesianism. For instance, the use of beta distributions is common but not required. Dempster–Shafer decision theory continues to evolve [20] while organizations like the Belief Functions and Applications Society provide annual conferences for the development and exploration of Dempster–Shafer belief functions. Our article has been the first investigation of the Dempster–Shafer theory in relation to the Ambiguity Dilemma, but it should not be the last. We hope that our results will serve as a stimulus for Dempster–Shafer theorists to explore their flexible formalism's capacity to mitigate or perhaps even avoid the Ambiguity Dilemma.

Acknowledgements

The work described in this article was partly supported by the 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. The Authors are grateful to the Johannes Kepler University Linz for providing open access for this article.

References

[1] F. Bacchus, H. E. Kyburg and M. Thalos. Against conditionalization. *Synthese*, **85**, 475–506, 1990.

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

[3] Y. Benétreau-Dupin. The Bayesian who knew too much. *Synthese*, **192**, 1527–1542, 2015.

[4] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 1985.

[5] F. Berto. Logic will get you from A to B, imagination will take you anywhere. *Noûs*, **58**, 717–729, 2023.

[6] H. Blau. Ploxoma: Testbed for uncertain inference. In *Learning from Data V, Volume 112 of Lecture Notes in Statistics,* D. Fisher and H. J. Lenz, eds, pp. 47–57. Springer-Verlag, New York, USA, 1996.

[7] R. Bradley. *Decision Theory with a Human Face*. Cambridge University Press, Cambridge, UK, 2017.

[8] S. Bradley. Imprecise probabilities, Accessed: 2023-05-01. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, ed., edn. Metaphysics Research Lab, Stanford University, Stanford CA, USA, spring, 2019.

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

[10] R. Bradley and K. Steele. Making climate decisions. *Philosophy Compass*, **10**, 799–810, 2015.

[11] M. E. Brady. J. M. Keynes's theoretical approach to decision-making under conditions of risk and uncertainty. *The British Journal for the Philosophy of Science*, **44**, 357–376, 1993.
[12] R. Carnap. *The Continuum of Inductive Methods*. The University of Chicago Press, Chicago IL, USA, 1952.
[13] R. Carnap. *The Logical Foundations of Probability*. The University of Chicago Press, Chicago IL, USA, 1962.
[14] W. F. Caselton and W. Luo. Decision making with imprecise probabilities: Dempster—Shafer theory and application. *Water Resources Research*, **28**, 3071–3083, 1992.
[15] M. E. Cattaneo and A. Wiencierz. Likelihood-based imprecise regression. *International Journal of Approximate Reasoning*, **53**, 1137–1154, 2012.
[16] M. Colombo, L. Elkin and S. Hartmann. Being realist about Bayes, and the predictive processing theory of mind. *The British Journal for the Philosophy of Science*, **72**, 185–220, 2021.
[17] D. R. Cox. *Principles of Statistical Inference*. Cambridge University Press, Cambridge, UK, 2006.
[18] B. Davidson and R. Pargetter. Guilt beyond reasonable doubt. *Australasian Journal of Philosophy*, **65**, 182–187, 1987.
[19] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, **38**, 325–339, 1967.
[20] T. Denœux. Decision-making with belief functions: a review. *International Journal of Approximate Reasoning*, **109**, 87–110, 2019.
[21] D. Ellsberg. Risk, Ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, **75**, 643–669, 1961.
[22] J. Franklin. Resurrecting logical probability. *Erkenntnis*, **55**, 277–305, 2001.
[23] J. Franklin. Case comment—United States v. Copeland, 369 f. supp. 2d 275 (e.d.n.y. 2005): Quantification of the 'proof beyond reasonable doubt' standard. *Law, Probability and Risk*, **5**, 159–165, 2006.
[24] K. Genin. Formal representations of belief. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, ed., fall 2022 edn. Accessed: 24-01-24. Metaphysics Research Lab, Stanford University, Stanford CA, USA, 2022.
[25] M. A. Hernán and J. M. Robins. *Causal Inference: What if*. Chapman & Hall/CRC, Boca Raton FL, USA, 2020.
[26] P. W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, **81**, 945–960, 1986.
[27] C. Howson and P. Urbach. *Scientific Reasoning: The Bayesian Approach*. Open Court Publishing, Chicago IL, USA, 2006.
[28] F. Huber. For true conditionalizers Weisberg's paradox is a false alarm. *Symposion*, **1**, 111–119, 2014.
[29] L. Hurwicz. The generalised Bayes-minimax principle: a criterion for decision-making under uncertainty. *Cowles Commission Discussion Paper: Statistics*, **335**, 1–7, 1951.
[30] E. T. Jaynes. Information Theory and Statistical Mechanics. I. *Physical Review*, **106**, 620–630, 1957.
[31] H. Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **186**, 453–461, 1946.
[32] J. M. Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge, UK, 1999.

[33] J. M. Joyce. How probabilities reflect evidence. *Philosophical Perspectives*, **19**, 153–178, 2005.
[34] J. M. Joyce. A defense of imprecise credences in inference and decision making. *Philosophical Perspectives*, **24**, 281–323, 2010.
[35] J. M. Keynes. *A Treatise on Probability*. Macmillan and Co., London, UK, 1921.
[36] D. Klein, O. Majer and S. Rafiee Rad. Non-classical probabilities for decision making in situations of uncertainty. *Roczniki Filozoficzne*, **68**, 315–343, 2021.
[37] H. E. Kyburg. Bets and beliefs. *American Philosophical Quarterly*, **5**, 54–63, 1968.
[38] H. E. Kyburg. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, **31**, 271–293, 1987.
[39] H. E. Kyburg. *Science and Reason*. Oxford University Press, Oxford, UK, 1990.
[40] H. E. Kyburg and C. M. Teng. Choosing among interpretations of probability. In: *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, UAI'99*, K. B. Laskey and H. Prade, eds, pp. 359–365, San Francisco CA, Morgan Kaufmann Publishers Inc., 1999.
[41] H. E. Kyburg and C. M. Teng. *Uncertain Inference*. Cambridge University Press, Cambridge, UK, 2001.
[42] P. S. Laplace. *Théorie Analytique Des probabilités*. Courcier, Imprimeur-Libraire pour les Mathématiques, rue du Jardinet, France, 1820.
[43] H. Leitgeb. Imaging all the people. *Episteme*, **14**, 463–479, 2016.
[44] I. Levi. On indeterminate probabilities. *The Journal of Philosophy*, **71**, 391–418, 1974.
[45] I. Levi. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, Cambridge MA, USA, 1980.
[46] R. Martin. False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, **113**, 39–73, 2019.
[47] R. Meester and T. Kerkvliet. The infinite epistemic regress problem has no unique solution. *Synthese*, **198**, 4973–4983, 2021.
[48] J. C. Mork. Uncertainty, credal sets and second order probability. *Synthese*, **190**, 353–378, 2013.
[49] D. A. Nance. *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*. Cambridge University Press, Cambridge, UK, 2016.
[50] J. D. Norton. Ignorance and indifference. *Philosophy of Science*, **75**, 45–68, 2008.
[51] J. D. Norton. Challenges to Bayesian confirmation theory. In: *Philosophy of Statistics, Volume 7 of Handbook of the Philosophy of Science*, P. S. Bandyopadhyay and M. R. Forster, eds, pp. 391–439. Elsevier, Oxford, UK, 2011.
[52] J. Pearl. On probability intervals. *International Journal of Approximate Reasoning*, **2**, 211–216, 1988a.
[53] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco CA, USA, 1988b.
[54] J. Pearl. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, **4**, 363–389, 1990.
[55] W. Peden. Imprecise probability and the measurement of Keynes's weight of arguments. *Journal of Applied Logics—IFCoLog Journal of Logics and their Applications*, **5**, 677–708, 2018.
[56] K. R. Popper. A third note on degree of corroboration or confirmation. *The British Journal for the Philosophy of Science*, **8**, 294–302, 1958.
[57] M. Radzvilas, W. Peden and F. De Pretis. A Battle in the Statistics Wars: a simulation-based comparison of Bayesian, Frequentist and Williamsonian methodologies. *Synthese*, **199**, 13689–13748, 2021.

[58] M. Radzvilas, W. Peden and F. De Pretis. Making decisions with evidential probability and objective Bayesian calibration inductive logics. *International Journal of Approximate Reasoning*, **162**, 1–37, 2023.

[59] M. Radzvilas, W. Peden and F. De Pretis. The ambiguity dilemma for imprecise Bayesians. *The British Journal for the Philosophy of Science*, 1–41, 2024. https://doi.org/10.1086/729618.

[60] J. Reiss. What's wrong with our theories of evidence? *Theoria: An international Journal for Theory History and Foundations of Science*, **29**, 283–306, 2014.

[61] J.-W. Romeijn. Conditioning and interpretation shifts. *Studia Logica*, **100**, 583–606, 2012.

[62] R. D. Rosenkrantz. *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*, vol. **115**. D. Reidel, Dordrecht, Netherlands, 1977.

[63] J. Runde. Keynesian uncertainty and the weight of arguments. *Economics and Philosophy*, **6**, 275–292, 1990.

[64] L. J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, **46**, 55–67, 1951.

[65] S. Seabold and J. Perktold. Statsmodels: Econometric and statistical Modeling with python. In: *Proceedings of the 9th Python in Science Conference*, S. van der Walt and J. Millman, eds, pp. 92–96. SciPy, Austin TX, USA, 2010.

[66] T. Seidenfeld. Direct inference and inverse inference. *The Journal of Philosophy*, **75**, 709, 1978.

[67] T. Seidenfeld. 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*, W. Harper and G. Wheeler, eds, pp. 267–279. College Publications, London, UK, 2007.

[68] T. Seidenfeld, J. B. Kadane and M. J. Schervish. On the shared preferences of two Bayesian decision makers. *The Journal of Philosophy*, **86**, 225–244, 1989.

[69] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton NJ, USA 1976.

[70] N. J. J. Smith. Respecting evidence: belief functions not imprecise probabilities. *Synthese*, **200**, 1–30, 2022.

[71] N. J. J. Smith. Acting on belief functions. *Theory and Decision*, **95**, 575–621, 2023.

[72] A. Spanos. Is frequentist testing vulnerable to the base-rate fallacy? *Philosophy of Science*, **77**, 565–583, 2010.

[73] J. Sprenger and S. Hartmann. *Bayesian Philosophy of Science*. Oxford University Press, Oxford, UK, 2019.

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

[75] V. K. Varma, I. Ferguson and I. Wild. Decision support system for the sustainable forest management. *Forest Ecology and Management*, **128**, 49–55, 2000.

[76] A. Wald. Statistical decision functions which minimize the maximum risk. *The Annals of Mathematics*, **46**, 265–280, 1945.

[77] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, UK, 1991.

[78] G. Wheeler and J. Williamson. Evidential probability and objective Bayesian epistemology. In *Philosophy of Statistics, Volume 7 of Handbook of the Philosophy of Science*, P. S. Bandyopadhyay and M. R. Forster, eds, pp. 307–331. Elsevier, Amsterdam, Netherlands, 2011.

[79] R. White. Evidential symmetry and mushy credence. In *Oxford Studies in Epistemology*, T. S. Gendler and J. Hawthorne, eds, vol. **3**, pp. 161–186. Oxford University Press, Oxford, UK, 2010.

[80] J. E. Wilcox. An argument for the principle of indifference and against the wide interval view. *Journal for General Philosophy of Science*, **51**, 65–87, 2019.
[81] J. Williamson. *In Defence of Objective Bayesianism*. Oxford University Press, Oxford, UK, 2010.
[82] J. Williamson. *Lectures on Inductive Logic*. Oxford University Press, Oxford, UK, 2016.
[83] R. R. Yager. Belief structures, weight generating functions and decision-making. *Fuzzy Optimization and Decision Making*, **16**, 1–21, 2016.
[84] S. L. Zabell. Confirming universal generalizations. *Erkenntnis*, **45**, 267–283, 1996.

Received 26 April 2024