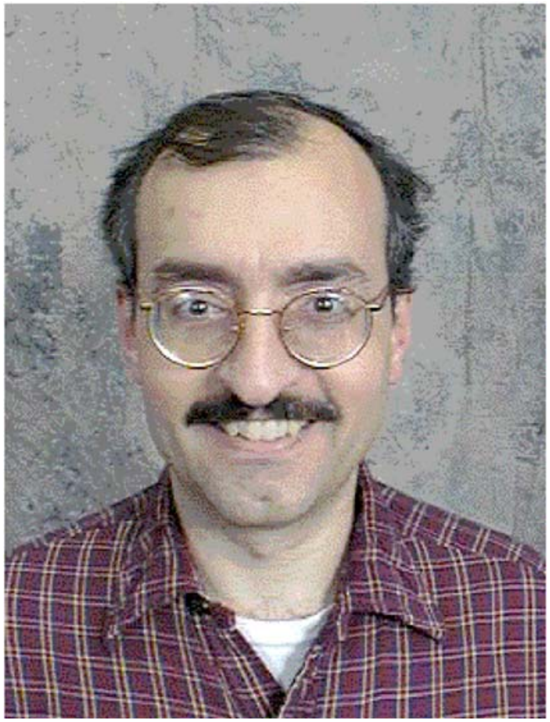
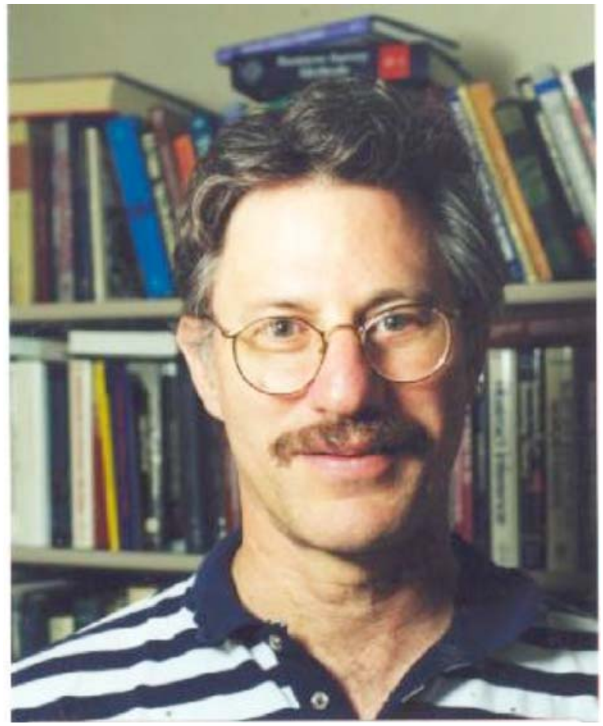
Coherence with Proper Scoring Rules

Mark Schervish, Teddy Seidenfeld, and Joseph (Jay) Kadane



Mark Schervish



Joseph ("Jay") Kadane

In this paper we explore three concepts of dominance – (de Finetti) *coherence* :

- **Coherence$_1$ for previsions of random variables with generalized betting;**
- **Coherence$_2$ for probability forecasts of events with Brier score penalty;**
- **Coherence$_3$ probability forecasts of events with various proper scoring rules.**

In order to show that these variations on *coherence* are equivalent we use Minimiax theory to relate *dominance* and *Bayes's decisions*, generally.

- **With each proper scoring rule, assuming that (simple) mixed strategies are available, all and only non-Bayes forecasts are dominated options.**

This extends de Finetti's result that Coherence$_1$ and Coherence$_2$ are equivalent, to include Coherence$_3$.

- **However, unless the proper scoring rules are continuous, Coherence$_3$ is not equivalent to the other two senses without the use of *incoherent* mixed strategy forecasts to dominate. That is, with discontinuous proper scores, no coherent forecast may dominate particular incoherent ones.**

1. Structural assumptions about preference used in this presentation.

***1.1 Act-state independence*: no cases of "moral hazards" are considered – so strict dominance is valid.**

***Reminder:* Consider the following binary state, two act problem, with outcomes ordinally (or cardinally) ranked so that more is better.**

	ω_1	ω_2
Act_1	**3**	**1**
Act_2	**4**	**2**

Act_2 strictly dominates Act_1. Nonetheless, if

$$Prob(\omega_i \mid Act_i) \approx 1 \quad (i = 1, 2),$$

then dominance carries no force.

A rational decision maker then prefers Act_1 to Act_2.

- **We consider decision problems <u>without</u> *moral hazards*.**

1.2 **The focus here is on *normal* form (aka "strategic form") games/decisions.**

The decision maker can commit, in advance, to contingency planning.

- **BUT conditional previsions/forecasts are represented using the device of a *called-off options*.**

***Example*: A bookie's called-off prevision on event *A* given event *B* – the fair betting odds – yields payoff to the bookie**

$$\mathrm{B}\, S_A[\ \mathrm{A} - P_B(A)\ ]$$

Where: **A is the indicator function for event *A*.**

B is the indicator function for event *B*.

S_A is the total stake (positive or negative) for the wager, fixed by a rival gambler.

$P_B(A)$ is the bookie's conditional prevision or called-off fair-betting odds on *A*, given *B*.

- **So, we bypass the difficult work that is needed when the conditioning event *B* is null: Coherence for called-off wagers impose no substantive constraint on the conditional prevision when the conditioning event is null.**

- **We make no assumption that *normal* and *extensive* form decisions are equivalent, and generally they will not be equivalent**

  Aside: The non-equivalence is salient in decision theories with Indeterminate/Imprecise Probabilities – IP Theory.

- **This issue is important for so-called "Dynamic Book" arguments**

2. de Finetti's *Coherence$_1$ of previsions*.

An important, and historically early application of strict dominance in decision making is de Finetti's criterion of *coherence$_1$ of previsions*.

Coherence$_1$ **– de Finetti's notion of coherence$_1$ begins with**

an arbitrary partition of states, $\Omega = \{\omega_i: i \in I\}$, and

a class of bounded real-valued variables, $\chi = \{X_j: j \in J\}$, defined on Ω.

For each random variable $X \in \chi$, the rational agent has a (two-sided) *prevision* P(X) which is to be interpreted as a *fair* price (both for *buying* and *selling*) gambles.

For all real $\beta > 0$, small enough so that the agent is willing to pay the possible losses, the agent is willing

to pay βP(X) in order *to buy* (i.e., to receive) βX in return.

and, is willing

to accept βP(X) in order *to sell* (i.e., to pay) βX in return.

The agent will accept the gamble

$$\beta[X - P(X)]$$

as a change in fortune, for all sufficiently small (positive or negative) β.

The agent is required to accept all <u>finite</u> sums of gambles of the preceding form.

That is, for all finite n and all small, real $\beta_1, \ldots, \beta_n$ and all $X_1, \ldots, X_n \in \chi$, the agent will accept the combination of gambles

$$\sum_{i=1}^{n} \beta_i[X_i - P(X_i)].$$

Where β_i is positive, the agent buys β_i-units of X_i for a price of $\beta_i P(X_i)$

where it is negative, the agent sells β_i-units of X_i for a price of $\beta_i P(X_i)$.

- **The previsions are *incoherent₁* if there is a *uniformly negative* acceptable finite combination of gambles.**

That is, if there exists a sum of the form above and $\varepsilon > 0$ such that, for each $\omega \in \Omega$,

$$\sum_{i=1}^{n} \beta_i[X_i(\omega) - P(X_i)] < -\varepsilon.$$

Otherwise the agent's previsions are *coherent$_1$*.

Where previsions are incoherent$_1$, the book that indicates this constitutes a combination of gambles that is uniformly, strictly dominated by *not-betting* (= 0).

de Finetti's *Coherence$_1$ Theorem*:

- **A set of previsions are coherent$_1$ if and only if they are the expected values for the respective random variables under a (finitely additive) probability distribution over Ω.**

- **When the variables are indicator functions for events (subsets of Ω), coherent previsions are exactly those in agreement with a (finitely additive) probability. And then the $|\beta_i|$ are the stakes in winner-take-all bets, where the previsions fix betting rates, P(X_i) : 1-P(X_i).**

De Finetti's result applies to conditional previsions, given an event *B*. These use called-off gambles of the form

$$B\,\beta[X - P(X)]$$

where *B* is the indicator function for the conditioning event *B*.

Then, with the proviso about non-null conditioning events, coherence$_1$ assures that coherent called off (2-sided) previsions are finitely additive conditional expectations, given the conditioning event.

When the random variables, X_i, include indicator functions for events, the resulting coherent$_1$ previsions include conditional probabilities for these events.

Note well: Called-off previsions correspond only to *normal form*, and not to *extensive form* decisions. There is no *dynamical coherence* in de Finetti's theory. His theory covers merely *static* aspects of coherence$_1$.

- **Thus, de Finetti's theory of coherence$_1$ does not require updating/learning by Bayesian conditional probabilities.**

3. de Finetti's *Coherence$_2$ of forecasts with Brier-score*.

In a 1981 note to the *BJPS* de Finetti explains some of his motives for introducing (in 1974) a second, but equivalent concept of coherence.

- **The betting context for coherence$_1$ also serves as an elicitation of the bookie's previsions. But the bookie's *announced* previsions may reflect strategic aspects of the game-like interaction with the rival gambler. The gambler plays second, after the bookie has announced her/his fair previsions. The bookie may anticipate this and play strategically.**
- **The bookie's previsions may change upon learning which bets the gambler has chosen.**

Using the "ruler" changes the length of the object measured.

Brier score – squared error loss.

The loss (negative utility) that attaches to a prevision P(*X*) is

$$-(X - P(X))^2.$$

The called-off loss for a conditional prevision P(X) given event *B* is

$$-B(X - P(X))^2.$$

The combined losses from a finite set of (called-off) previsions is the sum of the individual losses just as payoffs from bets are added together with coherence$_1$.

- **Coherence$_2$: A set of previsions is incoherent$_2$ if for some finite subset of those previsions, there exist a rival set of forecasts that dominace, by resulting in a (uniformly) smaller Brier score loss over all states.**

  The previsions are coherent$_2$, otherwise.

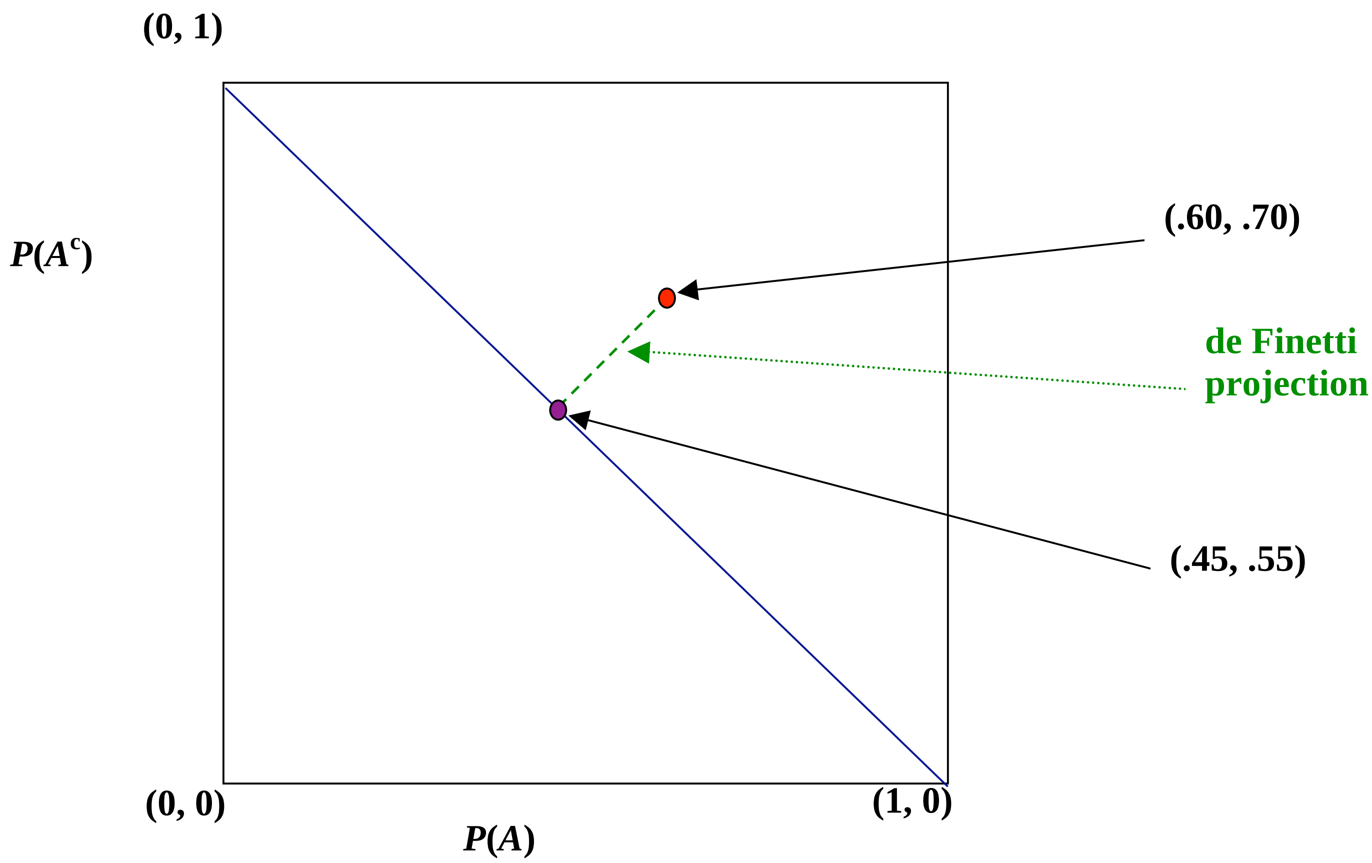
If a forecaster makes decisions by maximizing subjective expected utility, and since the mean of a distribution (uniquely) minimizes mean-squared error, then her/his announced previsions will be her/his "honest" previsions – her/his degrees of belief when forecasting events – including announcing conditional previsions for called-off forecasts.

- **Thus, Brier score is (strictly) proper – it elicits the forecaster's conditional expected value of the variable X, given event B.**

De Finetti established equivalence between coherence$_1$ and coherence$_2$ with an elegant, geometric argument.

- **The distance from an incoherent$_1$ forecast Q to a corner of the probability simplex is greater than the corresponding distance to that corner from the coherent$_1$ forecast that is the projection of Q onto the probability simplex.**

- **If P is coherent$_1$ forecast, hence a point in the probability simplex, there is no other point that has smaller distance to each corner of the simplex.**


(0, 1)
$P(A^c)$
(.60, .70)
de Finetti projection
(.45, .55)
(0, 0)
(1, 0)
$P(A)$

4. *Coherence$_3$ of forecasts with proper scores*.

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

A scoring rule for forecasting an event *A* given event *B*, the called-off score for the forecast P(*A* | *B*), is defined by two extended real-valued loss functions (g_0, g_1), with arguments from [0,1], as follows.

if *A* occurs, the loss is Bg_1(P(A|B))

and if A^c occurs, the loss is Bg_0(P(A|B)).

- **As before, with a finite set of called-off forecasts, even using different scoring rules for different forecasts, the combined score is the sum of the individual scores.**

- **A scoring rule (as a loss function) for forecasting A, given B, is (*strictly*) *proper* provided that a decision maker who maximizes subjective expected utility (uniquely) minimizes expected score by using her/his conditional probability P(A|B) as the called-off forecast.**

Thus (strictly) proper scores all elicit the same forecast as does Brier score. Savage (1971) and Schervish (1989) characterized the class of (strictly) proper scoring rules. For the results on coherence and dominance highlighted here, we do not need to review these characterizations of proper scoring rules.

Surprisingly, neither de Finetti nor Savage gave us more than a brief sketch of their thinking whether (strictly) proper scoring rules all support the same concept of coherence.

- **Do the same forecasts avoid dominance under all proper scoring rules?**

In what follows, we allow that the (strictly) proper scoring rule may vary with the event forecasted. $\{(g_{0,Ai,Bi}, g_{1,Ai,Bi}): i = 1, \ldots\}$

- ***Coherence$_3$*: A collection of conditional forecasts for events is incoherent$_3$ if some finite sub-collection of those forecasts are uniformly dominated by some rival set of conditional forecasts.**

  The forecasts are coherent$_3$, otherwise.

<u>Central Result</u>: Subject to 3 mild constraints on these proper scoring rules (Assumptions 1-3, see below) and assuming that the forecaster can use simple mixed strategy forecasts then,

(i) Coherence$_1$ and Coherence$_3$ are equivalent.

(ii) Moreover, if the scoring rules are continuous, each incoherent forecast is dominated by some (non-mixed) coherent forecast.

***Assumptions*:**

(1) For k = 0, 1 $g_k(x)$ is bounded below with $g_0(0) = g_1(1) = 0$.

(2) For k = 0, 1 $g_k(x)$ is continuous at $x = k$.

(3) For k = 0, 1 $g_k(x)$ is finite for $0 < x < 1$.

Aside: We give examples to support each of these assumptions.

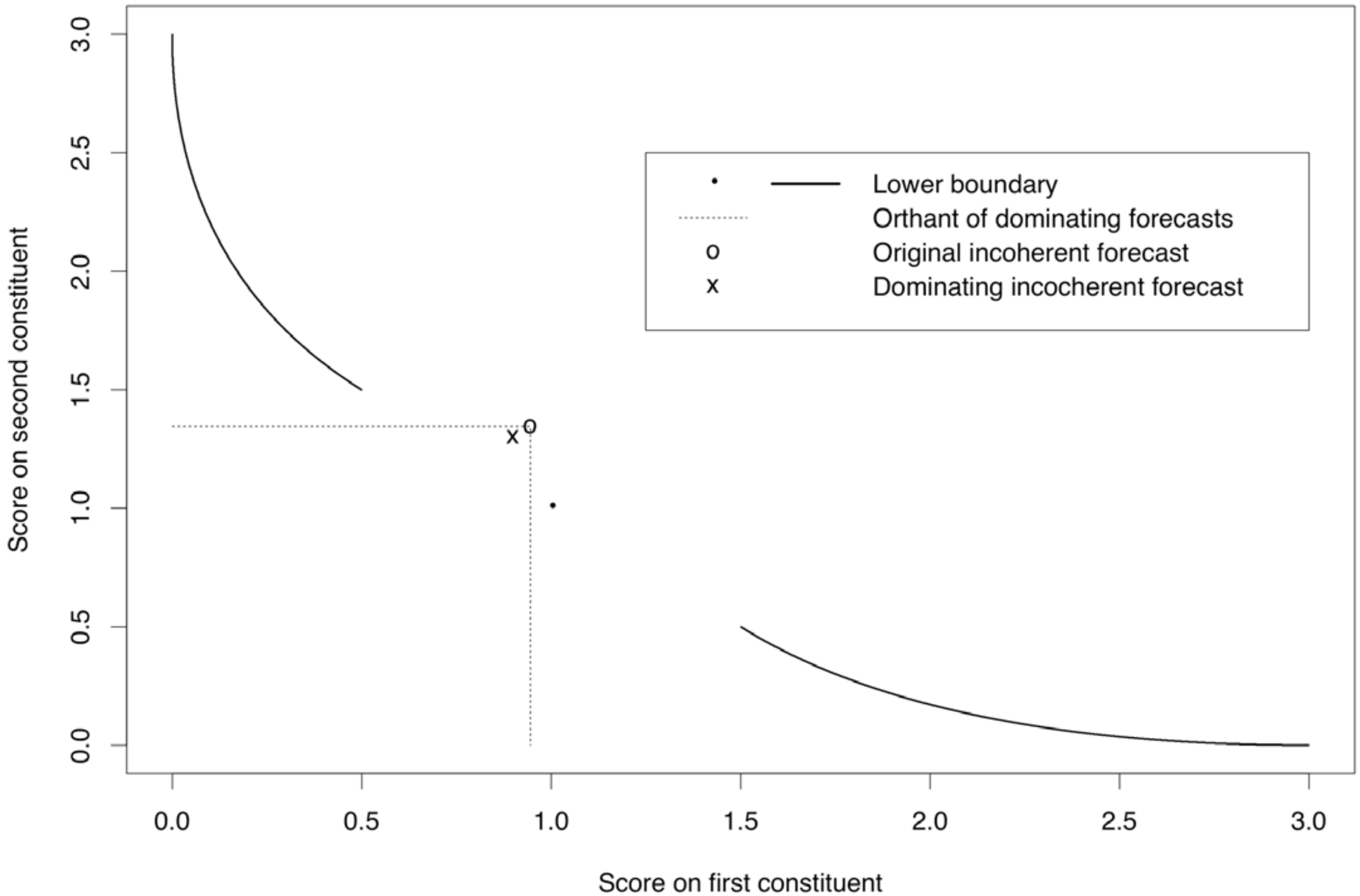
Here is an illustration (Example 3, p. 5) where in order to dominate an incoherent forecast, one must use a different incoherent forecast. This example uses a (strictly) proper scoring rule that is not continuous.

Consider a binary partition {*A*, *A*^c^} and a common scoring rule for forecasting both events.

$$g_0(x) = \begin{cases} x^2 & \text{if } x \leq \frac{1}{2} \\ x^2 + \frac{1}{2} & \text{if } x > \frac{1}{2} \end{cases}$$

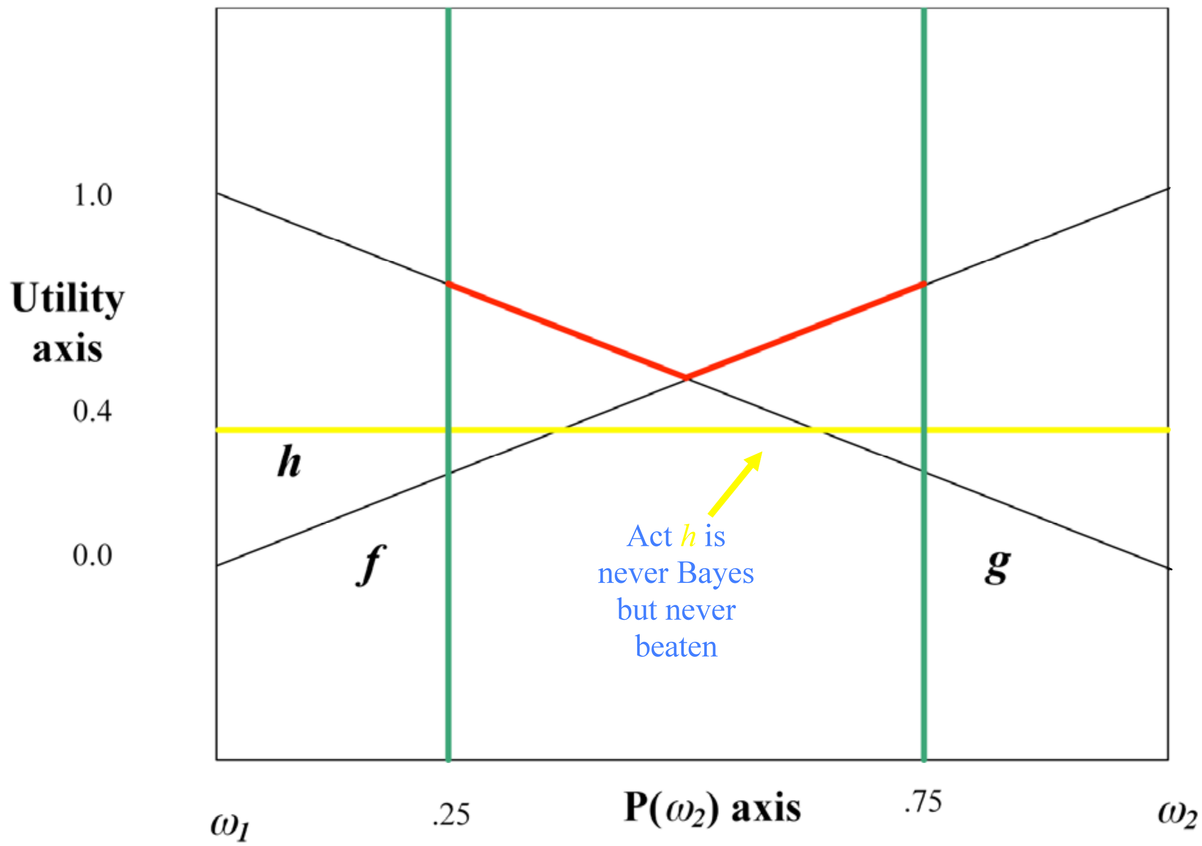
$$g_1(x) = \begin{cases} (1-x)^2 + \frac{1}{2} & \text{if } x \leq \frac{1}{2} \\ (1-x)^2 & \text{if } x > \frac{1}{2} \end{cases}$$

The incoherent forecast pair (0.6, 0.7) with scores (1.15, 0.95) is not dominated by any coherent forecast, but the incoherent forecast pair (0.55, 0.65) dominates, with scores (1.125, 0.925).


Lower boundary
Orthant of dominating forecasts
o Original incoherent forecast
x Dominating incocherent forecast
Score on second constituent
Score on first constituent
0.0
0.5
1.0
1.5
2.0
2.5
3.0

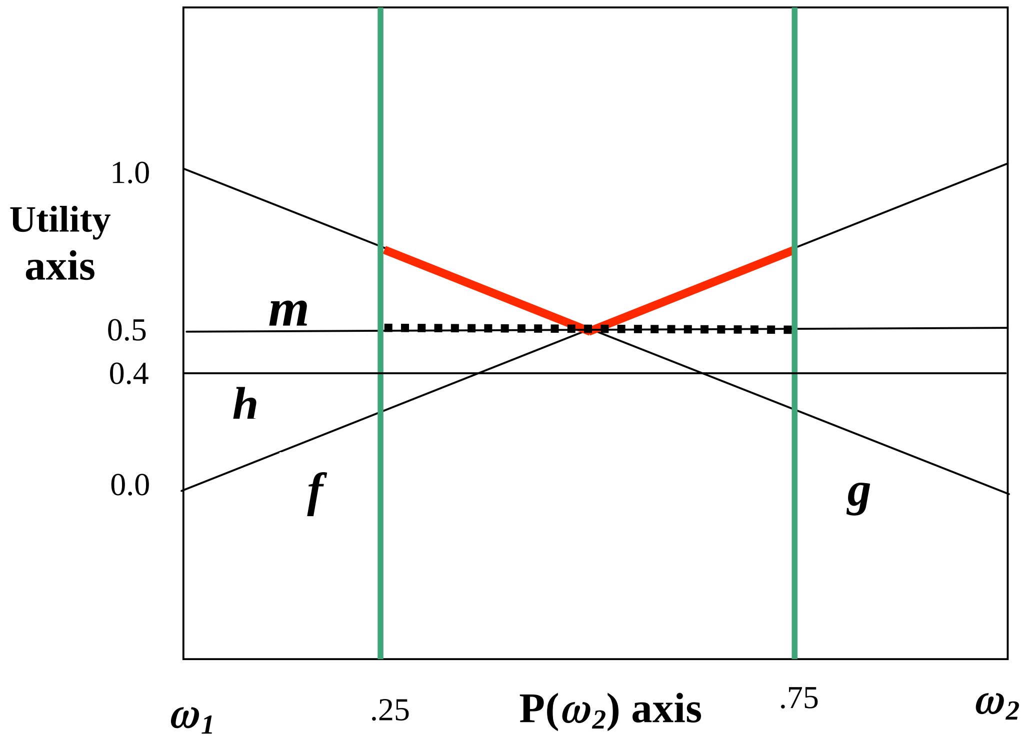
5. When non-Bayes decisions may be dominated.

Consider this decision problem on 2-states with three options.



Now, consider the mixed strategy option, *m* which is .5*f* ⊕ .5*g*.

Observe that *m* dominates *h*.



Our approach for showing the equivalence among the different senses of coherence is to generalize this picture to apply it to forecasting under proper scoring rules. Ours is NOT a generalization of de Finetti's geometric argument, which is the focus of a paper by Predd et al. (2007).

Instead we have generalized D.Pearce's (1984) game theoretic result about *rationalizable* strategies to cover certain semi-infinite 2-person 0-sum games. These are games where the Statistician has infinitely many pure strategy forecasts, and Nature has only finitely many pure strategies, corresponding to choosing from a finite partition of states.

- **When the scoring rules are continuous, the game's solution set is closed (below) and pure strategies win for the Statistician.**

- **When the proper scoring rules are discontinuous, the game has a value but the solution set may be open (below), as in the Example. Then there is no Minimax solution for the Statistician. Only other non-Bayes solutions may dominate a particular non-Bayes forecast, as they form the epsilon-minimax strategies for the corresponding game.**

6. Summary and review.

1. ***Structural assumptions for this presentation.***
   Moral hazards and dominance
   Normal versus extensive form decision making

2. **de Finetti's *coherence$_1$ of previsions*.**

3. **de Finetti's *coherence$_2$ of forecasts with Brier score*.**

4. ***Central result about the equivalence of all three coherence concepts***
   An example of forecasts with a discontinuous (strictly) proper scores where no coherent forecast dominates some incoherent forecasts.

5. ***On dominance and rationalizability* (D. Pearce's 1984 result) *extended to semi-infinite 0-sum games against Nature.***

Example: Consider a countably infinite state space $\Omega = \{\omega_1, \omega_2, ...\}$ with its powerset serving as the σ-field of sets. Let the class of forecast variables be $\mathcal{X} = \{W_i: i = 1, ...\}$, where W_i is the indicator for event $\{\omega_i\}$. Consider a (purely) finitely additive probability with 0 probability for each atom: $P(\omega_i) = 0$, $i = 1, ...$. These form a *coherent*$_1$ set of previsions. But to secure *coherence*$_1$ only finitely many previsions may be used in betting. On the contrary, as de Finetti noted, were countably many previsions used simultaneously, then for each $\omega \in \Omega$, choosing $\alpha_i = -1$, yields the constant payoff:

$$\Sigma_i -(W_i(\omega) - P[W_i]) = \Sigma_i -W_i(\omega) = -1,$$

which is a sure loss of -1 to the *bookie*.

However, the infinite set of forecasts $\{P[Wi] = 0\}$ has a constant Brier score
$-1 = \Sigma_i -(W_i(\omega) - P[W_i])^2$,
which is not dominated (let alone uniformly dominated) by any set of forecasts, $\{ P'[W_i] \}$, regardless whether or not these rival forecasts are *coherent*$_{1\text{-or-}2}$.

This example generalizes to the following result.

Proposition about *coherence*$_2$ of f.a. probability forecasts:

(*i*) Call a set of variables $\{X_i$: $i = 1, \ldots, \}$ *uniformly series-bounded,*

if there exists a real number *b*, so that for each $\omega \in \Omega$, $\Sigma_i |X_i(\omega)| < b$

When the class of forecast variables are *uniformly series-bounded*, it is *not* necessary to restrict Brier score to sums of *finitely* many unconditional forecasts in order that a set of finitely additive probability forecasts is undominated by a rival forecast set: *coherence*$_2$.

Note: *Series-boundedness* is satisfied when forecasting events that form a partition.

(*ii*) However, the previous result is modest. It does not generalize to include *called-off* forecasts. Then, because of non-conglomerablity of merely finitely additive conditional probabilities, their coherence$_2$ depends upon the clause that only finitely many Brier scores from called-off forecasts are summed up simultaneously.