Forecasting with Imprecise Probabilities

Teddy Seidenfeld teddy@stat.cmu.edu
Mark J. Schervish mark@cmu.edu
Joseph B. Kadane kadane@stat.cmu.edu

Carnegie Mellon University

Abstract

We review de Finetti's two coherence criteria for determinate probabilities: *coherence*$_1$ defined in terms of previsions for a set of events that are undominated by the status quo – previsions immune to a sure-loss – and *coherence*$_2$ defined in terms of forecasts for events undominated in Brier score by a rival forecast. We propose a criterion of IP-coherence$_2$ based on a generalization of Brier score for IP-forecasts that uses 1-sided, lower and upper, probability forecasts. However, whereas Brier score is a strictly proper scoring rule for eliciting determinate probabilities, we show that there is no *real-valued* strictly proper IP-score. Nonetheless, with respect to either of two decision rules – Γ-*Maximin* or (Levi's) **E**-a*dmissibility*-+-Γ-*Maximin* – we give a *lexicographic* strictly proper IP-scoring rule that is based on Brier score.

Keywords. Brier score, coherence, dominance, **E**-*admissibility*, Γ-*Maximin,* proper scoring rules.

1. Introduction

One important approach to the *foundations* for subjective probability is the strategy to reduce rational degrees of belief to normative decision theory. Savage's [17] is a classic among such theories. De Finetti's *Book* argument, dating from about 1930 and summarized in [3], is another. De Finetti considers personal *previsions*, which are an agent's fair prices for buying and selling random variables. These random variables are defined with respect to some common space of possibilities. De Finetti introduces a criterion of *coherent previsions:* that the agent's fair prices cannot be used to form a set of trades that result in a uniform sure loss with respect to that space of possibilities. Thus in de Finetti's theory, coherence is a normative decision theoretic constraint on an agent's previsions: avoid sure loss. He established the central result that a set of previsions is coherent in this sense just in case there is some (finitely additive) probability against which the prevision for a random variable is its expected value. When the random variables are indicator functions for events, coherent previsions are the agent's personal probabilities for those events, and the agent's fair prices are her/his coherent betting odds. De Finetti thus reduced the problem of rational degrees of belief to the problem of coherent previsions.

Starting in about 1960, de Finetti emphasized two coherence criteria – coherence$_1$ for previsions (as described above), and coherence$_2$ for forecasts assessed by Brier score. He established [3, 5] that these two criteria are equivalent for purposes of distinguishing between sets of previsions or sets of forecasts that are undominated versus those that are dominated. *Coherence* is the common requirement that a decision maker avoids dominated alternatives. A set of previsions are coherent$_1$ i.e., they are undominated by the alternative of the status-quo – there is no "Book" – if and only if those same quantities, when used as forecasts evaluated by Brier score, are coherent$_2$, i.e., they are undominated by any rival set of forecasts. In his later presentations de Finetti favored coherence$_2$ over coherence$_1$ because, in addition to providing an equivalent criterion for coherence, also proper scores provide a method for incentive compatible elicitation, unlike the situation with coherence$_1$ and the *prevision game*, as we call it. In Section 2, we make precise and explain these claims.

De Finetti's theory of coherent previsions, coherence$_1$, serves as the basis for numerous *IP* generalizations – see [13, 26, 27, 28] for examples. However, we know of no parallel development of IP theory based on proper scoring rules. It is our purpose in this essay to report some basic findings about scoring-rule based IP theory. In Section 3 we explain one approach to an IP version of coherence$_2$ and illustrate how that approach works. In Section 4 we present an impossibility result for a *real-valued* proper IP scoring rule. By contrast, we illustrate a strictly proper, lexicographic (non-standard) IP version of Brier score. In Section 5 we conclude with remarks about the approach begun here.

2. De Finetti's two criteria for coherence

2.1 Coherence$_1$ and Coherence$_2$. We begin our review of de Finetti's theory with a reformulation of his first coherence criterion, *coherence*$_1$, which constrains a rational agent's fair prices for buying and selling random variables. Coherence$_1$ requires that the rational agent's fair prices cannot result in a set of trades that result in a (uniform) sure loss. Coherent$_1$ previsions cannot be dominated by the status-quo, where there are no trades. We reformulate coherence$_1$ in the context of a 2-person, 0-sum game, the *prevision game*, in order to prepare the reader for concerns about strategic aspects of applying coherence$_1$. These strategic aspects can distort the rational agent's play, even though that play is coherent$_1$. That is, the rational agent may have incentives to play coherently$_1$ in the game while misidentifying his/her degrees of belief.

The existence of such strategic aspects in the prevision game help to motivate de Finetti's second coherence criterion, *coherence*$_2$. Coherence$_2$ constrains the rational agent's forecasts for the same set of random variables by requiring that, as assessed by Brier score, forecasts are undominated relative to each rival set of forecasts. As we explain, below, coherence$_2$ is an incentive compatible criterion for forecasting variables that provides an alternative foundation for subjective probability. It mitigates the strategic aspects of rational play that threaten to distort the agent's announced prices in the prevision game.

The *prevision game*, is formulated for a class of bounded variables, $\mathcal{X} = \{X_i: i \in \mathbf{I}\}$ each of which is measurable with respect to a space $\{\Omega, \mathcal{B}\}$, where $\mathbf{I}$ serves an index set. One player, the <u>*bookie*</u>, posts a *fair*, or *2-sided* prevision $P(X_i)$ for each $X_i \in \mathcal{X}$. The bookie's opponent, the <u>*gambler*</u>, may choose *finitely many* non-zero real numbers $\{\alpha_i\}$ where, when the state $\omega \in \Omega$ obtains, the bookie's payoff is $\Sigma_i\alpha_i(X_i(\omega) – P(X_i))$, and the gambler's payoff is the negative of this quantity, $-\Sigma_i\alpha_i(X_i(\omega) – P(X_i))$. That is, the bookie is obliged either to buy (if $\alpha > 0$), or to sell (if $\alpha < 0$) $|\alpha|$-many units of X at the price, $P(X)$. Hence, the previsions are described as being *2-sided* or *fair* buy/sell prices.

The bookie's previsions are <u>*incoherent*$_1$</u> if the gambler has a strategy that insures a uniformly negative payoff for the bookie, i.e., if there exist a <u>finite set</u> $\{\alpha_i\}$ and $\varepsilon > 0$ such that, for each $\omega \in \Omega$, $\Sigma_i\alpha_i(X_i(\omega) – P(X_i)) < -\varepsilon$. Otherwise, the bookie's previsions are <u>*coherent*$_1$</u>.

De Finetti's *Fundamental Theorem of Previsions*:

The bookie's previsions $\{P(X): X \in \mathcal{X}\}$ are coherent$_1$ if and only if there is a finitely additive probability P whose expected value for X, $\mathbf{E}_P[X]$, is the *bookie*'s prevision. That is:

- *Coherence*$_1$ obtains *if and only if* $\mathbf{E}_P[X] = P(X)$.

This result extends to include *coherence*$_1$ for conditional expectations given non-null events, using the device of called-off previsions. Let F be an event with $F(\omega)$ its indicator function. The bookie's called-off prevision, $P_F[X]$, for X given event F has payoff in state ω to the bookie: $F(\omega)\alpha(X(\omega) - P_F(X))$, which equals 0 – the transaction is called-off – in case event F fails. Assuming that the conditioning event is not null, i.e., $P(F) \neq 0$, then

- *Coherence*$_1$ for called-off previsions requires that $\mathbf{E}_P[X \mid F] = P_F[X]$.

When the conditioning event F is null, coherence$_1$ places no substantive constraints on the called-off prevision $P_F[X]$. That is $E_P[F(\omega)\alpha(X(\omega) - P_F(X))] = 0$ regardless the real-value of $P_F[X]$. This defect in de Finetti's formulation has been discussed many times in the literature, and with a variety of different proposals to remedy the situation. For different corrections to this defect in coherence$_1$ see [7, 13, 16, and 28]. In our opinion, the debate over conditional probability given a null event is not yet resolved. The correctives to de Finetti's theory engender other controversies. For example, each of these three proposals underwrites Dubins' [6] theory of full conditional probabilities. But Dubins' theory of conditional probability produces an asymmetric relevance relation. (See [2].) Because such controversies about conditioning on null events do not arise for the basic questions about IP-coherence addressed in this essay, we use de Finetti's original version of coherence$_1$ and sidestep the important challenge of developing a satisfactory theory of conditional probability given a null event.

The problems we do address here are prompted by de Finetti's [3, 4] observation that *strategic* aspects of betting may affect <u>elicitation</u> of a bookie's *fair* previsions using the prevision game. For example, when the bookie (believes he/she) knows the gambler's betting odds, then <u>announcing</u> a prevision is subject to strategic play in the game and may fail to reveal the bookie's fair prevision.

<u>*Example* 1</u>: Suppose the bookie's *fair* (2-sided) prevision for an event G is .50. But suppose the bookie is confident the

gambler's fair prevision for G is .75. So the bookie announces $P(G) = .70$, anticipating that the gambler will find it profitable to buy units of G at the inflated price. *Elicitation* using the prevision game fails to identify the bookie's fair price for G. ◊

Note: There are other issues concerning elicitation in the prevision game. Among these is the challenge of state-dependent utilities [20], which we mention in Section 5.

To mitigate strategic aspects of the prevision game, de Finetti turned to a different coherence criterion: probabilistic forecasting of random variables subject to Brier score. In this essay, where our central goal is to discuss extensions of coherence$_2$ to Imprecise Probabilities for events, we focus on forecasting events, represented by their indicator functions. $E(\omega) = 1$ if $\omega \in E$ and $E(\omega) = 0$ if $\omega \notin E$.

The bookie's previsions serve as probabilistic forecasts subject to Brier score: squared-error loss. The penalty for the forecast $P(E)$ when $\omega \in \Omega$ is given by two functions $\{g_1, g_0\}$ depending upon the state:

$g_1(P(E), \omega) = (1 - P(E))^2$ if event $\omega \in E$ obtains;

$g_0(P(E), \omega) = (0 - P(E))^2$ if event $\omega \in E^c$ obtains, which is summarized by the squared-error penalty score

$$(E(\omega) - P(E))^2$$

For the conditional (called-off) forecast $P_F(E)$, on condition that event F obtains, the score is

$$F(\omega)(E(\omega) - P(E))^2.$$

And just as in the prevision game, the score for a finite set of forecasts is the sum of the separate scores.

The coherence$_2$ criterion applies to forecasting real-valued random variables, not just indicator functions.
Definition: A forecast set $\{P(X): X \in \mathcal{X}\}$ is *coherent$_2$* if, for each finite subset of $\mathcal{X}$, there is no rival forecast set $\{P'(X): X \in \mathcal{X}\}$ whose scores uniformly dominates in Ω.

The two senses of coherence are equivalent, as de Finetti established [3, Sections 3.3-3.4].
Proposition 1: A set of previsions are coherent$_1$ in the prevision-game *if and only if* those same set previsions are a coherent$_2$ set of forecasts under Brier score.

Proof: Here is a geometric version of de Finetti's projection-argument that establishes coherence$_1$ and coherence$_2$ are equivalent coherence criteria. We sketch his argument applied to previsions/forecasts for a complementary pair events. We use the same geometric presentation in Section 3 in order to extend coherence$_2$ to an IP setting.

Let $\mathcal{X} = \{X_1, X_2\}$ be a pair of variables where X_1 is the indicator for an event A and X_2 is the indicator for the complementary event A^c. In Figure 1, below, a pair of forecasts, $\{Q(A), Q(A^c)\}$ with $0 \leq Q(A), Q(A^c) \leq 1$, is depicted by the point $(Q(A), Q(A^c))$ in the unit square. The Brier score for a pair of such forecasts depends upon the two possible values of the indicators $\{X_1, X_2\}$, which are represented by the two points: (1,0), if the event A obtains, and (0,1) if the event A^c obtains. The Brier score for the pair of forecasts $\{Q(A), Q(A^c)\}$ equals the square of the Euclidean distance between the point $(Q(A), Q(A^c))$ and the respective point, either (1,0) or (0,1), depending upon which of the two possible values of the indicators $\{X_1, X_2\}$ obtains.

A forecast pair $\{Q(A), Q(A^c)\}$ is incoherent$_2$ if there is some rival forecast pair $\{Q'(A), Q'(A^c)\}$ whose Brier score is smaller regardless the realized values of the variables $\{X_1, X_2\}$. Thus, the rival forecast pair dominates if and only if the distance between the two points $(Q'(A), Q'(A^c))$ and (1,0) is less than the distance between the two points $(Q(A), Q(A^c))$ and (1,0), and likewise for the respective distances to the point (0,1).

The coherent$_1$ forecasts lie along the reverse diagonal, the simplex on two states, where $Q(A) + Q(A^c) = 1$. No such point is dominated in Brier score by any other coherent$_1$ forecast, since moving along this line segment increases the distance, and hence increases the squared error relative to one endpoint or the other.

Example 2: Consider, the incoherent$_1$ previsions: $P(A) = .6$ and $P(A^c) = .7$. A *Book* is achieved against these previsions with the gambler's strategy $\alpha_1 = \alpha_2 = 1$. Then the net payoff to the bookie is -0.3 regardless which state ω obtains. In order to see that these are also incoherent$_2$ forecasts, review Figure 1. ◊

If the forecast previsions $\{Q(A), Q(A^c)\}$ are not coherent$_1$, they lie outside the probability simplex. Project these incoherent$_1$ forecasts into the simplex. As in Example$_2$, the point (.60, .70) projects onto the coherent$_1$ previsions at the point (.45, .55). By elementary properties of Euclidean projection, the resulting pair of coherent$_1$ forecasts, represented by the point (.45, 55), are closer to each endpoint of the simplex than is the pair of incoherent forecasts, represented by the point (.60, .70). Thus, the projected forecasts have a dominating Brier score with respect to the binary partition of possible values for the variables $\{X_1, X_2\}$. This establishes that the initial forecasts are incoherent$_2$. Since no coherent$_1$ forecast set can be so dominated, we have coherence$_1$ of the previsions if and only we have coherence$_2$ of the corresponding forecasts, as required by *Proposition 1*.

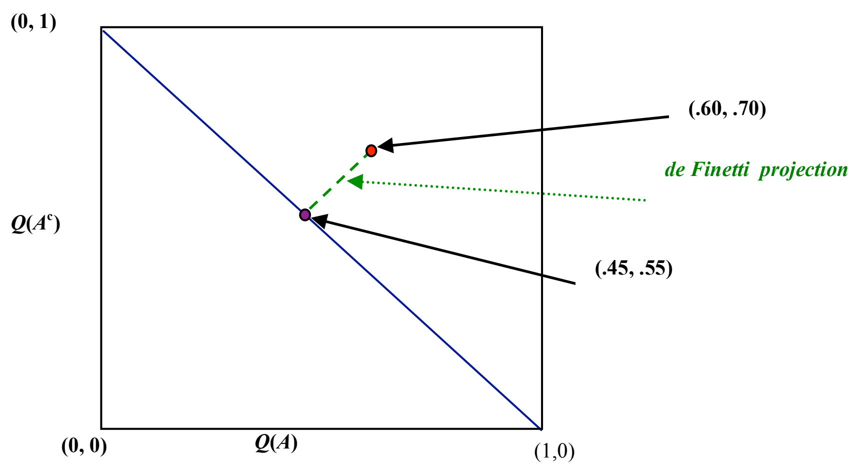

Figure 1

Notes: If either forecast is outside the unit interval, then it is outside the range for the variable being forecasted. Then it is trivial to dominate that single forecast with a rival forecast chosen to be closer to the nearest endpoint of the range of the variable in question. Also, just as coherence$_1$ fails to regulate called-off previsions given a null event, coherence$_2$ does not regulate called-off forecasts given a null event. See [7] for a parallel revision to coherence$_2$ in order to accommodate conditional forecasts given a null event.

2.2 Incentive Compatible Scoring

Brier score is just one of an infinite class of (*strictly*) *proper* scoring rules.

Definition: A scoring rule is (*strictly*) *proper* just in case a forecaster (uniquely) minimizes expected score by announcing her/his previsions.

Thus, forecasting with a (strictly) proper scoring rule avoids the problem of strategic behavior present in the prevision game: there is no opponent. Even allowing different proper scoring rules for different forecasts, by taking the combined score for a finite set of forecasts as the sum of the individual scores, the result is again (strictly) proper. Savage [18] and Schervish [19] characterize the (g_0, g_1) pairs for proper scoring rules. In [21] we establish that all (proper) scoring rules produce the same distinction between coherent$_1$ and incoherent$_1$ forecasts as with Brier score, both for unconditional forecasts and for conditional forecasts given a non-null event.

Proposition 2 [21]:

(*2.1*) When the scoring rule is proper, finite, and continuous, each incoherent$_1$ forecast set is dominated by some coherent$_1$ forecast set.

(*2.2*) When the scoring rule is proper, finite, but not continuous, each incoherent$_1$ forecast set is dominated, but not necessarily by a coherent$_1$ forecast set.

Notes: Result 2.1 can be established by a generalization of de Finetti's geometric argument, where the projection depends upon the scoring rule. See [15]. Gilio and Sanfilippo [8] use a strengthened coherence criterion to extend this analysis to continuous scoring rules when there is conditioning on null events. The demonstration of result (2.2) in [21] uses game-theoretic reasoning.

3. Coherence$_2$ with a Brier *IP* scoring rule.

One introduction to Imprecise Probabilities is provided by C.A.B.Smith's [26] modification of de Finetti's prevision game, which provides a criterion of IP-coherence$_1$ for (closed, convex) IP sets. Rather than requiring a 2-sided, *fair* price, the bookie may fix a pair of 1-sided previsions for each $X \in \mathcal{X}$: the bookie may fix separate *buy* and *sell* prices.

- The bookie announces one rate $\underline{P}(X)$ as a *buying* price for use when $\alpha > 0$, and a possibly different *selling* price $\overline{P}(X)$ for use when $\alpha < 0$.

The result is a generalized *Book* argument. See [279, chapter 2] for some history and basic results.

Proposition 3:
(*3.1*) A bookie's 1-sided previsions *avoid sure loss* if and only if there is a maximal, non-empty (closed, convex) set of finitely additive probabilities $\mathcal{P}$ where

$\underline{P}(X) \leq \textit{infimum}_{P \in \mathcal{P}} \mathbf{E}_P[X]$

And $\overline{P}(X) \geq \textit{supremum}_{P \in \mathcal{P}} \mathbf{E}_P[X]$.

When these inequalities are equalities, the 1-sided previsions are said to be *IP-coherent*$_1$.
(*3.2*) By requiring lower and upper previsions for sufficiently many variables (from the linear span of $\mathcal{X}$), the 1-sided previsions avoid sure loss if and only if they are also IP-coherent$_1$. See Theorem 1.ii of [23].

We offer a parallel version for defining IP-coherence$_2$ based on Brier score for 1-sided forecasts, as follows:
Use a *lower forecast* to assess a penalty score when the event forecasted *fails*;
Use an *upper forecast* to assess a penalty score when the event forecasted *obtains*.

Let $\{E_i: i = 1, \ldots, m\}$ be m events defined over a finite partition $\Omega = \{\omega_j: j = 1, \ldots, n\}$. The forecaster gives *lower* and *upper* probability forecasts $\{p_i, q_i\}$ for each event E_i.

Scoring forecasts with a Brier-styled IP scoring rule:
Fix a state $\omega \in \Omega$.
If $\omega \in E_i$ the score for the forecast of E_i is $(1-q_i)^2 = g_1(q_i, \omega)$
If $\omega \notin E_i$ the score for the forecast of E_i is $p_i^2 = g_0(p_i, \omega)$
That is, use the most favorable forecast value from the pair $\{p_i, q_i\}$ for determining the score. Just as with the other coherence criteria discussed here, the score for a set of forecasts is the sum of the individual forecast scores.

Dominance:
A forecast set $\mathcal{G}$ (*strictly*) *dominates* another $\mathcal{F}$ if, for each $\omega \in \Omega$, the score for $\mathcal{G}$ is (strictly) less than the score for $\mathcal{F}$.

But, since the vacuous $\{0 = p_i, q_i = 1\}$ forecast dominates each rival $\{0 < p_i', q_i' < 1\}$, we require an additional restriction on the class of competing forecasts in order to avoid triviality of the resulting theory of IP-coherence$_2$.
Note: This is analogous to a problem that is usually ignored within traditional IP theory. With 1-sided previsions, it remains IP-coherent$_1$ to be strategic: announce a lower buying (and/or a higher selling) price than one is prepared to accept. That is, knowing who is the *Gambler* in the 1-sided Prevision Game, the *Bookie* may play strategically and mimic having a less determinate IP-coherent$_1$ set of previsions in order to secure strictly favorable gambles.

We propose that *IP-coherence*$_2$ takes into account both a *rival model class M*, which identifies the competing class of rival forecasts, and an index of *relative imprecision* in a forecast set. By allowing the rival forecasts to be restricted to a particular class *M*, we offer a more general approach than when the rival class is fixed as the maximal class of all possible lower and upper forecasts. This flexibility permits, also, to link our approach to different theories of Robust Statistics, as illustrated in Example 3, below.

Stated informally, a set of 1-sided forecasts $\mathcal{F}$ are incoherent$_2$ when there exists a dominating set of forecasts $\mathcal{G}$ that are
(i) at least as precise/determinate as $\mathcal{F}$ and
(ii) where $\mathcal{G}$ belongs to the model class *M*.
We illustrate this idea by filling in the details of the two concepts: the *rival model class M* and *relative informativeness* between forecast sets using the ε-*contamination* class.

Example 3: Set *M* equal to the ε-*contamination* class, defined as follows. Let *P* be a particular probability distribution over $\Omega = \{\omega_1, \ldots, \omega_n\}$. Fix $0 \leq \varepsilon \leq 1$. Let $\boldsymbol{Q}$ be the simplex of all probability distributions on Ω. The ε-contamination model, $\mathcal{P}_\varepsilon$, with focus on the distribution *P*, is the set of probability distributions on Ω defined by $\mathcal{P}_\varepsilon = \{(1-\varepsilon)P + \varepsilon Q$:

$Q \in \boldsymbol{Q}\}$. This model is popular in studies of Bayesian Robustness. (See Huber [11, 12] and Berger [1].) Also, it is the model obtained from Harsanyi and Selten's [9] "trembling hand" strategies where P is the target strategy which can be achieved with probability 1-ε; otherwise, with probability ε strategy Q obtains. As a third reason for illustrating our ideas with the ε-contamination model is that the lower probability function from this model is also a Dempster-Shafer *Belief Function*, and updating the ε-contamination model either by Bayes' rule or by Dempster's rule yields the same results. For our purposes here, it is useful to know that this class is characterized by specifying (IP-coherent$_1$) lower probabilities *for atomic events*, and then using the largest closed convex set of distributions satisfying those bounds. (See Seidenfeld [22].)

IP-forecasts over a finite partition for Brier-styled, ε-contamination coherence$_2$:

Let $\mathcal{F} = \{ \{p_i, q_i\}: i = 1, \ldots, n\}$ be forecasts for each state $\omega_i \in \Omega = \{\omega_1, ..., \omega_n\}$.

Define $\mathcal{F}$'s score set $\mathcal{S}$ by an ordered n-tuple of n-dimensional points:
$\mathcal{S} = \{(q_1, p_2, \ldots, p_n), (p_1, q_2, \ldots, p_n), \ldots, (p_1, p_2, \ldots, q_n)\}$.
Thus, $\mathcal{S}$ contains at most n-many distinct points. Each point in $\mathcal{S}$ has n-many coordinates.

Observe that the *IP-Brier-style* score for $\mathcal{F}$ evaluated at state ω_j is the square of the Euclidean distance between the j^{th} point of $\mathcal{S}$ and the j^{th} corner of the probability simplex on Ω. Clearly, the *IP*-score for a forecast set can be improved merely by moving a lower forecast closer to 0, or by moving an upper forecast closer to 1. So, consider dominating forecast sets only when the dominating forecast has a score set that is *less indeterminate* than the score set for the dominated forecast. Here is a candidate for *relative indeterminacy* which, when combined with our Brier-style IP-score, allows a characterization of *ε-contamination* IP-coherence$_2$.

Definition: Forecast set $\mathcal{F}_2$ is at least as indeterminate as forecast set $\mathcal{F}_1$ (or $\mathcal{F}_1$ is at least as determinate as $\mathcal{F}_2$) if the convex hull of score set $\mathcal{S}_1$, $\boldsymbol{H}(\mathcal{S}_1)$, is isomorphic under rigid movements (where both shape and size are held fixed) to a subset of the convex hull of score set $\mathcal{S}_2$, $\boldsymbol{H}(\mathcal{S}_2)$.

Note that this relation of *relative imprecision*, or *relative indeterminacy*, is merely a partial order. We opt for such a concept so that relative indeterminacy may be extended to a variety of different real-valued indices of imprecision, e.g., by using generalized volume of the score set to quantify indeterminacy.

We use these notions to define IP-coherence$_2$ generally, and then continue with our illustration of IP-coherence$_2$ with respect to the ε-contamination model.
Definition: Given an *IP*-scoring rule, *a* set $\mathcal{F}$ of IP-forecasts is *IP-incoherent$_2$ with respect to the model M* provided that there is a dominating set of rival forecasts $\mathcal{G}$ from the model *M* where the set $\mathcal{G}$ is at least as determinate than the set $\mathcal{F}$. Say that $\mathcal{F}$ is IP-coherent$_2$ with respect to *M* if it is not IP-incoherent$_2$ with respect to *M*. For convenience we will write these as *M*-coherent$_2$ and *M*-incoherent$_2$

Observe that IP-incoherence$_2$ reduces to de Finetti's incoherence$_2$ when all forecasts in $\mathcal{F}$ are determinate, i.e., when $p_i = q_i$ for each forecasted event E_i $(i \in \mathbf{I})$, and when *M* is the class of all determinate, coherent$_1$ forecasts. To see this, assume that $|\Omega| = k$. Then the score set $\mathcal{S}$ is the ordered set with k-many repetitions of the same $|\mathbf{I}|$-dimensional point. Since the lower and upper $\mathcal{F}$ forecasts for an event are identical, the k-many points in $\mathcal{S}$ do not vary with ω. So a dominating rival forecast set $\mathcal{G} = \{p'_i, q'_i\}$ must also assign the same lower and upper values to each event E_i (that is, for each $i \in \mathbf{I}$, $p'_i = q'_i\}$, in order for $\mathcal{G}$ to be at least as determinate as $\mathcal{F}$. By *Proposition* 2.1, then if $\mathcal{G}$ dominates $\mathcal{F}$ the rival forecast set $\{p_i'\}$ establish that $\mathcal{F}$ is incoherent$_2$ and incoherent$_1$.

Next, we provide two basic results for IP-coherence$_2$ with respect to the ε-contamination model.

Proposition 4: Let $0 \le p_i \le q_i \le 1$, with n-many forecasts $\mathcal{F}$ solely *for atoms* of the algebra, the elements of the partition $\Omega = \{\omega_1, \ldots, \omega_n\}$.

(*4.1*) The score set $\mathcal{S}$ for $\mathcal{F}$ lies entirely within the probability simplex on Ω *if and only if* the lower and upper forecasts $\mathcal{F}$ match an ε-*contamination* model. And then $\mathcal{F}$ cannot be dominated by rival forecasts from a more determinate ε-contamination model.

(*4.2*) If all the elements of a score set $\mathcal{S}$, associated with forecast set $\mathcal{F}$, lie outside the probability simplex on Ω, there is a dominating ε-contamination forecast model $\mathcal{F}^*$ with greater determinacy than $\mathcal{F}$. $\mathcal{F}$ is IP-incoherent$_2$ against rivals from the ε-contamination model.

Proof:
Result (*4.1*) is established by elementary calculations. If and only if each point of the score set $\mathcal{S}$ belongs to the probability simplex then, when state ω_j obtains, corresponding to the j^{th} point of $\mathcal{S}$, $1 = q_j + \sum_{i \neq j} p_i$. This equality obtains for each $j = 1, \ldots, n$. Then there exists an $\varepsilon \geq 0$ such that for each $i = 1, \ldots, n$, $q_i = p_i + \varepsilon$, which defines an ε-contamination model. In the opposite direction, if forecasts for the atoms are based on an ε-contamination model, for $i = 1, \ldots, n$, $q_i = p_i + \varepsilon$, and then $1 = q_j + \sum_{i \neq j} p_i$ so that all of the score set $\mathcal{S}$ lies in the probability simplex.

Last, if $\mathcal{S}$ belongs to the probability simplex and a rival ε-contamination model $\mathcal{F}$' (with corresponding score set $\mathcal{S}$') dominates, then $\boldsymbol{H}(\mathcal{S})$ is a proper subset of $\boldsymbol{H}(\mathcal{S}')$ because for each $j = 1, \ldots, n$, the j^{th} point of $\mathcal{S}$' is closer to the j^{th} extreme point of the probability simplex than is the j^{th} point of $\mathcal{S}$. So, $\mathcal{F}$' is *less* determinate than $\mathcal{F}$. Thus $\mathcal{F}$ is IP-coherent$_2$ with respect to the ε-contamination model.

Result (*4.2*) follows by the *Brouwer Fixed-Point* Theorem. Begin with a forecast set $\mathcal{F} = \mathcal{F}_0$, whose score set $\mathcal{S}_0$ has each of its n-many ordered points outside the simplex of coherent$_1$ forecasts. Recursively create rival forecast sets as follow. Apply the (de Finetti) projection to each of these n-many ordered points of $\mathcal{S}_0$ taking them into the probability simplex of coherent$_1$ forecasts. This creates a set of (at most) n-points $\boldsymbol{T}_1 = \{t_1, \ldots, t_n\}$ where each $t \in \boldsymbol{T}_1$ is a probability distribution $P(\bullet)$ over Ω. Form the new forecast set $\mathcal{F}_1 = \{\{p_{1i}, q_{1i}\}: i = 1, \ldots, n\}$ where $p_{1i} = min_{t \in \boldsymbol{T}_1}\{P(\omega_i)\}$ and $q_{1i} = max_{t \in \boldsymbol{T}_1}\{P(\omega_i)\}$. This determines a new score set $\mathcal{S}_1$. Since none of the points in $\mathcal{S}_0$ belongs to the probability simplex, by the same reasoning used in de Finetti's analysis for *Proposition* 1, $\mathcal{F}_1$ dominates $\mathcal{F}_0$.

Just in case $\mathcal{S}_1$ lies in the simplex, when result (4.1) applies, the recursive procedure halts. Otherwise forecast set $\mathcal{F}_2$ is created from a projection of score set $\mathcal{S}_1$ into the probability simplex, etc.

Since Euclidean projections are continuous functions and the probability simplex is compact, the recursive process with forecast sets $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$. has a fixed point $\mathcal{F}^*$ in the class of ε-contamination models. By a simple adaptation of de Finetti's argument for *Proposition* 1, the forecast set $\mathcal{F}_{i+1}$ (weakly) dominates the forecast set $\mathcal{F}_i$ unless $\mathcal{F}_i$ is a fixed point of the process.

Note: It may be that $\mathcal{F}_{i+1}$ merely weakly dominates $\mathcal{F}_i$ for $i \geq 1$, since some but not all the points in $\mathcal{S}_1$ may lie in the probability simplex. However, since all the points of $\mathcal{S}_0$ lie outside the probability simplex, $\mathcal{F}_1$ dominates $\mathcal{F}_0$.

Last, the projection of a closed, convex set, e.g., the projection of $\boldsymbol{H}(\mathcal{S})$ into the probability simplex, is isomorphic to a subset of $\boldsymbol{H}(\mathcal{S})$. Thus, assuming that the each of the points of $\mathcal{S}_0$ is outside the probability simplex on Ω, the fixed point $\mathcal{F}^*$ of the process $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$, which belongs to the ε-contamination model class, strictly dominates $\mathcal{F}_0$, and is at least as determinate as $\mathcal{F}_0$. Hence, $\mathcal{F}_0$ is IP-incoherent$_2$ with respect to the ε-contamination class.▯

<u>*Example* 4</u>: Here is an illustration of *Proposition* 4, IP-coherence$_2$ with respect to the ε-contamination model, using 5 different forecast sets. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Forecasts are for the three atoms only. The five forecast sets $\mathcal{F}^j$ ($j = 1, \ldots, 5$) are presented in the form $\{\{p_i, q_i\}$ for ω_i: $i = 1, 2, 3\}$. The respective score sets have three points with coordinates $\{(q_1, p_2, p_3), (p_1, q_2, p_3), (p_1, p_2, q_3)\}$, as described above.

Figure 2 diagrams the convex hull of each score set and shows the shaded 2-dimensional, triangular simplex of probability functions on Ω.

Example$_4$: Score set S_2 dominates the other four score sets and is at least as determinate as each of them

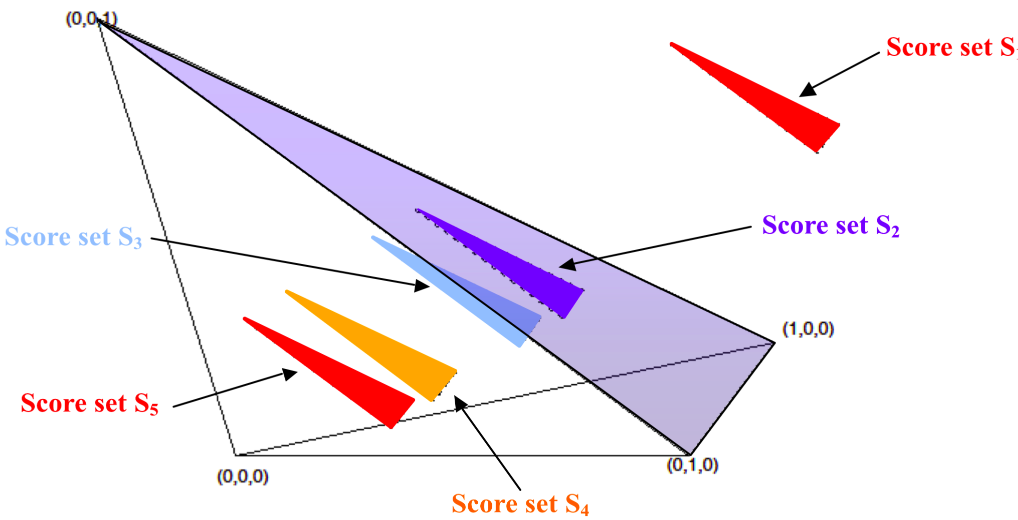

Figure 2 (for *Example* 4)
The convex hull of the five score sets are color coded. The simplex of probability distributions is shaded. Each score set projects onto $\mathcal{S}^2$, the score set for forecast set $\mathcal{F}^2$, corresponding to an ε-contamination model.

$\mathcal{F}^1$ = { {.55, .80}, {.55, .80}, {.55, .80}}
$\mathcal{S}^1$ = {(.80, .55, .55), (.55, .80, .55), (.55, .55, .80)}

$\mathcal{F}^2$ = { {.25, .50}, {.25, .50}, {.25, .50}}
$\mathcal{S}^2$ = {(.50, .25, .25), (.25, .50, .25), (.25, .25, .50)}

$\mathcal{F}^3$ = { {.20, .45}, {.20, .45}, {.20, .45}}
$\mathcal{S}^3$ = {(.45, .20, .20), (.20, .45, .20), (.20, .20, .45)}

$\mathcal{F}^4$ = { {.10, .35}, {.10, .35}, {.10, .35}}
$\mathcal{S}^4$ = {(.35, .10, .10), (.10, .35, .10), (.10, .10, .35)}

$\mathcal{F}^5$ = { {.05, .30}, {.05, .30}, {.05, .30}}
$\mathcal{S}^5$ = {(.30, .05, .05), (.05, .30, .05), (.05, .05, .30)}

The two forecast sets $\mathcal{F}^1$ and $\mathcal{F}^5$ are IP-incoherent$_1$ in accord with *Proposition* 3. Their 1-sided previsions lead to sure losses as, respectively, their lower (upper) forecasts are too great (too small). There is no determinate probability distribution agreeing with either set's lower and upper forecasts.

Forecast set $\mathcal{F}^2$ corresponds to an ε-contamination model with focus on the uniform probability P = (1/3, 1/3, 1/3) and ε = 1/4. The convex hull of the score set $\mathcal{S}^2$ lies in the probability simplex, as per *Proposition* (4.1). It is IP-coherent$_1$ and IP-coherent$_2$ with respect to the ε-contamination model class.

Forecast set $\mathcal{F}^3$ is IP-coherent$_1$ as it has lower and upper forecasts agreeing with a closed convex set of probabilities. Those values agree with a rival *ALUP* model, but not with an ε-contamination model. That is, $\mathcal{F}^3$ is IP-coherent$_2$ with respect to an IP-model class defined by specifying **a**tomic **l**ower and **u**pper **p**robabilities [ALUP], but not so with respect to the ε-contamination class, which is an IP-model class determined solely by atomic lower probabilities. (See Appendix 1 for additional details about the ALUP model.)

Forecast set $\mathcal{F}^4$ has lower and upper forecasts that do not match those from a closed convex set of probabilities. Its intervals are too wide. However, the uniform probability agrees with these forecasts, i.e., the probability values (1/3, 1/3, 1/3) fall inside the forecast intervals from $\mathcal{F}^4$. Thus, in accord with *Proposition* 3, the forecasts from $\mathcal{F}^4$ do not suffer a sure-loss in the 1-sided prevision game; however, $\mathcal{F}^4$ is IP-incoherent$_1$ and IP-incoherent$_2$ with respect to the ε-contamination model class.

As indicated in Figure 2, each of the other four convex hulls projects to $\boldsymbol{H}(\mathcal{S}^2)$. That is, the process described in the proof of Proposition (4.2) has $\mathcal{F}^2$ as its fixed point for each of the five forecast sets, and the process terminates after at most one projection. (See Appendix 2 for an illustration of Proposition (4.2) where the fixed point is merely a limit of the process.)

4. Incentive compatible IP-elicitation

Recall that de Finetti favored coherence$_2$ over coherence$_1$ because, in addition to serving as an equivalent criterion of coherence, Brier score provides a *strictly proper* score. A decision maker who maximizes expected utility against Brier score announces her/his previsions for random variables as their forecasts. Brier score is an incentive compatible elicitation for determinate probabilities. It eliminates some of the strategic aspects evident in the prevision game. That is, for a decision maker whose degrees of belief about events are represented by a single probability function $P(\bullet)$ and who maximizes expected utility, she/he has a unique strategy for announcing forecasts (and called-off forecasts) that minimize expected Brier score. Announce the probability $P(E)$ as the forecast of event E. If H is not-null, then announce the conditional probability $P(E\,|H)$ for the called-off forecast of event E, on condition that H obtains.

Recall that when H is null, coherence$_2$ places no restrictions on the called-off forecasts given H. There is no difference to the expected score contributed by any conditional forecast of E, called-off if H fails, regardless whether that forecast is or is not coherent$_2$. See [5] for an improved version of coherence$_2$. However, our presentation in this section is not affected by the open problem of how to resolve the problem of reducing conditional probability given a null event to a decision-theoretic criterion of coherence.

What can be done to extend Brier score to an incentive compatible IP-scoring rule? The question is ill-formed without a decision rule that extends maximizing expected utility to IP decisions. That is, whether a particular IP-scoring rule is *proper* or not, depends upon what decision rules we allow the IP decision maker to use. The IP community has not agreed on the answer to this question. Here, we consider only decision rules that reduce to the rule of maximizing expected utility when IP sets of probabilities collapse onto the special case of a singleton set, where upper and lower probabilities are identical and a single probability distribution represents uncertainty. Also, we require that decision rules respect the following weak form admissibility.

Let $\boldsymbol{S}(\mathcal{F}, \omega)$ be a real-valued IP-scoring rule for forecast set $\mathcal{F}$ in state ω. Recall that scores are given in the form of a loss so that smaller is better.

> *Admissibility Principle*: If for each $\omega \in \Omega$ $\boldsymbol{S}(\mathcal{F}, \omega) \leq \boldsymbol{S}(\mathcal{F}', \omega)$, then $\mathcal{F}$ is admissible in a pairwise choice between rival forecasts $\mathcal{F}$ and $\mathcal{F}'$. Moreover, if for each ω this inequality is strict then $\mathcal{F}'$ is inadmissible whenever $\mathcal{F}$ is an option.

In this section we report two results concerning existence of proper scoring rules for eliciting upper and lower probabilities for events, when the forecaster's opinion is represented by a closed, convex sets of probabilities on a finite state space and decisions conform to the Admissibility Principle. The first of the two, *Proposition* 5, establishes that there is no real-valued IP-counterpart to a continuous scoring rule, such as Brier score.

Proposition 5: There is no *real-valued* strictly proper IP continuous scoring rule.

By contrast, *Proposition* 6, if one considers scoring rules with non-standard values, then strictly proper IP-scoring rules exist for each of two IP-decision rules. The IP-decision rules we investigate in connection with *Proposition* 6 are summarized as follows, with additional details given in Section 4.2:

Γ-*Maximin*: The admissible options in a decision problem **D** are those that maximize their lower expected value.

E-*admissibility*: An option $X \in \mathbf{D}$ is **E**-*admissible* if for some $P \in \mathcal{P}$ and each $Y \in \mathbf{D}$, $\mathbf{E}_P[X] \geq \mathbf{E}_P[Y]$.
E-*admissibility-followed-by*-Γ-*Maximin*: Apply Γ-*Maximin* to the set of **E**-*admissible* options in **D**.

Proposition 6: Under either the Γ-*Maximin* decision rule, or using one of Levi's [13] lexicographic decision rules – **E**-*admissibility* followed by Γ-*Maximin* security – there is a strictly proper *lexicographic* IP-Brier scoring rule.

Next, we establish and explain these findings.

4.1 Proof of *Proposition* 5 The impossibility reported in this result is made evident by considering the demands on a real-valued strictly proper IP-scoring rule $\boldsymbol{S}(\mathcal{F}', \omega)$, for forecasting one event, E.
Let the interval $[p, q]$, $0 \leq p \leq q \leq 1$, represent the forecaster's uncertainty for E. In general, the IP-scoring rule may be written

$\boldsymbol{S}([x, x], \omega) = g_1([p, q], \omega)$ if $\omega \in E$ obtains,
and $\boldsymbol{S}([x, x], \omega) = g_0([p, q], \omega)$ if $\omega \in E^c$ obtains.

When $p = q$, in order to be strictly proper and real-valued, the scoring rule must satisfy Theorem 4.2 of Schervish [12]. Specifically, with $0 \leq x \leq 1$, the loss for the point forecast $\boldsymbol{S}([x, x], \omega)$, x satisfies

$$g_1(x) = g_1(1) + \int_x^1 (1-q)\lambda(dq) \quad \text{if } \omega \in E \text{ obtains;}$$

$$g_0(x) = g_0(0) + \int_0^x q\lambda(dq) \quad \text{if } \omega \in E^c \text{ obtains,}$$

where $g_1(1)$ and $g_0(0)$ are finite, and $\lambda(dq)$ is a measure on $[0, 1)$ that gives positive measure to every non-degenerate interval. Continuity of the scoring rule results from a continuous measure λ with no point masses. For example, Brier score results by letting λ have the constant density 2 on the unit interval.

When $p < q$, the impossibility of a strictly proper continuous IP-scoring rule is a consequence of the fact that, since λ is positive on non-degenerate sub-intervals of the unit interval [0,1] and continuous, and as there is no continuous 1-1 map between the unit square and unit interval, there will be rival interval forecasts $[p, q]$ and $[p', q']$ with

$g_1([p, q]) - g_1([p', q']) \geq 0$,
and $g_0([p, q]) - g_0([p', q']) \geq 0$.

Then the interval forecast $[p', q']$ is admissible against the rival interval forecast $[p, q]$. When the interval $[p, q]$ is the forecaster's IP-uncertainty for event E, she/he will not have reason to announce that interval as her/his forecast rather than the rival forecast $[p', q']$. Thus, the IP-scoring rule is not strictly proper. If each of the two inequalities is strict, as illustrated in Examples 5 and 6 (below), then the IP-scoring rule is not even proper.□

Example 5. We illustrate *Proposition* 5 using the ideas about IP-coherence$_2$ presented in Section 3. Consider Brier score adapted to a forecast interval $[p, q]$ using the favorable end of the forecast interval. That is, let

$\boldsymbol{b}([p,q], \omega) = g_1([p, q], \omega) = (1-q)^2$ if $\omega \in E$,
and $\boldsymbol{b}([p,q], \omega) = g_0([p, q], \omega) = p^2$ if $\omega \in E^c$.

Introduce a *real-valued* index of indeterminacy for a forecast set $\mathcal{F}$, $\boldsymbol{I}(\mathcal{F})$, where $\boldsymbol{I}$ agrees with the partial order of relative imprecision used to define IP-coherence$_2$. For instance, let $\boldsymbol{I}([p, q]) = q-p$. For real values x, y, let $\boldsymbol{H}(x,y)$ be a real-valued function increasing in each of its arguments, e.g., $\boldsymbol{H}(x,y) = x + y$. Define an IP-Brier score for forecast set $\mathcal{F}$ by $\boldsymbol{B}(\mathcal{F},\omega) = \boldsymbol{H}(\boldsymbol{b}(\mathcal{F},\omega), \boldsymbol{I}(\mathcal{F}))$. Then by *Proposition* 5, $\boldsymbol{B}$ is an improper-IP scoring rule. To complete the example, consider event E and compare the two interval forecasts [.25, .75] and [.50, .50].

Then $\boldsymbol{B}([.25, .75], \omega) = 1/16 + 1/2 = 9/16$
and $\boldsymbol{B}([.50, .50], \omega) = 1/4 + 0 = 1/4$,

all independent of ω. Hence, the interval forecast [.25, .75] is inadmissible under this IP-Brier scoring rule $\boldsymbol{B}$. That is, under this variant of IP-Brier scoring rule, a decision maker whose IP-uncertainty about the event E is given by the closed interval of probabilities [.25, .75] strictly prefers announcing the point-valued forecast [.50, .50] instead of her/his IP-interval, [.25, .75].

4.2 Proof of *Proposition* 6 First we review the two decision rules mentioned in the result. Let $\mathcal{P}$ be a closed, convex set of probabilities P on the space $\{\Omega, \mathcal{B}\}$. Let χ be the class of bounded random variables, X, each measurable with respect to this space. For each X, write $\underline{X}$ for the *infimum* over $\mathcal{P}$ of the expected value of X,

$$\underline{X} = inf_{P \in \mathcal{P}}\ \mathbf{E}_P[X],$$

which identifies the *lower expected value* for X with respect to $\mathcal{P}$. Identify a decision problem, **D**, with a closed subset of χ. That is, the options in a decision problem form a closed set of bounded variables.

The two IP-decision rules we investigate in *Proposition* 6 are defined as follows:

Γ-*Maximin*: The admissible options in **D** are those that maximize their lower expected value.
Note: By making both $\mathcal{P}$ and **D** closed sets, this *max-min* operation is well defined.

E-*admissibility*: An option $X \in \mathbf{D}$ is **E**-*admissible* if for some $P \in \mathcal{P}$ and each $Y \in \mathbf{D}$, $\mathbf{E}_P[X] \geq \mathbf{E}_P[Y]$.
E-*admissibility-followed-by*-Γ-*Maximin*: Apply Γ-*Maximin* to the set of **E**-*admissible* options in **D**.

In general, these decision rules have very different axiomatic characterizations. Γ-*Maximin* is represented by a real-valued ordering of χ using $\underline{X}$-values to index each option. But that ordering violates the independence axiom for preferences. **E**-*admissibility* is not represented by an ordering. In fact, it does not even reduce to pairwise comparisons. (See [24] for related discussion.) Nonetheless, next we construct a lexicographic IP-Brier score that is strictly proper under either of the two decision rules mentioned in *Proposition* 6.

Proposition 5 precludes a proper IP-scoring rule that elicits both endpoint of the interval forecast $[p,q]$ for event E. However, we may elicit either endpoint alone.
Define the *lower-Brier scoring rule*, $\underline{\mathbf{b}}([x,y], \omega) = \underline{\mathbf{b}}(x,\omega)$ as:

$g_1(x) = (1-x)^2$ if $\omega \in E$
$g_0(x) = 1 + x^2$ if $\omega \in E^c$.

and the *upper-Brier scoring rule*, $\overline{\mathbf{b}}([x,y], \omega) = \overline{\mathbf{b}}(y,\omega)$ as:

$\overline{g}_1(y) = (1-y)^2 + 1$ if $\omega \in E$
$\overline{g}_0(y) = y^2$ if $\omega \in E^c$.

Each of these is a strictly proper scoring rule for eliciting determinate forecasts. This follows immediately from Schervish's representation (above) where $g_1(1) = \overline{g}_0(0) = 0$, $g_1(0) = \overline{g}_1(1) = 1$, and $\lambda = 2$ is the uniform (Brier) score density for both rules.

Lemma 1: Under the Γ-*Maximin* decision rule, respectively, the lower- (upper-) Brier score is strictly proper for the lower (upper) endpoint of the IP-forecast $[p,q]$ of event E.
Proof of Lemma 1: We give the argument for the lower-Brier score. The reasoning for the upper-Brier score is similar. Let $p = min_{P \in \mathcal{P}}\ P[E]$ and $q = max_{P \in \mathcal{P}}\ P[E]$, so that $\forall P \in \mathcal{P}\ \ p \leq P(E) \leq q$, and these bounds are tight. The lower-Brier score of the forecast $[r, s]$ for E depends solely on r. The P-Expected score for forecast $[r, s]$ is:

$$\mathbf{E}_P[\underline{\mathbf{b}}[r, s]] = P(E)(1-r)^2 + (1-P(E))(1+r^2)$$
$$= (1-r)^2 + 2r(1-P(E)).$$

By simple dominance, $0 \leq r \leq 1$. For a given forecast r, this expected penalty score is greatest among $P \in \mathcal{P}$ at $P(E) = p$, when the expected score is $(1-r)^2 + 2r(1-p)$. But since lower-Brier score is strictly proper, this worst value is best, i.e., the worst of these expected scores is smallest uniquely for a forecast with $r = p$.□ Lemma 1

Lemma 2: Under the **E**-*admissibility-followed-by*-Γ-*Maximin* decision rule, respectively, the lower- (upper-) Brier score is strictly proper for the lower (upper) endpoint of the IP-forecast $[p,q]$ of event E.

Proof of Lemma 2: Again, we give the argument only for the lower-Brier score. Since lower-Brier score is a strictly proper scoring rule for determinate forecasts, the **E**-*admissible* forecasts are exactly those of the form $[r, s]$ where $p \leq r \leq q$, and $r \leq s$. That is, consider $P \in \mathcal{P}$ with $P(E) = r$. Only forecasts of the form $[r, s]$ maximizes the P-expected lower-Brier score. Hence, relative to lower-Brier score, the set of **E**-admissible forecasts with respect to the IP-set $\mathcal{P}$ are interval forecasts of the form $\{[r, s]: p \leq r \leq q$, and $r \leq s\}$. Then, by *Lemma* 1, the Γ-*Maximin* solution from this set is uniquely solved at $r = p$.□ Lemma 2

By *Proposition* 5, unfortunately, the real-valued composite score obtained by adding together these two scores, $\overline{\underline{\mathbf{b}}}([r,s]) = \underline{\mathbf{b}}([r,s]) + \overline{\mathbf{b}}([r,s]$, is not an IP-proper scoring rule, which we verify with Example 6.

Example 6: We illustrate the impropriety of the real-valued IP-score, $\overline{\underline{\mathbf{b}}}([r,s])$, in accord with *Proposition* 5. Consider an extreme case where the forecaster is maximally uncertain of event E, so that the vacuous probability interval [0, 1] represents her/his uncertainty. The precise forecast [.5, .5] has constant $\overline{\underline{\mathbf{b}}}$-score, i.e.,

$$\overline{\underline{\mathbf{b}}}([.5, .5], \omega) = 1 + ¼ + ¼ = 1.5,$$

which is independent of ω.
The straightforward forecast [0,1] has the constant score

$$\overline{\underline{\mathbf{b}}}([0, 1], \omega) = 1+1 = 2,$$

which also is independent of ω. So forecast [.5, .5] strictly dominates forecast [0,1] under the $\overline{\underline{\mathbf{b}}}$-scoring rule.◊

Instead of a real-valued IP-scoring rule, we use a 2-tier *lexicographical* (non-standard) composite scoring rule to combine these two scores in a manner that creates a strictly proper (but non-standard) IP-Brier score.
Definition: The two-tier, lexicographic IP-Brier score for the interval forecast $[p, q]$ of event E, which we write as $\mathbf{b}_{LU}([r, s])$, is the 2-tier lexicographic loss function

$$\mathbf{b}_{LU}([r, s], \omega) = < \underline{\mathbf{b}}([r, s], \omega),\ \overline{\mathbf{b}}([r, s], \omega) >.$$

That is, lexicographically, first apply the loss function $\underline{\mathbf{b}}([r, s])$, and among those forecasts have equal $\underline{\mathbf{b}}$-value, then apply the $\overline{\mathbf{b}}([r, s])$ loss function. By the preceding two lemmas, under the two decision rules named in *Proposition* 6, only the interval $[p, q]$ is $\mathbf{b}_{LU}$-optimal for forecasting event E when the forecaster's uncertainty for that event is the IP-interval $[p, q]$.□

Aside: It is evident that the order of the components is irrelevant in this 2-tiered, lexicographic IP-Brier score.

To elicit an IP-forecast set $\mathcal{F} = \{ \{p_i, q_i\}: i = 1, \ldots, n\}$ for the events $\{E_1, E_2, \ldots, E_n\}$ use, e.g., the $2n$ tiered lexicographic IP-Brier score

$$< \underline{\mathbf{b}}_1([r_1, s_1]), \overline{\mathbf{b}}_1([r_1, s_1]), \ldots, \underline{\mathbf{b}}_n([r_n, s_n]),\ \overline{\mathbf{b}}_n([r_n, s_n]) >.$$

Then the following is immediate from *Proposition* 6.

Corollary. The $2n$-tiered, lexicographic IP-Brier score is strictly proper under either the **Γ**-*Maximin* or **E**-*admissibility*-followed-by-**Γ**-*Maximin* decision rules.
As above, the order of the $2n$-terms in the lexicography is irrelevant.

In the light of *Proposition* 5, the theory of IP-coherence$_2$ developed in Section 3 does not produce a strictly-proper IP-scoring rule. In section 3 all IP-forecasts have *real-valued* scores. The strictly proper, *lexicographic* IP-scoring rules identified in *Proposition* 6 do not satisfy that structural assumption of the theory in Section 3. Thus, the analyses of Sections 3 and 4 do not yet provide a unified account of IP-coherence$_2$ and IP-proper scoring rules. In the next Section, we discuss our ideas about this challenge, which was pointed out to us by one of the Readers.

5. Summary

When coherence$_1$ of 2-sided previsions is not enough, and elicitation also matters, then Brier score offers an incentive compatible scoring rule with an equivalent coherence criterion: coherence$_2$ – avoid dominated forecasts. This is de Finetti's analysis, *Proposition* 1.

We extend Brier scoring to IP-coherence$_2$ of interval-valued forecasts, analogous to the familiar use of 1-sided (*lower* and *upper*) previsions for defining IP-coherence$_1$. Subject to an IP-scoring rule for forecasting events, the coherent forecaster gives lower and upper probabilistic forecasts for a particular set of events that characterize elements of an IP-model class M – e.g., the ε-*contamination* class is characterized by IP-forecasts for the atoms of the measure space – *Proposition* 4. Coherence$_2$ of the set of IP-forecasts requires that these lower and upper forecasts are not dominated by any *more determinate IP* model within the model class M, subject to the same *IP* scoring rule.

However, a distinguishing feature between coherence$_1$ and coherence$_2$, namely that Brier score is incentive compatible for elicitation of 2-sided (real-valued) forecasts for events, does not extend to 1-sided forecasts. That is, according to *Proposition* 5, there is no strictly proper, continuous real-valued IP-scoring rule for events. However, by relaxing the conditions on scoring rules to permit a lexicographic utility, subject to either of two IP-decision rules that we investigate, there do exist strictly proper IP-scoring rules for eliciting closed, interval-valued probability forecasts.

There are numerous open questions relating to the preliminary work reported in this paper. We list four topics on which we are currently at work.

1) The central results reported here about IP-coherence$_2$ (Section 3) and IP-incentive compatible scoring (*Proposition* 6 of Section 4) are based on Brier scoring, which is the basis of de Finetti's approach to coherence$_2$. As we showed in [21], each strictly proper scoring rule can serve in place of Brier score to provide a foundation for subjective probability based on forecasting rather than on the prevision game. We conjecture that the positive results reported here about IP-coherence$_2$ can be duplicated using other strictly proper scoring rules in place of Brier score.

2) As noted in Section 2, neither coherence$_1$ nor coherence$_2$ constrains, respectively, a called-off prevision for an event or a called-off forecast for an event, given a null-event. However, lexicographic expected utility [14] is one approach among several others available [7, 16, 28] for improving the treatment of 2-sided conditional probability with called-off previsions given a null-event. (See [2] for a review of some of the open issues.) *Proposition* 6 identifies a class of lexicographic scoring rules that satisfy IP-propriety with respect to interval valued forecasts for events. But those lexicographic scores do not conform to the structural assumption of the theory of IP-coherence$_2$ developed in Section 3, which uses only real-valued scores.

Can we use lexicographic scoring rules to provide a unified theory that provides both an IP-coherence$_2$ (including called-off forecasts given a non-empty event) and an IP-incentive compatible scoring rule?

3) A different challenge to elicitation, even when probability is determinate, is the problem posed by state-dependent utilities. This arises in the choice of the *numeraire* that is to be used, either with outcomes of previsions for coherence$_1$, or in scoring forecasts for coherence$_2$. (See [20] for discussion of the problem in the setting of coherence$_1$.)

Does forecasting afford any advantage over betting in this context and is there a difference also with IP-elicitation?

4) De Finetti's theory of coherence is designed to accommodate all finitely additive probabilities. That is, countable additivity is not a requirement of coherence$_1$ or coherence$_2$. This is achieved by insisting that incoherence, i.e., a failure of simple dominance, is achieved using only finitely many previsions or only finitely many forecasts at one time. In other words, a coherent set of previsions or forecasts may be dominated when more than finitely many are combined at once, even though they cannot be dominated when only finitely many are combined. It is interesting, we find, that even with determinate probabilities, coherence$_1$ and coherence$_2$ are not equivalent in this regard. There are settings where countably many coherent$_2$ forecasts may be combined and remain (simply) undominated by all rival forecasts, though these same previsions may result in a sure-loss when countably many are combined into a single option [25].

In order to accommodate all finitely additive probabilities, when does IP-coherence$_2$ depend upon the restriction that violations of dominance matter only when finitely many forecasts are scored at the same time?

Acknowledgements

Earlier versions of these results were presented at the University of Warwick's *Subjective Bayes Workshop*, the *Purdue Wimer Memorial Lectures* workshop, and CMU's *Games and Decisions* discussion group, and we thank the participants at these meetings for their helpful comments. In particular we appreciate suggestions from Timos Athanasiou, Luca Rigotti, and Kevin Zollman. Also, we appreciate the constructive advice of two, careful, anonymous Readers.

Appendix 1

The Atomic Lower-Upper Probability [*ALUP*] class. This IP-class consists of closed, convex sets of probabilities defined by lower and upper probabilities for atomic events. That is an ALUP model is the largest (closed) convex set of distributions that satisfy such bounds, where the bounds are achieved by the lower and upper probability values given for the atoms of the space. See [10] for discussion about this IP-class of models.

IP-coherence$_2$, where rival forecasts are taken from the ALUP class, arises when the forecaster is called upon to give lower-and-upper forecasts for each atom, ω, and *for the complement to each atom*, ω^c, in the space. That is, in order to duplicate *Proposition* 4 for the ALUP class the forecaster is called upon to give $2n$-many forecasts when $\Omega = \{\omega_1, \ldots, \omega_n\}$. Example 7 illustrates this.

<u>*Example* 7</u> (a continuation of *Example* 4): An illustration of ALUP-coherence$_2$. We provide 3 forecast sets for the

atoms, and their complements in a space defined by $\Omega = \{\omega_1, \omega_2, \omega_3\}$. That is, each forecast set includes IP-forecasts for 6 events. Forecast sets $\mathcal{F}^j$ (j = 2, 3, 4) are given as 6 pairs: $\{p_i, q_i\}$ for ω_i, ω_i^c i = 1, 2, 3. Each of the corresponding 3 score sets is comprised by 3 points, corresponding to the 3 states in Ω. Each point in a score set has 6 coordinates, corresponding to the scores for forecasts of $(\omega_1, \omega_1^c, \omega_2, \omega_2^c, \omega_3, \omega_3^c)$.

	ω_1	ω_1^c	ω_2	ω_2^c	ω_3	ω_3^c
$\mathcal{F}^2$ =	{{.25, .50}	{.50, .75}	{.25, .50}	{.50, .75}	{.25, .50}	{.50, .75}}

$\mathcal{S}^2$ = (.50, .50, .25, .75, .25, .75) for ω_1
(.25, .75, .50, .50, .25, .75) for ω_2
(.25, .75, .25, .75, .50, .50) for ω_2

	ω_1	ω_1^c	ω_2	ω_2^c	ω_3	ω_3^c
$\mathcal{F}^3$ =	{ {.20, .45}	{.55, .80}	{.20, .45}	{.55, .80}	{.20, .45}	{.55, .80} }

$\mathcal{S}^3$ = (.45, .55, .20, .80, .20, .80) for ω_1
(.20, .80, .45, .55, .20, .80) for ω_2
(.20, ,80, .20, .80 .45, .55)} for ω_3

	ω_1	ω_1^c	ω_2	ω_2^c	ω_3	ω_3^c
$\mathcal{F}^4$ =	{ {.10, .35}	{.65, .90}	{.10, .35}	{.65, .90}	{.10, .35}	{.65, .90} }

$\mathcal{S}^4$ = (.35, .65, .10, .90, .10, .90) for ω_1
(.10, .90 .35, .65, .10, .90) for ω_2
(.10, .90, .10, .90, .35, .65)} for ω_3

Forecast sets $\mathcal{F}^2$ and $\mathcal{F}^3$ are ALUP-*coherent*. There do not exist more precise forecast sets from the ALUP-model that dominate either of these sets of forecasts. Their score sets lie in the probability simplex for these 6 events.

Forecast set $\mathcal{F}^4$ is ALUP-incoherent. A de Finetti projection of $\mathcal{S}^4$ produces a more determinate rival ALUP forecast with dominating IP Brier score. In fact, the projection produces a more informative ε-*contamination* model that dominates. The respective IP-Brier scores for $\mathcal{F}^4$ and for $\mathcal{F}^2$ are independent of ω: For $\mathcal{F}^4$ the score is a constant penalty of 0.885. For $\mathcal{F}^2$ it is a constant penalty of 0.750.

Appendix 2

Example 8 – This construction provides a more complicated illustration of *Proposition* 4 where the fixed point $\mathcal{F}^*$ of the process is a limit of the recursive procedure given in the proof of (4.2). Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Forecast sets $\mathcal{F}_j$ are of the form $\{\{p_i, q_i\}$: for events ω_i: i = 1, 2, 3}.

$\mathcal{F} = \mathcal{F}_0$ = { {.25, .60}, {.20, .50}, {.10, .40} }

$\mathcal{S} = \mathcal{S}_0$ = {(.60, .20, .10), (.25, .50, .10), (.25, .20, .40)}

(Step 1) Project score set $\mathcal{S}_0$ to form set

$\boldsymbol{T}_1$ = { (.6$\bar{3}$, .2$\bar{3}$,, .1$\bar{3}$,), (.30, .55, .15), (.30, .25, .45)}

Form the new forecast and score sets $\mathcal{F}_1$, $\mathcal{S}_1$ based on the probabilities in set $\boldsymbol{T}_1$

$\mathcal{F}_1$ = { {.30, .6$\bar{3}$ } {.2$\bar{3}$, .55} {.1$\bar{3}$, .45} }

$\mathcal{S}_1$ = {(.6$\bar{3}$ .2$\bar{3}$ .1$\bar{3}$) (.30, .55, .1$\bar{3}$) (.30, .2$\bar{3}$, .45)}

(Step 2) Project set $\mathcal{S}_1$ to form set

$\boldsymbol{T}_2$ = { (.6$\bar{3}$, .2$\bar{3}$, .1$\bar{3}$) (.30$\bar{5}$, .5$\bar{5}$, .1$\bar{5}$) (.30$\bar{5}$, .2$\bar{5}$, .4$\bar{5}$)}

Form the new forecast and score sets $\mathcal{F}_2$, $\mathcal{S}_2$ based on the probabilities in set $\boldsymbol{T}_2$

$\mathcal{F}_2$ = { {.30$\bar{5}$, .63$\bar{3}$ } {.23$\bar{3}$, .55$\bar{5}$ } {.13$\bar{3}$, .45$\bar{5}$ } }

$\mathcal{S}_2 = \{(.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.30\overline{5}, .5\overline{5}, .1\overline{3})\ (.30\overline{5}, .2\overline{3}, .4\overline{5})\}$

(Step 3) Project $\mathcal{S}_2$ to form set

$\boldsymbol{T}_3 = \{\ (.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.30\overline{740}, .55\overline{740}, .13\overline{740})\ (.30\overline{740}, .23\overline{740}, .45\overline{740})\}$

Form the new forecast and score sets $\mathcal{F}_3$, $\mathcal{S}_3$ based on the probabilities in set $\boldsymbol{T}_3$

$\mathcal{F}_3 = \{\ \{.30\overline{740}, .6\overline{3}\}\ \{.2\overline{3}, .55\overline{740}\}\ \{.1\overline{3}, .45\overline{740}\}\ \}$

$\mathcal{S}_3 = \{(.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.30\overline{740}, .55\overline{740}, .1\overline{3})$

$(.30\overline{740}, .2\overline{3}, .45\overline{740})\}$

(Step 4) Project $\mathcal{S}_4$ to form set

$\boldsymbol{T}_4 \approx \{\ (.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.308, .558, .134)\ (.308, .234, .458)\}$

Form the new forecast and score sets $\mathcal{F}_4$, $\mathcal{S}_4$ based on the probabilities in set $\boldsymbol{T}_4$

$\mathcal{F}_4 = \{\ \{.308, .6\overline{3}\}\ \{.2\overline{3}, .558\}\ \{.1\overline{3}, .458\}\ \}$

$\mathcal{S}_4 = \{(.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.308, .558, .1\overline{3})\ (.308, .2\overline{3}, .458)\}$

Iterate the process which converges to forecast set

$\mathcal{F}^* = \{\ \{.308\overline{6}, .6\overline{3}\}\ \{.2\overline{3}, .558\}\ \{.1\overline{3}, .458\}\ \}$

and score set

$\mathcal{S}^* = \{(.6\overline{3}, .2\overline{3}, .1\overline{3})\ (.308\overline{6}, .558, .1\overline{3}\ (.308\overline{6}, .2\overline{3}, .458)\}$

$\mathcal{F}^*$ is an ε-contamination model whose IP-Brier score dominates $\mathcal{F}$'s score. $\mathcal{F}^*$ has greater *informativeness* (greater *determinacy*) than forecast $\mathcal{F}$ as the hull $\boldsymbol{H}(\mathcal{S}^*)$ is isomorphic to a proper subset of the hull $\boldsymbol{H}(\mathcal{S})$.

References

[1] Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis* (2nd edition). Springer-Verlag. New York.

[2] Cozman, F. and T.Seidenfeld (2009) Independence for Full Conditional Measures, Graphoids and Bayesian Networks. In *Foundations of the Formal Sciences VI*. B.Lowe, E.Pacuit, and J-W Romeijn (eds.). College Publications: London.

[3] de Finetti, B. (1974) *Theory of Probability* (vol. 1). John Wiley: New York.

[4] de Finetti, B. (1981) The role of *dutch books* and *proper scoring rules*. *Brit. J. Phil. Sci.* 32: 55-56.

[5] de Finetti, B (2008) *Philosophical Lectures on Probability* (A.Mura, Ed.) Springer: United States.

[6] Dubins, L.E. (1975) *Finitely additive conditional probabilities, conglomerability and disintegrations*. Annals of Probability **3**: 89-99.

[7] Gilio, A. (1996) Algorithms for Conditional Probability Assessments. In *Bayesian Analysis in Statistics and Econometrics*. D.A.Berry, K.MChaloner, and J.K.Geweke (eds.) John Wiley: New York, pp. 29-39.

[8] Gilio, A and Sanfilippo, G. (2011) Coherent conditional probabilities and proper scoring rules. In *ISIPTA-11, Proceedings of the Seventh International Symposium on Imprecise Probability*. (www.sipta.org)

[9] Harsanyi, J. and Selten, R. (1988) *A General Theory of Equilibrium Selection in Games*. MIT Press. Cambridge.

[10] Herron, T., T.Seidenfeld, and L.Wasserman (1997) Divisive Conditioning: Further Results on Dilation. *Phil. Sci.*, 64: 411-444.

[11] Huber, P.J. (1973) The use of Choquet capacities in statistics. Bull. Int. Stat. Inst. 45: 181-191.

[12] Huber, P.J. (1981) *Robust Statistics*. Wiley: New York.

[13] Levi, I. (1974) On Indeterminate Probabilities. *J.Phil* 71: 391-418.

[14] Levi, I. (1980) *The Enterprise of Knowledge*. MIT Press: Cambridge.

[15] Predd, J., R.Seiringer, E.H.Lieb, D.Osherson, V.Poor, and S.Kulkarni (2009) Probabilistic coherence and proper scoring rules. *IEEE Trans. Information Theory* 55; 4786-4792.

[16] Regazzini, E. (1987) De Finetti's Coherence and Statistical Inference. *Ann. Stat.* **15**: 845–864.

[17] Savage, L.J. (1954) *The Foundations of Statistics*. Wiley: New York.

[18] Savage, L.J. (1971) Elicitation of personal probabilities and expectations. *J. Amer. Stat. Assoc.* 66: 783-801.

[19] Schervish, M.J. (1989) A general method for comparing probability assessors. *Ann. Stat.* 17: 1856-1879.

[20] Schervish, M.J., T.Seidenfeld and J.B.Kadane (1990) State-dependent Utilities. *J.A.S.A.* **85**: 840-847.

[21] Schervish, M.J, T.Seidenfeld, and J.B.Kadane (2009) Proper Scoring Rules, Dominated Forecasts, and Coherence. *Decision Analysis* **6**, #4: 202-221.

[22] Seidenfeld, T. (1996) Some Static and Dynamic Aspects of Robust Bayesian Theory. Proceedings of the 1996 Workshop, *Applications and Theory of Random Sets*. IMA, Univ. Minnesota.

[23] Seidenfeld, T., M.J.Schervish, and J.B.Kadane (1990) Decisions without Ordering. In *Acting and Reflecting*. W.Sieg (ed.) Kluwer Publishing: Dordrecht, pp. 143-170.
[24] Seidenfeld, T., M.J.Schervish, and J.B.Kadane (2010) Coherent choice functions under uncertainty. *Synthese* 172, #1: 157-176. Presented at ISIPTA-07.
[25] Seidenfeld, T., M.J.Schervish, and J.B.Kadane (05/2011) Dominating Countably Many Forecasts. Technical Report, Statistics Department, CMU.
[26] Smith, C.A.B. (1961) Consistency in Statistical Inference and Decision. *J.R.S.S*. B **23**: 1-25.
[27] Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London.
[28] Williams, P.M. (1975) Notes on conditional previsions*. I. J. Approximate Reasoning* **4**: 366–383.