

# Fisher’s underworld and the behavioral–statistical reliability balance in scientific inference

Ryan Martin\*

December 9, 2023

## Abstract

That science and other domains are now largely data-driven means virtually unlimited opportunities for statisticians. With great power comes responsibility, so it’s imperative that statisticians ensure that the methods being developing to solve these problems are reliable. But reliable in what sense? This question is problematic because different notions of reliability correspond to distinct statistical schools of thought, each with their own philosophy and methodology, often giving different answers in applications. To achieve the goal of reliably solving modern problems, I argue that a balance in the behavioral–statistical priorities is needed. Towards this, I make use of Fisher’s “underworld of probability” to motivate a new property called *invulnerability* that, roughly, requires the statistician to avoid the risk of losing money in a long-run sense. Then I go on to make connections between invulnerability and the more familiar behaviorally- and statistically-motivated notions, namely coherence and (frequentist-style) validity.

*Keywords and phrases:* Bayesian; coherence; Dempster’s rule; false confidence; frequentist; imprecise probability; inferential model; invulnerability; validity.

## 1 Introduction

Science is now largely data-driven. This means that statisticians—or people doing statistical analyses—play an important role in the advancement of scientific knowledge. But substantial care is needed in every aspect of a scientific investigation, statistical and non-statistical, so simply “doing statistical analyses” isn’t enough. It’s imperative that the process by which data and other available information are fused in order to quantify uncertainty about the scientifically relevant unknowns be *reliable*. The statement I just made is uncontroversial, but only because I gave no specifics that one could disagree with. Attempts to be more specific face immediate setbacks, for example:

*...as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel. (Savage 1972, p. 2)*

---

\*Department of Statistics, North Carolina State University, [rgmarti3@ncsu.edu](mailto:rgmarti3@ncsu.edu)

*Should probability enter to capture degrees of belief about claims? ... Or to ensure we won't reach mistaken interpretations of data too often in the long run of experience?* (Mayo 2018, p. xi)

These different roles mentioned by Mayo that probability plays correspond to distinct statistical schools of thought, so this hits at the very core of the subject.

*Two contending philosophical parties, the Bayesians and the frequentists, have been vying for supremacy over the past two-and-a-half centuries... Unlike most philosophical arguments, this one has important practical consequences. The two philosophies represent competing visions of how science progresses and how mathematical thinking assists in that progress.* (Efron 2013)

The longstanding stalemate between the two schools of thought, their distinct views on probability, and their competing notions of reliability suggests that neither is fully satisfactory. A resolution, therefore, can only come through compromise, through balancing the distinct priorities. This balancing act requires new perspectives on probability, and the present paper's contribution is along these lines.

The “underworld” in the present paper's title refers to one of Sir Ronald Fisher's papers, entitled “The underworld of probability” (Fisher 1957). There, Fisher argues that probability has levels or ranks: that there can be probability statements, probability statements about probability statements, etc., and, furthermore, that one agent might be able to make probabilistic wagers, not on the outcome of the experiment directly, but on whether or not another agent's probabilistic wagers on that outcome pay off. For illustration, Fisher considers the familiar situation involving the toss of a six-sided die, with one of the faces identified as *Ace*. What follows is roughly how Fisher (1957) described the first few levels of the underworld:

- *Level 1 uncertainty.* Agent 1 is sure that the probability of tossing an *Ace* is  $\frac{1}{6}$  but, of course, is uncertain about whether a particular toss of the die will be an *Ace*. He is, however, willing to accept bets against an *Ace* at, say, odds 4:1.
- *Level 2 uncertainty.* Agent 2 isn't sure what the probability of tossing an *Ace* is because he doesn't know the specific die that's to be tossed. He is, however, sure that at least 10% of the dice in the box from which Agent 1's die will be randomly chosen have probability  $\geq \frac{1}{5}$  of turning up *Ace*. So, Agent 2's assessment is that Agent 1's odds against an *Ace* may be favorable to his opponents; therefore, Agent 2 would accept bets at odds 9:1 that Agent 1—with his “accept bets against an *Ace* at odds 4:1” rule—will have negative capital in the long run.

Note that, in the Level 2 case, the agent may or may not have his own probability assessment of tossing an *Ace*. That is, the description above gives insufficient information to pin down a precise estimate of Agent 2's probability of an *Ace*. Agent 2 is saying only that (a) he thinks Agent 1 is wrong and (b) he can formulate odds at which he'd bet against the success of Agent 1. Of course, this process need not stop at Agent 2, it can go even deeper into the underworld, getting less familiar:

- *Level 3 uncertainty.* Agent 3 “does not know and does not think he knows... the probability of throwing an *Ace*, but believes he knows how the probability is distributed in a consignment of boxes, one of which has been used to supply the die in

question, may know with confidence what odds he can profitably accept in betting against the success of [Agent 2]’s wager” (Fisher 1957, p. 203).

This process could continue to Level 4, 5, etc., but Level 3 is a deep enough descent into the underworld for our purposes. Again, Agent 3 need not weigh in on the questions directly relevant to Agents 1–2, his sole objective is capitalizing on a perceived fallibility of Agent 2’s assessments, who in turn is hoping to take advantage of a perceived fallibility of Agent 1’s assessments. That is, Agent 3 has side-bets pertaining to the success of Agent 2’s side-bets on the success of Agent 1’s bets on the die toss.

How is this practically relevant? My claim—which I believe was also Fisher’s—is that Agents 2–3 in the above scenario have counterparts in the statistical inference problem. To make sense of this claim, let  $\Theta$  denote the probability of tossing an Ace. Just like in Fisher’s scenario above, this value is uncertain to Agent 2, but suppose he is able to quantify his uncertainty by making (possibly imprecise) probability statements about  $\Theta$  based on the available information, which might include data, expert opinions, personal beliefs, etc. This makes Agent 2 like the statistician in a statistical inference problem. He is, of course, free to quantify his uncertainty about  $\Theta$  as he pleases, but if his resulting inferences are to be meaningful in any real-world sense, then they must be capable of standing up to scrutiny from the relevant parties, e.g., society, the scientific community, etc. In real applications, we only get one shot to draw inferences and, since the truth about  $\Theta$  is rarely if ever revealed, this scrutiny can’t be based on whether Agent 2 is “correct” or not; it must be based on the reliability of his *methods*. Then Agent 3 in Fisher’s scenario is the scrutinizer, and if he can identify fallibility or vulnerability in the statistician’s uncertainty quantification, i.e., that he can profit off of the perceived systematic errors, then the statistician’s probability assessments have been proven unsound or, more precisely, the methods employed to arrive at these probability statements have been proven unreliable. It’s in this sense that Fisher’s underworld is practically relevant: it aids in understanding and formulating on a common ground the distinct (probabilistic) objectives of the various parties who have a stake in scientific inference.

In light of these observations, the statistician concerned with the reliability of his inferences might aim to ensure that no scrutinizer can identify and capitalize off of a lurking vulnerability in his methods. This puts a constraint on the kinds of methods the statistician can use and what mathematical form his uncertainty quantification takes. The goal of this paper is to carefully formulate this notion of *invulnerability* and to understand the aforementioned constraints and how they can be satisfied.

Following some background on imprecise probability and statistical inference in Section 2—in particular, on *inferential models* (IMs, Martin and Liu 2013, 2015b) for quantifying statistical uncertainty—I’ll define a notion of invulnerability for IMs that aligns with the above intuition, i.e., scrutinizers shouldn’t be able to capitalize off of fallibility in the statistician’s probability assessments. Then Proposition 1 shows that an appropriately defined generalized Bayes IM is invulnerable. This draws an unexpected connection between Fisher’s unique brand of non-subjective, precise-probabilistic reasoning and the largely subjective considerations that dominate the imprecise-probabilistic literature.

The trouble is that generalized Bayes is rather conservative, so it fails to strike the behavioral–statistical reliability balance. If invulnerability needs to be relaxed a bit in order to accommodate the statistical reliability considerations, then a relevant question is: in what direction should it be relaxed? In Section 4, I introduce the (partial-prior)

validity condition that I’ve been advocating for recently (e.g., Martin 2022a,b) and show that validity is a sufficient condition for no sure-loss and a necessary condition for invulnerability. These results together motivate relaxing invulnerability in the direction of validity. Moreover, Proposition 3 generalizes and also gives a new monetary-loss interpretation to the *false confidence theorem* in Balch et al. (2019), Martin (2019), and elsewhere. That is, the result here shows that if validity fails and the statistician’s IM is afflicted with false confidence, then he’s at risk of losing money in the long-run. I hope that this new perspective will help make clear to those who deny the false confidence phenomenon (e.g., Cunen et al. 2020) what’s at risk.

When the available prior information about the relevant unknown is vacuous, validity is equivalent to strong validity and, there, the construction of valid IMs is well-studied. When the partial prior information is not vacuous, all the IMs I’ve developed recently (e.g., Martin 2022b) are strongly valid, and it’s not yet clear what’s the relationship between invulnerability and strong validity. In Section 5, I offer a partial prior IM construction that is valid but not strongly valid; this is based on expressing a (strongly) valid vacuous-prior IM and the partial prior information as belief functions (e.g., Cuzzolin 2021; Shafer 1976) and combining them using Dempster’s rule of combination. Then Proposition 4 establishes that this Dempster’s rule-based IM is valid. A simple numerical example illustrates how this valid partial prior IM helps to strike a behavioral–statistical balance between the vacuous prior IM that ignores the available information and the overly-conservative generalized Bayes IM.

## 2 Problem setup

### 2.1 Model formulation

Let  $(Y, \Theta)$  be a pair of uncertain variables, taking values in the possibility space  $\mathbb{Y} \times \mathbb{T}$ . The pair will be modeled by a general lower and upper prevision pair  $(\underline{P}_{Y,\Theta}, \overline{P}_{Y,\Theta})$  whose domain  $\mathfrak{F}$  is the linear space of bounded gambles. The “gamble” terminology comes from the interpretation of  $f(Y, \Theta)$  as the uncertain payoff an agent receives when the value of  $(Y, \Theta)$  is revealed. Since the lower and upper previsions are linked via conjugacy, i.e.,  $\underline{P}_{Y,\Theta}(f) = -\overline{P}_{Y,\Theta}(-f)$  for each  $f$ , it suffices to discuss just one of them, say, the lower prevision  $\underline{P}_{Y,\Theta}$ . When the gamble is an indicator function,  $f = 1_B$  for  $B \subseteq \mathbb{Y} \times \mathbb{T}$ , I’ll write  $\underline{P}_{Y,\Theta}(B)$  instead of  $\underline{P}_{Y,\Theta}(1_B)$ . In this case, conjugacy takes the form

$$\underline{P}_{Y,\Theta}(B) = 1 - \overline{P}_{Y,\Theta}(B^c), \quad B \subseteq \mathbb{Y} \times \mathbb{T}.$$

The standard subjective interpretation of an agent’s lower/upper prevision pair is as bounds on the prices at which he’s willing to buy/sell gambles concerning the uncertain  $(Y, \Theta)$ . Indeed, as Miranda and de Cooman (2014, Sec. 2.2) explain,

$$\begin{aligned} \underline{P}_{Y,\Theta}(f) &= \text{agent’s supremum acceptable buying price for the gamble } f \\ \overline{P}_{Y,\Theta}(f) &= \text{agent’s infimum acceptable selling price for the gamble } f. \end{aligned}$$

In case the above two prices happen to be the same, that common value would be interpreted as the “fair price” for the gamble  $f$ . Generalizing this “fair price” notion, a gamble

$f$  is *almost desirable* if  $\underline{P}_{Y,\Theta}(f) \geq 0$ . For more on lower previsions and desirability, see Walley (1991, Sec. 3.7.3), Miranda and de Cooman (2014, Sec. 2.2.3), and Quaeghebeur (2014, Sec. 1.6.3). Restricting gambles to indicators, one could interpret  $\underline{P}_{Y,\Theta}(B)$  and  $\overline{P}_{Y,\Theta}(B)$  as subjective measures of the degree of support and plausibility, respectively, for the assertion or hypothesis “ $(Y, \Theta) \in B$ .” This is consistent with the interpretation in Dempster (1967, 1968) and Shafer (1976), even if the lower/upper previsions here don’t completely agree with their mathematical formulation.

To ensure that a lower prevision deserves its “supremum buying price” interpretation, certain mathematical structure is needed. Following Walley (1991, Sec. 2.5.1), who built on de Finetti (1937, 1972), I say that the lower prevision  $\underline{P}_{Y,\Theta}$  is *coherent* if

$$\sup_{y,\theta} \left[ \sum_{k=1}^K \{f_k(y, \theta) - \underline{P}_{Y,\Theta}(f_k)\} - m\{f_0(y, \theta) - \underline{P}_{Y,\Theta}(f_0)\} \right] \geq 0$$

all integers  $m, K \geq 0$  and all  $f_0, f_1, \dots, f_K \in \mathfrak{F}$ . (1)

The case where  $m = 0$  corresponds to what’s called *no sure-loss*, which implies that any finite sum of almost desirable gambles is almost desirable. The general coherence condition (1) simply means that the agent’s “supremum” buying price  $\underline{P}_{Y,\Theta}(f_0)$  for  $f_0$  can’t be raised by simultaneously considering a positive linear combination of other almost desirable gambles (e.g., Miranda and de Cooman 2014, Sec. 2.2.1). Since virtually all the standard imprecise probability models are coherent, I’ll assume throughout this paper, with (virtually) no loss of generality, that  $\underline{P}_{Y,\Theta}$  is a coherent lower prevision.

There are a number of key consequences of coherence, and here I’ll mention just a few. First, Proposition 4.7 in Troffaes and de Cooman (2014) says that coherence implies  $\underline{P}_{Y,\Theta}(f) \leq \overline{P}_{Y,\Theta}(f)$  for any gamble  $f \in \mathfrak{F}$ , which justifies the “lower” and “upper” adjectives. Second, if the lower prevision is coherent and agrees with its corresponding upper prevision, i.e., if  $\underline{P}_{Y,\Theta} = \overline{P}_{Y,\Theta}$ , then the common function, say  $P_{Y,\Theta}$  is called a linear prevision, or precise probability model, and  $P_{Y,\Theta}$  reduces to an expectation operator with respect to a finitely additive probability (e.g., Walley 1991, Sec. 2.8.1). Third,  $\underline{P}_{Y,\Theta}$  equals the lower envelope of a closed and convex (credal) set of linear previsions (e.g., Walley 1991, Prop. 3.3.3). For more details, see Miranda and de Cooman (2014).

A common example in the statistical literature is to model  $Y$ , given  $\Theta = \theta$ , by a precise probability/linear prevision  $P_{Y|\theta}$  and a vacuous prior for  $\Theta$ . While the “no prior” case can’t be accommodated within precise probability theory, it can easily be handled using imprecise probability. Indeed, a minimal extension (Walley 1991, Sec. 6.7) of these individual lower previsions to a coherent lower prevision for  $(Y, \Theta)$  is given by

$$\underline{P}_{Y,\Theta}(f) = \inf_{\theta \in \mathbb{T}} P_{Y|\theta}\{f(Y, \theta)\}, \quad f \in \mathfrak{F},$$

i.e., the infimum of the  $P_{Y|\theta}$ -expectation of  $f(Y, \theta)$  over all  $\theta \in \mathbb{T}$ . For example, if  $f = 1_B$  is an indicator gamble and  $B = A \times H$  is a rectangle in  $\mathbb{Y} \times \mathbb{T}$ , then

$$\underline{P}_{Y,\Theta}(B) = \inf_{\theta \in H} P_{Y|\theta}(A) \quad \text{and} \quad \overline{P}_{Y,\Theta}(B) = \sup_{\theta \in H} P_{Y|\theta}(A).$$

Despite being common in the statistics literature, the vacuous or “no prior” case above is rather extreme. The general imprecise probability formulation is able to handle all the

intermediate/less-extreme cases where, for example, something is known about  $\Theta$  that's relevant to the data analysis, but not enough to justify a precise prior distribution for  $\Theta$  and the corresponding Bayesian analysis. This, of course, is related to robustness (e.g., Huber 1981), especially Bayesian prior robustness (e.g., Berger 1984), but there are other motivations for imprecision in statistical applications, as explained in, e.g., Walley (1991, Ch. 7) and Augustin et al. (2014a, Ch. 7); see, also, Martin (2022a,b).

## 2.2 Statistical inference

In statistical applications, the pair  $(Y, \Theta)$  decomposes as an observable data point  $Y$  and an (unobservable) unknown to be inferred. When the value  $y$  of  $Y$  is observed, the goal is to quantify uncertainty about  $\Theta$ , given  $Y = y$ .

*Statisticians want numerical measures of the degree to which data support hypotheses.* (Hacking 1976, p. 122)

Towards this, the statistician will choose a mapping  $y \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$  that takes a given value of  $y$  and returns a data-dependent lower/upper prevision for  $\Theta$ ; the domain  $\mathfrak{F}_y$  is the linear space consisting of bounded gambles  $f_y(\theta) = f(y, \theta)$  on  $\mathbb{T}$ . I refer to this mapping as an *inferential model*, or IM, and I assume that the collection  $\{\underline{\Pi}_y : y \in \mathbb{Y}\}$  are separately coherent lower previsions (e.g., Walley 1991, Def. 6.2.2). That is, for each  $y$ ,  $\underline{\Pi}_y$  is a coherent lower prevision in the sense that

$$\sup_{\theta} \left[ \sum_{k=1}^K \{f_{k,y}(\theta) - \underline{\Pi}_y(f_{k,y})\} - m \{f_{0,y}(\theta) - \underline{\Pi}_y(f_{0,y})\} \right] \geq 0$$

all integers  $m, K \geq 0$  and all  $f_{0,y}, f_{1,y}, \dots, f_{K,y} \in \mathfrak{F}_y$ ,

and, moreover,  $\underline{\Pi}_y(\{y\} \times \mathbb{T}) = 1$ . Of course, the IM's output can (and will) depend on aspects of the posited model  $(\underline{P}_{Y,\Theta}, \overline{P}_{Y,\Theta})$ , but I'll suppress that dependence in the notation. The lower/upper prevision will be used as follows: for any relevant hypothesis " $\Theta \in H$ " about the unknown, the statistician evaluates  $\underline{\Pi}_y(H)$  and  $\overline{\Pi}_y(H)$  and, based on these magnitudes, makes their judgments/inferences.

Interpretation of the IM is the same as the lower/upper previsions described above. That is, the IM's numerical output represents lower and upper bounds on the prices deemed acceptable for relevant gambles, i.e., given  $Y = y$ ,

$$\begin{aligned} \underline{\Pi}_y(f_y) &= \text{statistician's supremum buying price for the gamble } f_y \\ \overline{\Pi}_y(f_y) &= \text{statistician's infimum selling price for the gamble } f_y. \end{aligned}$$

In other words, given  $Y = y$ , the statistician is willing to pay at most  $\underline{\Pi}_y(f_y)$  dollars to a speculator in exchange for  $f_y(\Theta)$  dollars; similarly, he's willing to accept at least  $\overline{\Pi}_y(f_y)$  dollars from the speculator in exchange for  $f_y(\Theta)$  dollars. In the familiar case of a precise probability, the lower and upper previsions are the same and their common value,  $\Pi_y(f_y)$ , say, would be the statistician's fair price for  $f_y$ . When restricted to indicator gambles,  $\underline{\Pi}_y(H)$  and  $\overline{\Pi}_y(H)$  could represent, respectively, the statistician's degree of belief and of plausibility in the truthfulness of the claim " $\Theta \in H$ ," given  $Y = y$ .

From the aforementioned broad form of (imprecise-)probabilistic uncertainty quantification, it's possible to derive statistical procedures for hypothesis testing and point/set estimation, if desired. For example, to test the hypothesis " $\Theta \in H$ ," corresponding to a fixed subset  $H \subseteq \mathbb{T}$ , it would be natural to "reject  $H$ " based on data  $Y = y$  if and only if  $\bar{\Pi}_y(H)$  is smaller than some specified threshold. It's important to have access to these procedures, but it won't be relevant to the present discussion, so I won't get into these details; the interested reader can consult, e.g., Walley (2002), Martin and Liu (2013), Augustin et al. (2014b), and/or Martin (2021, 2022a,b).

There are an unlimited number of IMs for the statistician to choose from. Common choices of IM construction in the statistical literature, mostly in the case of vacuous prior information about  $\Theta$ , include Fisher's fiducial argument (e.g., Fisher 1935; Seidenfeld 1992; Zabell 1992), default-prior Bayes (e.g., Jeffreys 1946), and generalized fiducial (e.g., Hannig et al. 2016). Still in the case of vacuous prior information, there are the random set-driven IM constructions in Martin and Liu (2013, 2015a,b,c) and the more direct approaches in Martin (2015, 2018); see, also, Cella and Martin (2022). A generalization allowing non-vacuous prior information, is presented in Martin (2022b).

Arguably the most natural IM construction is based on updating the posited model for  $(Y, \Theta)$  to account for the observed  $Y = y$ , analogous to Bayes's rule in ordinary/precise probability. As expected, this is more complicated than the traditional presentations of Bayes's theorem, and Chapters 6–7 of Walley (1991) are devoted to conditioning in the context of general lower/upper previsions. Walley's focus is largely on what kind of properties are required of the posited (joint/marginal) lower prevision  $\underline{P}_{Y,\Theta}$  to ensure the existence (and uniqueness) of a collection  $\{\underline{P}_{\Theta|y} : y \in \mathbb{Y}\}$  of conditional lower previsions that are separately coherent and also jointly coherent with  $\underline{P}_{Y,\Theta}$ . Here's a summary.

Consider the partition of  $\mathbb{Y} \times \mathbb{T}$  determined by the  $y$ -slices

$$\mathcal{B} = \{B_y = \{y\} \times \mathbb{T} : y \in \mathbb{Y}\}. \quad (2)$$

Throughout, when discussing the generalized Bayes IM, I'll (sometimes silently) assume that the posited lower prevision  $\underline{P}_{Y,\Theta}$  is  $\mathcal{B}$ -conglomerable, i.e.,

for any countable collection  $\{B_{y_k} : k = 1, 2, \dots\}$  of (distinct) subsets in the partition  $\mathcal{B}$ , if  $\underline{P}_{Y,\Theta}(1_{B_{y_k}} f) > 0$  for all  $k$ , then  $\underline{P}_{Y,\Theta}(1_{\bigcup_k B_{y_k}} f) \geq 0$ .

In words, this says that, if the restriction of  $f$  to each set in the partition is desirable, then  $f$  itself must also be almost desirable. Conglomerability has a number of far-reaching consequences and Walley (1991, Sec. 6.8.4) argues that full conglomerability (i.e.,  $\mathcal{B}$ -conglomerability for every partition  $\mathcal{B}$ ) ought to be a requirement for a rational model  $\underline{P}_{Y,\Theta}$ . Moreover, conglomerability is a relatively weak requirement: it's automatic when  $\mathcal{B}$  is finite, which is guaranteed when  $\mathbb{Y}$  is finite, and also when  $\underline{P}_{Y,\Theta}(B) = 0$  for all  $B \in \mathcal{B}$ , which is often the case if  $\mathbb{Y}$  is uncountable. For more on conglomerability and its consequences, see Dubins (1975), Schervish et al. (1984), Kadane et al. (1986), Seidenfeld et al. (1998), Miranda and Zaffalon (2013, 2017), and Berti et al. (2017).

The key point is that conglomerability is a necessary and sufficient condition for existence of separately coherent conditional lower previsions  $\{\underline{P}_{\Theta|y} : y \in \mathbb{Y}\}$  that are also jointly coherent with  $\underline{P}_{Y,\Theta}$ ; see Theorem 6.8.2(a) in Walley (1991), Section 5.3 in Berti et al. (2017), and elsewhere. Beyond existence, Walley's Theorem 6.8.2(c) describes the

minimal conditional lower prevision  $\{\underline{\Gamma}_y : y \in \mathbb{Y}\}$  that's coherent with  $\underline{P}_{Y,\Theta}$ , which I refer to here as the *generalized Bayes IM*, defined as follows. For each  $y \in \mathbb{Y}$ :

- if  $\underline{P}_{Y,\Theta}(\{y\} \times \mathbb{T}) = 0$ , then

$$\underline{\Gamma}_y(f_y) = \inf_{\theta \in \mathbb{T}} f(y, \theta), \quad f \in \mathfrak{F};$$

- if  $\underline{P}_{Y,\Theta}(\{y\} \times \mathbb{T}) > 0$ , then  $\underline{\Gamma}_y(f_y)$  is given by the *generalized Bayes rule* (e.g., Walley 1991, Theorem 6.4.2), i.e.,

$$\underline{\Gamma}_y(f_y) = \min \left\{ \frac{P_{Y,\Theta}(f 1_{\{y\} \times \mathbb{T}})}{P_{Y,\Theta}(1_{\{y\} \times \mathbb{T}})} : P_{Y,\Theta} \in \mathcal{C}(\underline{P}_{Y,\Theta}) \right\}, \quad f \in \mathfrak{F},$$

where  $\mathcal{C}(\underline{P}_{Y,\Theta})$  is the (credal) set of all linear previsions  $P_{Y,\Theta}$  that dominate the lower prevision  $\underline{P}_{Y,\Theta}$ . The ratio in the above display is the familiar expectation of the gamble  $f_y(\Theta)$  with respect to the Bayesian posterior for  $\Theta$ , given  $Y = y$ .

### 3 Invulnerability

Let's return now to Fisher's underworld, but specifically in the context of statistical inference, where Agent 2 is the statistician and Agent 3 is the scrutinizer. The scrutinizer knows both the data  $y$  and the statistician's IM, so in principle he could probe the statistician's assessments and determine if there's any vulnerability, if there are any weaknesses he can potentially profit off of in side-bets. It's not all about the money for the scrutinizer; he's also playing an important role in the scientific process:

*Still greater onus is laid on men at large to criticize and probe the evidence given by [statisticians], to question whether they were men trained to observe...*  
(Pearson 2007, p. 54)

There are a couple relevant questions concerning the meaning of "profit off of" that the reader may be asking his/herself, so let me address those here first before moving on.

- *How is Fisher's "profit in the long-run" notion relevant to scientific inference?*

Arguments for the single- $y$  perspective, opposed to a long-run-over-multiple- $y$  perspective, are initially compelling. That is, why should other values of  $Y$  that weren't observed have any bearing on the inferences drawn? But the issue here is that, as Popper argued, individual probability statements can't be falsified, so the statistician—even with his IM that's coherent for the given  $y$ —can't be wrong. Moreover,  $\Theta$  will never be revealed, so even the inferences drawn by the statistician can never be proven wrong. This creates a false sense of security, reinforces attitudes like in the familiar meme "Being a statistician means never having to say you're certain," and breeds carelessness. The "long-run" perspective, however, puts the burden on the statistician to choose an IM whose uncertainty quantification is provably reliable, i.e., his inferences tend to be right. Finally, the aforementioned reliability is just a probability evaluation with respect to the statistician's posited model which, by the way, doesn't require a "long-run" interpretation.

- *Why does the statistician care if a scrutinizer can play side-bets and profit off of his perceived vulnerabilities?*

If it's not the statistician's money going into the scrutinizer's pocket, then why should he care? I'll show below that the statistician's interpretation of his IM output (cf. Section 2.2) obligates him to participate in these side-bets with the scrutinizer. So, if there are vulnerabilities in the statistician's uncertainty quantification, then *it is his money* going in the scrutinizer's pocket! It's precisely this connection to the scrutinizer that forces the statistician to have skin in the game (e.g., Crane 2018; Taleb 2018), i.e., something he personally stands to lose if he's wrong. This skin in the game is broadly beneficial to society—it forces the statistician to be careful.

To formalize this, recall that, for any  $H \subseteq \mathbb{T}$  and  $\beta \in [0, 1]$ , if data  $Y$  is such that  $\underline{\Pi}_Y(H) > \beta$ , then the statistician would be willing to pay the scrutinizer  $\beta$  for  $\$1(\Theta \in H)$ . So, if the scrutinizer believes that the statistician's IM tends to assign too large of  $\underline{\Pi}_Y(H)$  values, and can roughly gauge how large is “too large,” then he has a chance to profit off of the statistician's miscalculation. Towards this, define the function

$$f^{H,\beta}(y, \theta) = \{1(y \in \mathbb{Y}, \theta \in H) - \beta\} 1\{\underline{\Pi}_y(H) > \beta, \theta \in \mathbb{T}\}. \quad (3)$$

Then  $f^{H,\beta}(Y, \Theta)$  represents the statistician's payoff on this  $(H, \beta)$ -dependent gamble, which is an uncertain variable both as a function of  $(Y, \Theta)$  and as a function of  $\Theta$  with  $Y = y$  fixed. It's not difficult to verify that,

$$\underline{\Pi}_y(f_y^{H,\beta}) \geq 0, \quad \text{for all } H \subseteq \mathbb{T}, \text{ all } \beta \in [0, 1], \text{ and all } y \in \mathbb{Y},$$

where

$$f_y^{H,\beta}(\theta) = f^{H,\beta}(y, \theta), \quad \theta \in \mathbb{T}.$$

That is, for any  $(H, \beta)$  pair, the gamble  $f_y^{H,\beta}$  is almost desirable to the statistician, relative to his IM's fixed- $y$  assessments, for each  $y$ . It's in this sense that the statistician is inclined to participate in the scrutinizer's side-bets, exposing him to potential loss in the event that his IM is vulnerable. The following definition spells this out explicitly.

**Definition 1.** Given a coherent model  $(\underline{P}_{Y,\Theta}, \bar{P}_{Y,\Theta})$  for  $(Y, \Theta)$ , the statistician's IM is *invulnerable* if the gamble  $f^{H,\beta}$  in (3), which depends implicitly on the aforementioned IM, is almost desirable with respect to  $\underline{P}_{Y,\Theta}$  for all  $H$  and all  $\beta$ , i.e., if

$$\underline{P}_{Y,\Theta}(f^{H,\beta}) \geq 0 \quad \text{for all } H \subseteq \mathbb{T} \text{ and } \beta \in [0, 1];$$

otherwise, the IM is *vulnerable*.

To the most cautious of statisticians, namely, those who seek protection against all scrutinizers, invulnerability would be a very natural condition. Basically, invulnerability says that a scrutinizer can offer no gamble of the form (3) that's acceptable to the statistician as a function of  $\Theta$  for fixed  $y$  but unacceptable on average as a function of  $(Y, \Theta)$ . There are some connections to coherence, which I'll discuss below.

For some further intuition behind the invulnerability condition, I'd like to make an analogy to Kerckhoff's principle in cryptography:

*[D]esign [crypto]systems under the assumption that the enemy will immediately gain full familiarity with them. (Shannon 1949)*

That is, Kerckhoff's principle says assume that attackers are familiar with your encryption system and design it to be secure regardless of attackers' familiarity. This is the guiding principle behind commonly used public key encryption systems. Like Kerckhoff, invulnerability implies protection from loss even when the scrutinizer has full knowledge of the statistician's IM and the time/skills to probe for potential vulnerabilities.

Despite being a natural or intuitive requirement, there are relevant questions concerning invulnerability that need to be answered. The first basic question is if an invulnerable IM even exists, and the following proposition answers this in the affirmative. Indeed, under mild conditions, the generalized Bayes IM in Section 2.2 is invulnerable.

**Proposition 1** (generalized Bayes is invulnerable). *Suppose that the model  $\mathbb{P}_{Y,\Theta}$  for  $(Y, \Theta)$  is  $\mathcal{B}$ -conglomerable, where  $\mathcal{B}$  is the  $y$ -slice partition in (2). Let  $\{\underline{\Gamma}_y : y \in \mathbb{Y}\}$  denote the generalized Bayes IM, whose existence is implied by the assumed conglomerability. If the statistician's IM  $y \mapsto (\underline{\Pi}_y, \bar{\Pi}_y)$  satisfies  $\underline{\Pi}_y \leq \underline{\Gamma}_y$  for each  $y$ , then it's invulnerable in the sense of Definition 1.*

*Proof.* For any fixed  $H \subseteq \mathbb{T}$  and  $\beta \in [0, 1]$ , define the function

$$\phi(y, \theta) = 1\{\underline{\Pi}_y(H) > \beta, \theta \in \mathbb{T}\}.$$

Then, using Proposition 2.2(v) in Miranda and de Cooman (2014),

$$\begin{aligned} \mathbb{P}_{Y,\Theta}(f^{H,\beta}) &= \mathbb{P}_{Y,\Theta}[\{1(\Theta \in H) - \beta\} \phi(Y, \Theta)] \\ &= \mathbb{P}_{Y,\Theta}[\{1(\Theta \in H) - \underline{\Gamma}_Y(H) + \underline{\Gamma}_Y(H) - \underline{\Pi}_Y(H) + \underline{\Pi}_Y(H) - \beta\} \phi(Y, \Theta)] \\ &\geq \mathbb{P}_{Y,\Theta}[\{1(\Theta \in H) - \underline{\Gamma}_Y(H)\} \phi(Y, \Theta)] \\ &\quad + \mathbb{P}_{Y,\Theta}[\{\underline{\Gamma}_Y(H) - \underline{\Pi}_Y(H)\} \phi(Y, \Theta)] \\ &\quad + \mathbb{P}_{Y,\Theta}[\{\underline{\Pi}_Y(H) - \beta\} \phi(Y, \Theta)]. \end{aligned}$$

Now the goal is to show that all three terms in the lower bound above are non-negative. First, by definition of  $\phi$ , it follows immediately that

$$\mathbb{P}_{Y,\Theta}[\{\underline{\Pi}_Y(H) - \beta\} \phi(Y, \Theta)] \geq 0,$$

which takes care of the last/third term. Second, since  $\phi$  is  $\mathcal{B}$ -measurable, i.e., constant on  $y$ -slices in  $\mathbb{Y} \times \mathbb{T}$ , it follows from Property 8 in Section 6.3.5 in Walley (1991) that

$$\mathbb{P}_{Y,\Theta}[\{1(\Theta \in H) - \underline{\Gamma}_Y(H)\} \phi(Y, \Theta)] \geq 0,$$

which takes care of the first term. Third, and finally, since  $\underline{\Gamma}_y(H) \geq \underline{\Pi}_y(H)$  for each  $y$  and non-negative gambles are almost desirable, it follows that

$$\mathbb{P}_{Y,\Theta}[\{\underline{\Gamma}_Y(H) - \underline{\Pi}_Y(H)\} \phi(Y, \Theta)] \geq 0.$$

Since all three terms in the lower bound are non-negative, it must be that  $f_{H,\beta}$  is almost desirable, which proves the claim.  $\square$

Proposition 1 provides a sufficient condition for invulnerability; I don't know yet if there are other IMs that are invulnerable too. Regardless, this establishes an apparently new property satisfied by the generalized Bayes IM, and draws an apparently previously unknown path between Fisher and imprecise probability through the underworld.

The only downside is that the generalized Bayes IM tends to be quite conservative. For example, in the extreme case with a vacuous prior for  $\Theta$ , the generalized Bayes rule returns a vacuous posterior. It's exactly this conservatism that ensures its coherence and invulnerability. But even Walley wasn't fully committed to generalized Bayes:

*One difficulty in using generalized Bayes rule is that the updated provisions it defines may be highly imprecise... This suggests that generalized Bayes rule should be regarded as one of many possible strategies for updating beliefs. It is the only reasonable strategy that uses just the unconditional assessments, but other strategies, which use other assessments, are equally valid and may lead to greater precision. (Walley 1991, p. 286)*

In a similar vein, I propose seeking a balance of the behavioral–statistical reliability properties, and this apparently requires a relaxation of the invulnerability constraint. This begs the two-part question: a relaxation how and in which direction?

## 4 Validity

Taking a step back, recall that the statistician's basic goal is to offer “reliable” inference. Invulnerability is a special type of reliability, achieved by the generalized Bayes IM and perhaps others, but may not be the only meaningful such notion. What about more basic notions of frequentist-style error rate control? My previous work along these lines was largely motivated by the following remark:

*Even if an empirical frequency-based view of probability is not used directly as a basis for inference, it is unacceptable if a procedure... of representing uncertain knowledge would, if used repeatedly, give systematically misleading conclusions. (Reid and Cox 2015)*

That is, since the statistician's IM is a “procedure of representing uncertain knowledge,” it's “unacceptable” if there's no way to control the error at which wrong inferences are drawn. The following *validity* notion from Martin (2022a) makes this idea precise.

**Definition 2.** Relative to the posited model  $(\underline{P}_{Y,\Theta}, \overline{P}_{Y,\Theta})$ , the statistician's IM  $y \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$  for inference on  $\Theta$  is *valid* if

$$\overline{P}_{Y,\Theta}\{\underline{\Pi}_Y(H) > 1 - \alpha, \Theta \notin H\} \leq \alpha, \quad \text{for all } H \subseteq \mathbb{T} \text{ and all } \alpha \in [0, 1], \quad (4)$$

The interpretation of this result is as follows. If one is reasoning with data based on the magnitudes of the IM's lower and upper probability output, as is recommended, then there's a risk of making erroneous inference if

$$\underline{\Pi}_Y(H) \text{ is large and } \Theta \notin H.$$

There's a similar risk of error if  $\bar{\Pi}_Y(H)$  is small and  $\Theta \in H$ , but this is handled via the conjugacy in the IM's output. To avoid Reid and Cox's "systematically misleading conclusions," the validity condition ensures that the probability of the aforementioned risk-creating event be controlled. That's exactly what (4) does. It's important that " $\alpha$ " appears both inside and outside the probability statement, so that the threshold used to determine if " $\underline{\Pi}_Y(H)$  is large" is explicitly linked to error rate control achieved.

The notion of validity in Definition 2 generalizes that condition with the same name advocated for in, e.g., Martin and Liu (2013, 2015b). This previous work focused on the case of a precise model  $P_{Y|\theta}$  for  $(Y | \Theta = \theta)$  with a vacuous prior for  $\Theta$ . For such a model, the condition in (4) specializes to

$$\sup_{\theta \notin H} P_{Y|\theta} \{ \underline{\Pi}_Y(H) > 1 - \alpha \} \leq \alpha, \quad \text{for all } H \subseteq \mathbb{T} \text{ and all } \alpha \in [0, 1], \quad (5)$$

which is the same as that in Martin and Liu (2013, Sec. 3.3). What distinguishes the new version of validity from the old is that the latter can accommodate more flexibility in the model, allowing for error rate control with respect to more general priors.

Although the validity condition has a strong statistical reliability flavor, it's worth asking—in the spirit of striking a balance—if validity has any behavioral reliability implications as well. Along these lines, the next result says that *validity implies no sure-loss*. More specifically, if one interprets the IM as a rule for updating the prior assessments,

$$\underline{P}_\Theta(H) := \underline{P}_{Y,\Theta}(\mathbb{Y} \times H) \quad \text{and} \quad \bar{P}_\Theta(H) := \bar{P}_{Y,\Theta}(\mathbb{Y} \times H), \quad H \subseteq \mathbb{T},$$

to a data-dependent "posterior" assessment  $(\underline{\Pi}_y, \bar{\Pi}_y)$  given data  $Y = y$ , then one would hope that there's no sure-loss, i.e., the IM should satisfy

$$\inf_{y \in \mathbb{Y}} \underline{\Pi}_y(H) \leq \bar{P}_\Theta(H), \quad \text{for all } H \subseteq \mathbb{T}. \quad (6)$$

In words, if (6) fails, so that the above relation is " $>$ " for some  $H$ , then the scrutinizer could buy the gamble  $\$1(\Theta \in H)$  from the statistician for  $\$\bar{P}_\Theta(H)$  before  $Y$  is observed, and then sell it back to the statistician for  $\$\underline{\Pi}_y(H)$  after  $Y = y$  is observed and, regardless of the value  $y$ , will make a profit of

$$\underline{\Pi}_y(H) - \bar{P}_\Theta(H) > 0 \quad \text{dollars.}$$

A specific behavioral implication of validity is in Proposition 2 below, which confirms that if the IM is valid, then sure-loss in the sense described here is avoided.

**Proposition 2** (validity implies no sure-loss). *If the IM is valid in the sense of Definition 2, then there's no sure-loss relative to the prior assessments, i.e., (6) holds.*

*Proof.* Suppose that there's a sure-loss. That is, there exists an  $H \subseteq \mathbb{T}$  and  $\varepsilon > 0$  such that (6) fails, i.e.,  $\underline{\Pi}_y(H) > \bar{P}_\Theta(H) + \varepsilon$  for all  $y$ . Take  $\alpha = 1 - \bar{P}_\Theta(H) - \varepsilon$ , so that

$$\begin{aligned} \bar{P}_{Y,\Theta} \{ \underline{\Pi}_Y(H) > 1 - \alpha, \Theta \notin H \} &= \bar{P}_{Y,\Theta} \{ \underline{\Pi}_Y(H) > \bar{P}_\Theta(H) + \varepsilon, \Theta \notin H \} \\ &= \bar{P}_\Theta(H^c) \\ &= 1 - \underline{P}_\Theta(H) \\ &\geq 1 - \bar{P}_\Theta(H) \\ &= \alpha + \varepsilon \\ &> \alpha, \end{aligned}$$

where the second equality follows since  $\underline{\Pi}_y(H)$  is lower-bounded by  $\overline{P}_\Theta(H) + \varepsilon$  uniformly in  $y$ . Therefore, validity (4) fails, and the claim follows by contraposition.  $\square$

The next result connects validity to invulnerability. In contrast with Proposition 1, I conclude here that validity is necessary for invulnerability. That is, if the statistician’s IM isn’t valid, then he’s exposed to potential loss by a clever/lucky scrutinizer.

**Proposition 3** (invulnerability implies validity). *Given the model  $\underline{P}_{Y,\Theta}$ , if the statistician’s IM  $y \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$  for  $\Theta$  is invulnerable in the sense of Definition 1, then it’s also valid in the sense of Definition 2.*

*Proof.* Suppose that (4) fails, and let  $(H, \alpha)$  be such that

$$\overline{P}_{Y,\Theta}\{\underline{\Pi}_Y(H) > 1 - \alpha, \Theta \notin H\} > \alpha.$$

Set  $\beta = 1 - \alpha$ , and consider the gamble  $f^{H,\beta}$  as defined above; recall that, by definition,  $f_y^{H,\beta}$  is acceptable to the statistician for any given  $y$ . Rewrite  $f_y^{H,\beta}$  as

$$\begin{aligned} f_y^{H,\beta}(\theta) &= \{1(\theta \in H) - \beta\} 1\{\underline{\Pi}_y(H) > \beta\} \\ &= \{\alpha - 1(\theta \notin H)\} 1\{\underline{\Pi}_y(H) > 1 - \alpha\}. \end{aligned}$$

Define a new gamble

$$\check{f}_y^{H,\beta}(\theta) = \alpha - 1\{\underline{\Pi}_y(H) > 1 - \alpha, \theta \notin H\},$$

and notice that  $f_y^{H,\beta}(\theta) \leq \check{f}_y^{H,\beta}(\theta)$ . Since the former gamble is acceptable for each  $y$ , it must be that the latter gamble is too. But it’s also clear that

$$\underline{P}_{Y,\Theta}(\check{f}^{H,\beta}) = \alpha - \overline{P}_{Y,\Theta}\{\underline{\Pi}_Y(H) \geq 1 - \alpha, \Theta \notin H\},$$

and the right-hand side is negative for this choice of  $(H, \alpha)$ . So, if validity fails, then the IM is vulnerable and, therefore, by contraposition, validity implies no sure-loss.  $\square$

Proposition 3 has close connections to the *false confidence theorem* in Balch et al. (2019). There we showed that, in the case of vacuous prior information about  $\Theta$ , if the IM  $y \mapsto (\underline{\Pi}_y, \overline{\Pi}_y)$  is precise, i.e., if  $\underline{\Pi}_y = \overline{\Pi}_y$ , then it fails to be valid. Consequently, for any threshold  $\alpha$ , there exists  $H \subseteq \mathbb{T}$  such that

$$P_{Y|\theta}\{\underline{\Pi}_Y(H) > 1 - \alpha\} > \alpha, \quad \text{for some } \theta \notin H.$$

That is, there exists a false hypothesis that the IM tends to assign (relatively) high confidence/support/belief to, hence the name “false confidence.” And this has nothing to do with, say, choosing a “wrong prior” in a Bayesian analysis—it’s a consequence of unjustified precision, of pushing the likelihood too far (as I explain in Martin 2023b), that applies to any default-prior Bayes or fiducial-like solution. What’s new and interesting/important here is the behavioral connection: in the vacuous prior case, a precise IM is invalid, hence vulnerable, and hence a statistician adopting such an IM is at risk of losing money to a scrutinizer. Having exposed the behavioral/monetary consequences of unjustified imprecision, I hope that those who dismissed the false confidence phenomenon (e.g., Cunen et al. 2020) can now more clearly see the potential consequences.

This, of course, isn't the first such criticism of unjustifiably precise statistical solutions when the prior information is vacuous. Chapter 7.4 in Walley (1991) has a number of such results; see Seidenfeld (1982) and Caprio and Seidenfeld (2023) for related investigations into fiducial and generalized fiducial inference, respectively; and see, also, Martin (2022a, 2023a). What's new and interesting is the connection made here to the seemingly entirely frequentist-motivated validity property.

Proposition 3 isn't specifically focused on the vacuous prior case, so it also offers a particular generalization of the original false confidence theorem's take-away message. That is, if the statistician's IM fails to be valid even in the more general sense of Definition 2 above (from Martin 2022a), which allows for partial prior information on  $\Theta$ , then there's vulnerability and a corresponding risk of monetary loss to a savvy scrutinizer. This underscores how much care is needed in both the modeling—including prior elicitation—and IM construction steps in a statistical analysis.

Finally, in light of the above observations, the validity property emerges as perhaps the “correct” direction to pursue as it concerns striking a balance between behavioral and statistical reliability. Indeed, the above results reveal the following relationships:

$$\begin{aligned} \text{generalized Bayes IM} &\in \{\text{invulnerable IMs}\} \\ &\subseteq \{\text{valid IMs}\} \\ &\subseteq \{\text{no-sure-loss IMs}\}. \end{aligned}$$

(By the way, Martin (2022a), Corollary 3, established that the generalized Bayes IM is valid, but Propositions 1 and 3 here together offer a new proof.) So, if generalized Bayes and invulnerability more generally are taken to be worthy considerations, but the concern is conservatism due to an over-emphasis on behavioral reliability, then I think it makes sense to seek a better balance by relaxing the requirements, by expanding the class of IMs under consideration. Since one wouldn't want to go so far as to allow IMs that suffer from sure-loss, restricting attention to valid IMs, which have been previously investigated, seems like a very reasonable strategy.

## 5 Striking a balance?

### 5.1 Valid IM construction

Following the suggestion at the end of the previous section, the present goal is to construct an IM that's valid and different from generalized Bayes. Previous work, e.g., in the Martin and Liu (2015b) monograph, focuses on the construction of valid IMs different from generalized Bayes, but there's a subtle point: those previous efforts all focused on the vacuous prior case that's by far the most common in the statistical literature. Those IMs that are valid relative to a vacuous prior are also valid relative to a more precise partial prior (Martin 2022a, Proposition 4), but they actually satisfy a stronger form of validity that's not obviously linked to invulnerability via Proposition 3; see Section 6 below. For now, I want to focus on the present goal, namely, the construction of a partial prior IM that's valid in the sense of Definition 2 above.

What I'm going to present here is one natural construction of a valid, partial prior IM. I make no claims that this is the only such construction or the best; further investigation is needed to properly formulate and hopefully answer the question of “best.”

As mentioned above, it’s now well-known how to construct an IM that’s valid relative to a vacuous prior. So, can one suitably combine that vacuous-prior IM with the available partial prior information? This is similar to how Bayesian inference combines the data-driven likelihood function with a prior density, or how frequentists penalize their objective functions to encourage certain structure in their point estimates. Following the ideas in Martin (2022a), I’ll consider using *Dempster’s rule* to combine the vacuous-prior IM and partial prior information (e.g., Dempster 1967, 2008; Shafer 1976, 2016).

Start with a classical/precise statistical model,  $P_{Y|\theta}$ , for  $(Y \mid \Theta = \theta)$ ; the (partial) prior for  $\Theta$  will be discussed below. Express this precise model  $P_{Y|\theta}$  via an *association* or a “data-generating equation”

$$Y = a(\theta, U), \tag{7}$$

where  $U$  has a known distribution  $P_U$ . This form is familiar in the context of simulating data from a known distribution. That is, we can generate data  $Y$  from distribution  $P_{Y|\theta}$ , with known  $\theta$ , by first generating  $U$  from distribution  $P_U$  and then plugging the corresponding  $(\theta, U)$  into the expression (7) to get  $Y$ . Any model  $P_{Y|\theta}$  can be represented by an expression like (7), but that representation is not unique. This setup aligns with Fisher’s fiducial approach and the generalizations by Fraser (1968), Dempster (2008), and Hannig et al. (2016). The IM construction in Martin and Liu (2015b) differs from these others through its use of a random set  $\mathcal{U} \sim P_{\mathcal{U}}$  for quantifying uncertainty about the unobservable auxiliary variable  $U \sim P_U$  in (7). Define  $\mathcal{U}$  such that the  $y$ -dependent random set on the parameter space  $\mathbb{T}$ , given by

$$\mathbb{T}_y(\mathcal{U}) = \bigcup_{u \in \mathcal{U}} \{\theta : y = a(\theta, u)\}, \quad \mathcal{U} \sim P_{\mathcal{U}},$$

is non-empty for all  $y$ . Then define the IM’s lower and upper probabilities for quantifying uncertainty about  $\Theta$ , given  $Y = y$ , as

$$\begin{aligned} \underline{\Pi}_y^{\text{vac}}(H) &= P_{\mathcal{U}}\{\mathbb{T}_y(\mathcal{U}) \subseteq H\} \\ \bar{\Pi}_y^{\text{vac}}(H) &= P_{\mathcal{U}}\{\mathbb{T}_y(\mathcal{U}) \cap H \neq \emptyset\}, \quad A \subseteq \mathbb{T}. \end{aligned}$$

The  $\underline{\Pi}_y^{\text{vac}}$  is a belief function and, if the random set  $\mathcal{U}$  is nested, i.e., for any pair of realizations one is a subset of the other, then  $\underline{\Pi}_y^{\text{vac}}$  is *consonant* (e.g., Shafer 1976, Ch. 10). I mention consonance because this structure appears to be fundamental as it concerns the IM’s efficiency; see Theorem 4.3 in Martin and Liu (2015b) and Section 6 below. Importantly, under very mild conditions on the random set  $\mathcal{U}$ , it follows that the IM constructed above is valid relative to a vacuous prior for  $\Theta$ , i.e., (5) holds. For conditions on  $\mathcal{U}$  and a proof of the previous claim, see Martin and Liu (2013, Sec. 3).

The IM construction just described would be appropriate if prior information about  $\Theta$  were vacuous. If prior information is not vacuous, then it wouldn’t be unreasonable to encode this too as a belief function (e.g., Cuzzolin 2021; Dencœux 1999) and combine it with the  $y$ -dependent belief function above using Dempster’s rule. A convenient way to express this prior belief function is by introducing a (non-empty) random set  $\mathcal{T} \sim P_{\mathcal{T}}$  on  $\mathbb{T}$ , independent of  $U$  and  $\mathcal{U}$ , so that prior uncertainty can be quantified as

$$\underline{P}_{\Theta}(H) = P_{\mathcal{T}}\{\mathcal{T} \subseteq H\} \quad \text{and} \quad \bar{P}_{\Theta}(H) = P_{\mathcal{T}}\{\mathcal{T} \cap H \neq \emptyset\}, \quad H \subseteq \mathbb{T}.$$

Again, I make no claims that this would be an appropriate model for quantifying uncertainty about  $\Theta$  in every case, but it is quite flexible compared to the all-or-nothing vacuous or precise approaches that statisticians are currently taking.

Now, given data  $Y = y$ , we can combine the available prior information, encoded in  $\mathcal{T}$ , with the output of the IM construction above, expressed in terms of the random set  $\Theta_y(\mathcal{U})$ , using Dempster's rule to get some new IM output, namely,

$$\underline{\Pi}_y(H) = \frac{P_{\mathcal{T} \times \mathcal{U}}\{\emptyset \neq \mathbb{T}_y(\mathcal{U}) \cap \mathcal{T} \subseteq H\}}{P_{\mathcal{T} \times \mathcal{U}}\{\mathbb{T}_y(\mathcal{U}) \cap \mathcal{T} \neq \emptyset\}}, \quad H \subseteq \mathbb{T}, \quad (8)$$

where  $P_{\mathcal{T} \times \mathcal{U}}$  is the joint distribution for the (independent) random sets  $\mathcal{T}$  and  $\mathcal{U}$ . The corresponding upper probability/plausibility is

$$\overline{\Pi}_y(H) = \frac{P_{\mathcal{T} \times \mathcal{U}}\{\mathbb{T}_y(\mathcal{U}) \cap \mathcal{T} \cap H \neq \emptyset\}}{P_{\mathcal{T} \times \mathcal{U}}\{\mathbb{T}_y(\mathcal{U}) \cap \mathcal{T} \neq \emptyset\}}, \quad H \subseteq \mathbb{T}. \quad (9)$$

Note that this IM is different from both generalized Bayes and Dempster's solutions. The difference from generalized Bayes is probably obvious, but the difference from Dempster's solution stems from introducing the random set  $\mathcal{U}$  that's necessary to achieve the vacuous-prior validity in (5) which, in turn, appears to be necessary for proving the more general validity claim below. There's much more that could be said about the proposed IM in (8), but my focus here is just on proving its validity.

Towards this, I'm going to make two assumptions concerning the mathematical structure of the available prior information.

- A1. Assume that the support of  $\mathcal{T}$  is finite, i.e.,  $\mathcal{T}$  has only finitely many focal elements. This implies that  $P_{\mathcal{T}}$  can be described by a mass function  $m$ :

$$\overline{P}_{\Theta}(H) = P_{\mathcal{T}}(\mathcal{T} \cap H \neq \emptyset) = \sum_{T: T \cap H \neq \emptyset} m(T), \quad H \subseteq \mathbb{T}.$$

This is just for mathematical simplicity, the validity result shouldn't hinge on countable versus uncountable support for  $\mathcal{T}$ .

- A2. Assume that  $\mathcal{T}$  is nested—that is, for any  $T, T'$  in the support of  $\mathcal{T}$ , either  $T \subseteq T'$  or  $T \supseteq T'$ . This implies that the prior belief function  $\underline{P}_{\Theta}$  is consonant. Admittedly, this is a non-trivial restriction, but there are many practical cases where this would be appropriate. Like above, I wouldn't expect that consonant versus non-consonant would be relevant to the validity result, but I don't currently know how to prove the claim without consonance.

**Proposition 4.** *Let the model  $(\underline{P}_{Y|\Theta}, \overline{P}_{Y|\Theta})$  be determined by the conditional  $P_{Y|\theta}$  model for  $(Y \mid \Theta = \theta)$  and the partial prior  $(\underline{P}_{\Theta}, \overline{P}_{\Theta})$  for  $\Theta$ . Under Assumptions A1–A2, the Dempster's rule-based IM construction in (8) is valid in the sense of Definition 2.*

*Proof.* Take any fixed focal element  $T$  of  $\mathcal{T}$ . By monotonicity of  $H \mapsto \overline{\Pi}_y(H)$  for each  $y$ , it's easy to confirm that

$$P_{Y|\theta}\{\overline{\Pi}_Y(H) \leq \alpha\} \leq P_{Y|\theta}\{\overline{\Pi}_Y(H \cap T) \leq \alpha\}. \quad (10)$$

Next, note the following lower bound on  $\bar{\Pi}_Y(H \cap T)$  from (9):

$$\begin{aligned}
\bar{\Pi}_y(H \cap T) &= \frac{\mathbb{P}_{\mathcal{T} \times \mathcal{U}}\{\Theta_y(\mathcal{U}) \cap \mathcal{T} \cap (H \cap T) \neq \emptyset\}}{\mathbb{P}_{\mathcal{T} \times \mathcal{U}}\{\Theta_y(\mathcal{U}) \cap \mathcal{T} \neq \emptyset\}} \\
&\geq \mathbb{P}_{\mathcal{T} \times \mathcal{U}}\{\Theta_y(\mathcal{U}) \cap \mathcal{T} \cap (H \cap T) \neq \emptyset\} \\
&= \sum_{T': T' \cap H \cap T \neq \emptyset} m(T') \mathbb{P}_{\mathcal{T} \times \mathcal{U}}\{\Theta_y(\mathcal{U}) \cap (T' \cap H \cap T) \neq \emptyset\} \\
&= \mathbb{P}_{\mathcal{U}}\{\Theta_y(\mathcal{U}) \cap (H \cap T) \neq \emptyset\} \bar{\mathbb{P}}_{\Theta}(H) \\
&= \bar{\Pi}_y^{\text{vac}}(H \cap T) \bar{\mathbb{P}}_{\Theta}(H)
\end{aligned} \tag{11}$$

The first inequality follows because the denominator is a probability, hence  $\leq 1$ ; the next equality is by the total probability formula with respect to the distribution of  $\mathcal{T}$ ; the last equality follows by the assumed nestedness of  $\mathcal{T}$ , i.e., that  $T'$  has non-empty intersection with  $H \cap T$  if and only if  $T'$  has non-empty intersection with  $H$ .

Now suppose—without loss of generality—that the focal elements of  $\mathcal{T}$  are ordered as  $T_k \subset T_{k+1}$ , for  $k \geq 1$ . Define the expansion

$$H = \bigcup_{k \geq 1} \underbrace{H \cap (T_k \setminus T_{k-1})}_{H_k},$$

where the  $H_k$ 's are disjoint. Notice that  $H_k \subseteq H \cap T_k$ , so if  $\theta \in H_k$ , then the hypothesis  $H \cap T_k$  is “true” and it follows from the vacuous-prior validity result (5) that

$$\sup_{\theta \in H_k} \mathbb{P}_{Y|\theta}\{\bar{\Pi}_Y^{\text{vac}}(H \cap T_k) \leq \rho\} \leq \min(1, \rho), \quad \text{for any } \rho > 0. \tag{12}$$

Finally, with a slight abuse of notation, denote the extension of the prior upper probability/plausibility function for  $\Theta$  to an upper prevision (supported on a suitable set of bounded gambles on  $\mathbb{T}$ ) by the same symbol  $\bar{\mathbb{P}}_{\Theta}$ . This extension is achieved through a simple Choquet integral as shown in, e.g., Denceaux (2019). Now, take any linear prevision  $\mathbb{Q}$  dominated by the upper prevision  $\bar{\mathbb{P}}_{\Theta}$ , i.e.,  $\mathbb{Q}(\cdot) \leq \bar{\mathbb{P}}_{\Theta}(\cdot)$ . Then

$$\begin{aligned}
\mathbb{Q}[1(\theta \in H) \mathbb{P}_{Y|\theta}\{\bar{\Pi}_Y(H) \leq \alpha\}] &= \sum_k \mathbb{Q}[1(\theta \in H_k) \mathbb{P}_{Y|\theta}\{\bar{\Pi}_Y(H) \leq \alpha\}] \\
&\leq \sum_k \mathbb{Q}[1(\theta \in H_k) \mathbb{P}_{Y|\theta}\{\bar{\Pi}_Y(H \cap T_k) \leq \alpha\}] \\
&\leq \sum_k \mathbb{Q}[1(\theta \in H_k) \mathbb{P}_{Y|\theta}\{\bar{\Pi}_Y^{\text{vac}}(H \cap T_k) \leq \alpha \bar{\mathbb{P}}_{\Theta}(H)^{-1}\}] \\
&\leq \sum_k \min\{1, \alpha \bar{\mathbb{P}}_{\Theta}(H)^{-1}\} \mathbb{Q}(H_k) \\
&\leq \alpha \bar{\mathbb{P}}_{\Theta}(H)^{-1} \mathbb{Q}(H),
\end{aligned}$$

where the equality is by the disjoint-union representation  $H = \bigcup_k H_k$  and finite additivity, the first inequality is by monotonicity (10), the second inequality is by (11), the third inequality is by (12), and the fourth inequality by the disjoint-union representation again. By definition,  $\mathbb{Q}(H) \leq \bar{\mathbb{P}}_{\Theta}(H)$ , so the upper bound is no more than  $\alpha$ . Since  $\bar{\mathbb{P}}_{Y,\Theta}$  is

coherent, it's the upper envelope of the linear previsions  $P_{Y,\Theta}$  it dominates. For the present model—precise likelihood, imprecise prior—that upper envelope corresponds to a supremum over all the priors  $Q$  for  $\Theta$  dominated by  $\bar{P}_\Theta$ . Therefore,

$$\bar{P}_{Y,\Theta}\{\bar{\Pi}_Y(H) \leq \alpha, \Theta \in H\} = \sup_{Q: Q \leq \bar{P}_\Theta} Q[1(\Theta \in H) P_{Y|\Theta}\{\bar{\Pi}_Y(H) \leq \alpha\}],$$

and, since the upper bound is no more than  $\alpha$ , the validity claim follows.  $\square$

To recap, I'm proposing a simple IM construction based on treating the vacuous-prior IM and the partial prior as independent belief functions—determined by their own independent random sets—and combining them via Dempster's rule. This isn't the only IM construction available, but it's a simple and reasonable idea. Proposition 4 establishes that, besides being “reasonable,” this Dempster's rule-based IM construction is valid in the sense of Definition 2, at least under certain conditions on the prior specification.

## 5.2 Illustration

To illustrate the potential balance that this alternative to the generalized Bayes IM construction, I'll consider a simple example. Model  $(Y | \Theta = \theta)$  as  $P_{Y|\theta} = \mathbf{N}(\theta, 1)$ , a normal distribution with mean  $\theta$  and variance 1. I'll start here with a quick review of the vacuous prior IM for  $\Theta$ , given  $Y = y$ , derived in Martin and Liu (2013). Write the association (7) as  $Y = \theta + U$ , where  $U \sim \mathbf{N}(0, 1)$ . Model the unobserved value of  $U$  by a belief function determined by the random set

$$\mathcal{U} = \{u \in \mathbb{R} : |u| \leq |U^*|\}, \quad U^* \sim \mathbf{N}(0, 1),$$

a symmetric interval around the origin with random length. Then

$$\mathbb{T}_y(\mathcal{U}) = \bigcup_{u \in \mathcal{U}} \{\theta : y = \theta + u\} = [y - |U^*|, y + |U^*|], \quad U^* \sim \mathbf{N}(0, 1).$$

For a class of hypotheses  $(-\infty, \theta]$ ,  $\theta \in \mathbb{R}$ , it's easy to verify that

$$\begin{aligned} \underline{\Pi}_y^{\text{vac}}((-\infty, \theta]) &= \mathbf{P}(y + |U^*| \leq \theta) \\ \bar{\Pi}_y^{\text{vac}}((-\infty, \theta]) &= \mathbf{P}(y - |U^*| \leq \theta). \end{aligned}$$

Individually, these functions have the properties of a cumulative distribution function as  $\theta$  varies, for fixed  $y$ , but of course there's an ordering that holds pointwise in  $\theta$ . Figure 1 plots these lower and upper distribution functions for a few different values of  $y$ .

Next, suppose the statistician is “90% sure that  $\Theta$  is no more than 7.” This incomplete prior information can be modeled by a belief function determined by the random set  $\mathcal{T}$

$$\mathcal{T} = \begin{cases} (-\infty, 7] & \text{with probability 0.9} \\ \mathbb{R} & \text{with probability 0.1,} \end{cases}$$

so the mass function is

$$m((-\infty, 7]) = 0.9 \quad \text{and} \quad m(\mathbb{R}) = 0.1.$$

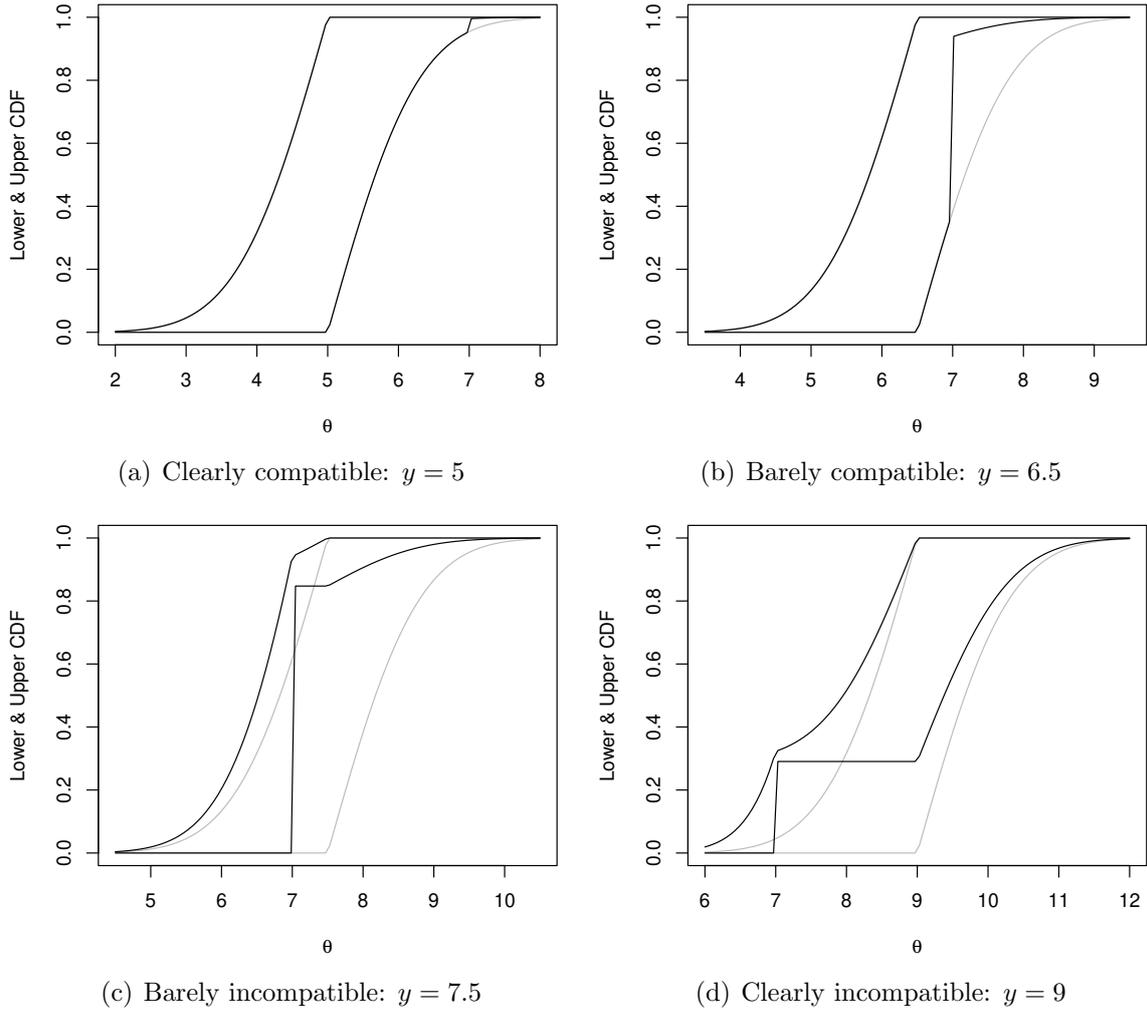


Figure 1: Plots of the lower and upper distribution functions for the vacuous-prior IM (gray) and Dempster’s rule-based IM (black) for four values of  $y$  corresponding to four different degrees of compatibility with the partial prior information.

It’s not too difficult to carry employ Dempster’s rule to combine this prior with the vacuous-prior IM described above—the expressions are messy and not worth presenting here—and then find the corresponding lower and upper distributions. Figure 1 plots these lower and upper distribution functions for a few different values of  $y$ .

In Figure 1, the  $y$  values range from being clearly compatible with the prior information, barely compatible, barely incompatible, and clearly incompatible. In the clearly-compatible case, there’s virtually no difference between the two IMs’ results, with the Dempster’s rule-based solution being slightly more precise. In the barely-compatible case, the increased precision is more pronounced compared to the previous case because there’s more to be gained from the prior information in this boundary case. When the data is (barely or clearly) incompatible with the prior information, the prior’s effect is more significant and, moreover, the difference between the two IMs can no longer be described in terms of relative precision. To understand the effect in the incompatible cases, it helps

to think about a statistical summary, namely, the interval determined by, say, the 2.5th and 97.5th percentiles of the upper and lower distribution functions, respectively. For the vacuous-prior IM, these intervals have length roughly 4. In the barely-incompatible case, the incompatibility can easily be chalked up to sampling variation, so the prior steps in and tightens up the aforementioned interval to have length less than 3.5. In the clearly-incompatible case, on the other hand, the discrepancy can't easily be explained by sampling variability, so, thanks to the validity property (Proposition 4), the IM responds by stretching out the aforementioned interval to roughly length 5.

For comparison, the generalized Bayes IM described in Section 2.2 is vacuous in this case. That is, thanks to the invulnerability and coherence constraints, the generalized Bayes IM is so conservative that it's unable to learn based on the observation  $Y = y$ . The Dempster's rule-based IM sacrifices on invulnerability—but not on validity—in exchange for the ability to make non-trivial inferences in common cases like these where detailed prior information is lacking. This illustrates the balance that I'm hoping to achieve by relaxing the invulnerability requirement in the direction of valid-but-vulnerable IMs.

## 6 Conclusion

In this paper, I start by considering Fisher's so-called "underworld of probability," which describes—in terms of bets—the different levels or perspectives on uncertainty. Each of these levels corresponds to an agent having different priorities and different forms of information available. While Fisher doesn't say so directly, I believe that his motivation for writing about the underworld was to make an analogy between agents in his formulation and (some of) the key players in scientific inference, namely,

- the statistician, or the one who quantifies uncertainty for the purpose of drawing inferences about scientifically relevant unknowns based on observed data, and
- the scrutinizers of scientific inference, e.g., society and/or the scientific community, the ones who stand to lose if the statistician's inferences happen to be erroneous.

Understanding the various priorities is imperative because scientific advances are now largely data-driven. I use Fisher's underworld considerations as a jumping off point to define a notion of *invulnerability* that ensures no scrutinizer can identify and profit off of fallibility in the statistician's uncertainty quantification and inference. Then Proposition 1 establishes that invulnerability can be attained, in particular, by what I call here the generalized Bayes inferential model (IM). This establishes an unexpected and apparently new connection between Fisher's nuanced/elusive brand of probabilistic reasoning and the behavioral considerations of de Finetti, Walley, Williams, and many other contributors to the (imprecise) probability literature.

While having statistical motivations, the close connection between invulnerability and behavioral reliability considerations—and the well-known conservatism of generalized Bayes—suggests a relaxation in some direction might be needed to achieve the desired balance. Towards this, Proposition 3 shows that the validity property (Definition 2) that I've been advocating for recently is a necessary condition for invulnerability. This, together with the dual facts that generalized Bayes IMs are valid and valid IMs avoid sure loss, encourages relaxing the (apparently conservative) notion of invulnerability in the

direction of validity. Section 5 offers a particular IM construction—building on my vacuous prior IM constructions in previous work—that incorporates partial prior information via a suitable belief function and Dempster’s rule of combination. Then Proposition 4 shows that, under certain conditions on the prior belief function, this partial prior IM is valid. A simple example illustrates this proposal and shows how this IM construction might strike the desired balance: there’s clear improvement compared to the generalized Bayes IM when the prior information is very limited, and even improvements compared to the vacuous prior IM when the data is compatible with the prior information.

More recently, I’ve advanced a very general partial prior IM construction (Martin 2022b, 2023c) that achieves an even stronger notion of validity, what I call *strong validity*. As the name suggests, strong validity is stronger than validity, which is easy to see from its representation as a “uniform version” of (4):

$$\bar{P}_{Y,\Theta}\{\underline{\Pi}_Y(H) > 1 - \alpha \text{ for some } H \text{ with } H \not\subseteq \Theta\} \leq \alpha, \quad \alpha \in [0, 1].$$

See Cella and Martin (2023) for some important implications of this uniformity. It turns out that strong validity is closely tied to consonance, possibility theory, and the construction of what I refer to as a *possibilistic IM* relies on outer consonant approximations (e.g., Dubois and Prade 1990; Montes et al. 2019) and the so-called imprecise-probability-to-possibility transform (e.g., Dubois et al. 2004; Hose 2022). Consequently, possibilistic IMs have structure that the generalized Bayes and Dempster’s rule-based IMs don’t, and I argue in Martin (2022b) that despite being largely statistically motivated and satisfying a number of important statistical properties, it has some desirable behavioral properties as well. To summarize, the Venn diagram in Figure 2 illustrates that both strongly valid and invulnerable IMs are special cases of valid IMs, and suggests an interesting open question: are there any possibilistic IMs that are also invulnerable? If so, then that IM would arguably settle the statistical–behavioral reliability balance question.

## Acknowledgments

Thanks to Teddy Seidenfeld for suggesting Fisher’s “underworld paper” to me, and also for his critique following my *SIPTA Seminar* presentation (<https://www.youtube.com/watch?v=1VL5P1YcXdo>). And special thanks to Enrique Miranda for some corrections and helpful suggestions concerning an earlier draft of this manuscript. This work is partially supported by the U.S. National Science Foundation, grant SES–2051225.

## References

- Augustin, T., Coolen, F. P. A., de Cooman, G., and Troffaes, M. C. M., editors (2014a). *Introduction to Imprecise Probabilities*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- Augustin, T., Walter, G., and Coolen, F. P. A. (2014b). Statistical inference. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 135–189. Wiley, Chichester.

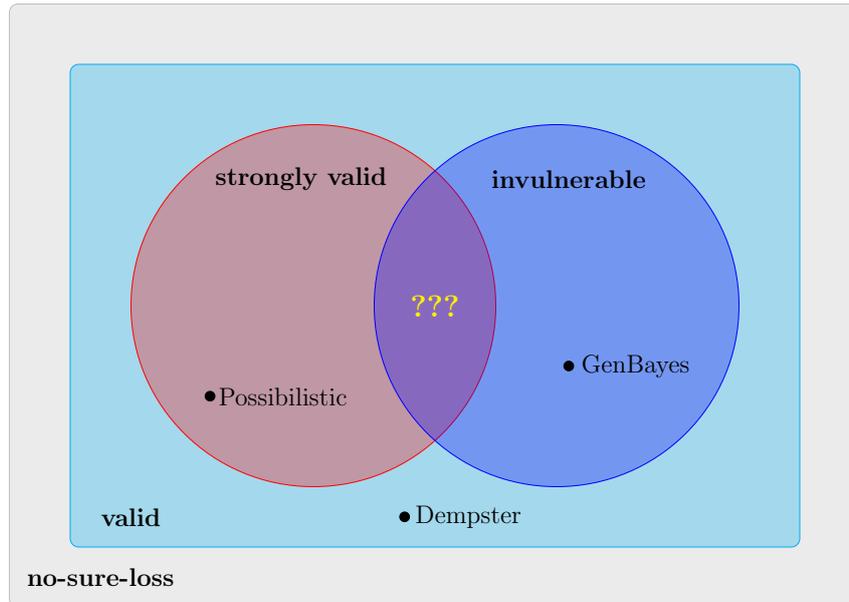


Figure 2: Venn diagram to illustration the results presented in this paper. The specific IM constructions—generalized Bayes, Dempster’s rule, and possibilistic—are highlighted by points in respective regions they fall into.

- Balch, M. S., Martin, R., and Ferson, S. (2019). Satellite conjunction analysis and the false confidence theorem. *Proc. Royal Soc. A*, 475(2227):2018.0565.
- Berger, J. O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses*, volume 4 of *Stud. Bayesian Econometrics*, pages 63–144. North-Holland, Amsterdam. With comments and with a reply by the author.
- Berti, P., Miranda, E., and Rigo, P. (2017). Basic ideas underlying conglomerability and disintegrability. *Internat. J. Approx. Reason.*, 88:387–400.
- Caprio, M. and Seidenfeld, T. (2023). Constriction for sets of probabilities. In Miranda, E., Montes, I., Quaeghebeur, E., and Vantaggi, B., editors, *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*, pages 84–95. PMLR.
- Cella, L. and Martin, R. (2022). Direct and approximately valid probabilistic inference on a class of statistical functionals. *Internat. J. Approx. Reason.*, 151:205–224.
- Cella, L. and Martin, R. (2023). Possibility-theoretic statistical inference offers performance and probativeness assurances. *Internat. J. Approx. Reason.*, 163:109060.
- Crane, H. (2018). The fundamental principle of probability: Resolving the replication crisis with skin in the game. <https://researchers.one/articles/18.08.00013>.
- Cunen, C., Hjort, N. L., and Schweder, T. (2020). Confidence in confidence distributions! *Proc. Roy. Soc. A*, 476:20190781.

- Cuzzolin, F. (2021). *The Geometry of Uncertainty*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer, Cham.
- de Finetti, B. (1937). La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, 7(1):1–68.
- de Finetti, B. (1972). *Probability, Induction and Statistics*. John Wiley & Sons, London-New York-Sydney. Wiley Series in Probability and Mathematical Statistics.
- Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38:325–339.
- Dempster, A. P. (1968). A generalization of Bayesian inference. (With discussion). *J. Roy. Statist. Soc. Ser. B*, 30:205–247.
- Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. *Internat. J. Approx. Reason.*, 48(2):365–377.
- Dencœux, T. (1999). Reasoning with imprecise belief structures. *Internat. J. Approx. Reason.*, 20(1):79–111.
- Dencœux, T. (2019). Decision-making with belief functions: a review. *Internat. J. Approx. Reason.*, 109:87–110.
- Dubins, L. E. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probability*, 3:89–99.
- Dubois, D., Foulloy, L., Mauris, G., and Prade, H. (2004). Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliab. Comput.*, 10(4):273–297.
- Dubois, D. and Prade, H. (1990). Consonant approximations of belief functions. *Internat. J. Approx. Reason.*, 4(5-6):419–449.
- Efron, B. (2013). A 250-year argument: belief, behavior, and the bootstrap. *Bull. Amer. Math. Soc. (N.S.)*, 50(1):129–146.
- Fisher, R. A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics*, 6:391–398.
- Fisher, R. A. (1957). The underworld of probability. *Sankhya*, 18:201–210.
- Fraser, D. A. S. (1968). *The Structure of Inference*. John Wiley & Sons Inc., New York.
- Hacking, I. (1976). *Logic of Statistical Inference*. Cambridge University Press, Cambridge-New York-Melbourne, paperback edition.
- Hannig, J., Iyer, H., Lai, R. C. S., and Lee, T. C. M. (2016). Generalized fiducial inference: a review and new results. *J. Amer. Statist. Assoc.*, 111(515):1346–1361.
- Hose, D. (2022). *Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering*. PhD thesis, University of Stuttgart.

- Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Ser. A*, 186:453–461.
- Kadane, J. B., Schervish, M. J., and Seidenfeld, T. (1986). Statistical implications of finitely additive probability. In *Bayesian inference and decision techniques*, volume 6 of *Stud. Bayesian Econometrics Statist.*, pages 59–76. North-Holland, Amsterdam.
- Martin, R. (2015). Plausibility functions and exact frequentist inference. *J. Amer. Statist. Assoc.*, 110(512):1552–1561.
- Martin, R. (2018). On an inferential model construction using generalized associations. *J. Statist. Plann. Inference*, 195:105–115.
- Martin, R. (2019). False confidence, non-additive beliefs, and valid statistical inference. *Internat. J. Approx. Reason.*, 113:39–73.
- Martin, R. (2021). An imprecise-probabilistic characterization of frequentist statistical inference. [arXiv:2112.10904](https://arxiv.org/abs/2112.10904).
- Martin, R. (2022a). Valid and efficient imprecise-probabilistic inference with partial priors, I. First results. [arXiv:2203.06703](https://arxiv.org/abs/2203.06703).
- Martin, R. (2022b). Valid and efficient imprecise-probabilistic inference with partial priors, II. General framework. [arXiv:2211.14567](https://arxiv.org/abs/2211.14567).
- Martin, R. (2023a). Fiducial inference viewed through a possibility-theoretic inferential model lens. In Miranda, E., Montes, I., Quaeghebeur, E., and Vantaggi, B., editors, *Proceedings of the Thirteenth International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*, pages 299–310. PMLR.
- Martin, R. (2023b). A possibility-theoretic solution to Basu’s Bayesian–frequentist via media. *Sankhya A*, to appear, [arXiv:2303.17425](https://arxiv.org/abs/2303.17425).
- Martin, R. (2023c). Valid and efficient imprecise-probabilistic inference with partial priors, III. Marginalization. [arXiv:2309.13454](https://arxiv.org/abs/2309.13454).
- Martin, R. and Liu, C. (2013). Inferential models: a framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.*, 108(501):301–313.
- Martin, R. and Liu, C. (2015a). Conditional inferential models: combining information for prior-free probabilistic inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 77(1):195–217.
- Martin, R. and Liu, C. (2015b). *Inferential Models*, volume 147 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.

- Martin, R. and Liu, C. (2015c). Marginal inferential models: prior-free probabilistic inference on interest parameters. *J. Amer. Statist. Assoc.*, 110(512):1621–1631.
- Mayo, D. G. (2018). *Statistical Inference as Severe Testing*. Cambridge University Press, Cambridge.
- Miranda, E. and de Cooman, G. (2014). Lower previsions. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 28–55. Wiley, Chichester.
- Miranda, E. and Zaffalon, M. (2013). Conglomerable coherence. *Internat. J. Approx. Reason.*, 54(9):1322–1350.
- Miranda, E. and Zaffalon, M. (2017). Full conglomerability. *J. Stat. Theory Pract.*, 11(4):634–669.
- Montes, I., Miranda, E., and Vicig, P. (2019). Outer approximating coherent lower probabilities with belief functions. *Internat. J. Approx. Reason.*, 110:1–30.
- Pearson, K. (2007). *The Grammar of Science*. Cosimo, New York, 3rd edition. Reprint of the 1911 edition.
- Quaeghebeur, E. (2014). Desirability. In *Introduction to Imprecise Probabilities*, Wiley Ser. Probab. Stat., pages 1–27. Wiley, Chichester.
- Reid, N. and Cox, D. R. (2015). On some principles of statistical inference. *Int. Stat. Rev.*, 83(2):293–308.
- Savage, L. J. (1972). *The Foundations of Statistics*. Dover Publications, Inc., New York, revised edition.
- Schervish, M. J., Seidenfeld, T., and Kadane, J. B. (1984). The extent of nonconglomerability of finitely additive probabilities. *Z. Wahrsch. Verw. Gebiete*, 66(2):205–226.
- Seidenfeld, T. (1982). Paradoxes of conglomerability and fiducial inference. In *Logic, Methodology and Philosophy of Science, VI (Hannover, 1979)*, volume 104 of *Stud. Logic Found. Math.*, pages 395–412. North-Holland, Amsterdam.
- Seidenfeld, T. (1992). R. A. Fisher’s fiducial argument and Bayes’ theorem. *Statist. Sci.*, 7(3):358–368.
- Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (1998). Non-conglomerability for finite-valued, finitely additive probability. *Sankhyā Ser. A*, 60(3):476–491.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J.
- Shafer, G. (2016). Dempster’s rule of combination. *Internat. J. Approx. Reason.*, 79:26–40.
- Shannon, C. E. (1949). Communication theory of secrecy systems. *Bell System Tech. J.*, 28:656–715.

- Taleb, N. N. (2018). *Skin in the Game: Hidden Asymmetries in Daily Life*. New York: Random House.
- Troffaes, M. C. M. and de Cooman, G. (2014). *Lower Previsions*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*, volume 42 of *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London.
- Walley, P. (2002). Reconciling frequentist properties with the likelihood principle. *J. Statist. Plann. Inference*, 105(1):35–65.
- Zabell, S. L. (1992). R. A. Fisher and the fiducial argument. *Statist. Sci.*, 7(3):369–387.