DECISION ANALYSIS
Vol. 6, No. 4, December 2009, pp. 202–221
ISSN 1545-8490 | EISSN 1545-8504 | 09 | 0604 | 0202

informs®

DOI 10.1287/deca.1090.0153

Proper Scoring Rules, Dominated Forecasts, and Coherence

Mark J. Schervish, Teddy Seidenfeld, Joseph B. Kadane
Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
{mark@stat.cmu.edu, teddy@stat.cmu.edu, kadane@stat.cmu.edu}

The concept of coherent probabilities and conditional probabilities through a gambling argument and through a parallel argument based on a quadratic scoring rule was introduced by de Finetti (de Finetti, B. 1974. *The Theory of Probability*. John Wiley & Sons, New York). He showed that the two arguments lead to the same concept of coherence. When dealing with events only, there is a rich class of scoring rules that might be used in place of the quadratic scoring rule. We give conditions under which a general strictly proper scoring rule can replace the quadratic scoring rule while preserving the equivalence of de Finetti's two arguments. In proving our results, we present a strengthening of the usual minimax theorem. We also present generalizations of de Finetti's fundamental theorem of probability to deal with conditional probabilities.

Key words: Brier score; finite additivity
History: Received on November 14, 2008. Accepted on June 22, 2009, after 1 revision. Published online in *Articles in Advance* October 13, 2009.

1. Introduction

1.1. Background

There are two different elicitation methods that lead to subjective probability. The first is through fair betting odds. What is a price at which you would buy or sell a ticket that would pay \$1 if it rains tomorrow in your city, and nothing if it does not? If you say \$0.30, then we take your probability of rain to be 30%. Put in mathematical terms, if A is the event in question, and I_A is the indicator random variable for the event A (1 if A occurs and 0 if A^c occurs), then $c[I_A - P(A)]$ is the payoff for buying c tickets if $c > 0$, or selling c tickets if $c < 0$. Here, $P(A)$ is that number (0.3 in the example above) that leaves you, as the Decision maker, indifferent as to whether c is positive or negative. Then $P(A)$ is taken to be your probability of A.

The second elicitation method uses the Brier (1950) score (squared-error loss) to find a value for $P(A)$. Suppose that a forecast p for the event A is penalized by the loss $(I_A - p)^2$. Then the expected loss is

$$P(A)(1-p)^2 + (1-P(A))p^2.$$

Simple differentiation shows that the unique optimal choice of p is $p = P(A)$.

Both of these methods can be generalized to deal with conditional probabilities. The first would seek a number $P(A \mid B)$ such that you are indifferent to the sign of c if faced with a gamble paying $cI_B[I_A - P(A \mid B)]$. The second would minimize the loss $I_B[I_A - p]^2$. Note that when $B = \Omega$ (the sure event), the conditional case specializes to the unconditional one. Because our results apply to the conditional case, we use it throughout. Conditional probabilities are a useful, if not vital, tool for eliciting joint distributions and tails of distributions. Kadane et al. (1980) showed how conditional probabilities can help to elicit prior distributions in regression problems.

Both of these elicitation methods were proposed by de Finetti (1974), and each of them has its own disadvantages. de Finetti (1981, 1974, p. 93) was uncomfortable with the first method because the decision maker might try to guess whether an "opponent" would choose $c > 0$ or $c < 0$, and then might modify the elicited $P(A)$ accordingly. This introduces an unwelcome strategic aspect to the first method.

The second method is peculiar because it appears to rely on the Brier score so directly. Would some other loss function destroy the equivalence between the two methods? For example, §3 of Schervish (1989)

discusses how a decision maker can construct a loss function that reflects what is important to her/him in the elicitation of each probability. It would be comforting if all such loss functions inherited the important properties of squared-error loss while allowing more flexibility to choose a relevant loss. We give some examples below (Examples 3 and 4) to illustrate why and how a decision maker might choose alternative loss functions. The results given in this paper, growing out of several previous research efforts, aim to give conditions on loss functions under which the two elicitation methods coincide. Essentially, the methods coincide if the loss function corresponds to a *strictly proper scoring rule* (a generalization of the Brier score as defined in Definition 3) that satisfies a few mild assumptions, which we state in §2. We give several examples in §4 to show what can go wrong if the assumptions do not hold.

1.2. Notation and Definitions

The uses of the Brier score and betting to elicit a probability extend naturally to the elicitation of an arbitrary number of probabilities and/or conditional probabilities. Conditional probability can be thought of as a function of two events, i.e., $P(A \mid B)$. When $B = \Omega$, we can refer to $P(A \mid \Omega) = P(A)$ as a *marginal* probability. For the remainder of this paper, when we refer to probabilities unqualified, we mean both conditional and marginal probabilities.

To be able to deal with arbitrary collections of marginal and conditional probabilities simultaneously, we introduce some notation. Let $\aleph$ be an index set, and let $\mathscr{X} = \{(A_\alpha, B_\alpha): \alpha \in \aleph\}$ be a set of pairs of events. For each $\alpha \in \aleph$, A_α is a subset of Ω, and B_α is a nonempty subset of Ω. For each $(A, B) \in \mathscr{X}$, the decision maker is required to provide a real-valued (conditional) probability $P(A \mid B)$.

DEFINITION 1 (COHERENCE$_1$). A set of elicited probabilities is *incoherent*$_1$ if there exists a finite subset $\{\alpha_1, \ldots, \alpha_k\}$ of $\aleph$ and corresponding values $\{c_{\alpha_1}, \ldots, c_{\alpha_k}\}$ so that the net payoff to the decision maker is uniformly negative in all states $\omega \in \Omega$. That is, there exists $\epsilon > 0$ such that, for all $\omega \in \Omega$,

$$\sum_{i=1}^{k} c_{\alpha_i} I_{B_{\alpha_i}}[I_{A_{\alpha_i}}(\omega) - P(A_{\alpha_i} \mid B_{\alpha_i})] < -\epsilon. \quad (1)$$

The elicited probabilities are called *coherent*$_1$ otherwise. If (1) occurs, we say that *book has been made* against the decision maker.

Thus, coherence$_1$ is the requirement that the decision maker's elicited probabilities cannot be (uniformly) dominated by the status quo, corresponding to the state of neither buying nor selling such gambles. Coherence$_1$ is a rationality condition on probabilities elicited by the first method of §1.1. Throughout this paper, we will refer to elicited probabilities as *forecasts* for the following reason. When elicited probabilities are incoherent$_1$, they do not satisfy the axioms of probability theory. Hence, it seems awkward to call them probabilities.

EXAMPLE 1. Let C be an event, and suppose that an agent provides the following forecasts: $P(C \mid \Omega) = 0.3$ and $P(C^c \mid \Omega) = 0.5$. In the notation of Definition 1, we have $k = 2$, $A_1 = C$, $B_1 = \Omega$, $A_2 = C^c$, and $B_2 = \Omega$. Choosing $c_1 = c_2 = -1$ in (1) gives us

$$\sum_{i=1}^{2} c_i I_{B_i}[I_{A_i}(\omega) - P(A_i \mid B_i)] = -I_C(\omega) + 0.3 - I_{C^c}(\omega) + 0.5 = -0.2$$

for all ω. Hence, the forecasts are incoherent$_1$.

There is an analogous rationality condition on forecasts that are elicited by the second method of §1.1. Let $L = I_B(I_A - p)^2$ stand for the loss (or score) suffered by an agent who forecasts p as the conditional probability of A given B. The loss given to a finite set of forecasts $\{P(A_{\alpha_1} \mid B_{\alpha_1}), \ldots, P(A_{\alpha_k} \mid B_{\alpha_k})\}$ is the sum of the losses for the individual forecasts.

DEFINITION 2 (COHERENCE$_2$). A set of forecasts is *incoherent*$_2$ if for some finite subset of those forecasts there exists an alternative set of forecasts that result in a (uniformly) smaller loss in all the states $\omega \in \Omega$. The set of forecasts is called *coherent*$_2$ otherwise.

Thus, coherence$_2$ is the requirement that no finite subset of the decision maker's forecasts can be (uniformly) dominated by a rival set of forecasts in terms of squared-error loss.

EXAMPLE 2 (CONTINUATION OF EXAMPLE 1). The sum of the squared-error losses for the two stated forecasts in Example 1 is

$$[I_C(\omega) - 0.3]^2 + [I_{C^c}(\omega) - 0.5]^2 = \begin{cases} 0.74 & \text{if } \omega \in C, \\ 0.34 & \text{if } \omega \in C^c. \end{cases} \quad (2)$$

The following coherent$_1$ forecasts result in uniformly smaller loss: $P(C \mid \Omega) = 0.4$ and $P(C^c \mid \Omega) = 0.6$. The sum of the losses for these forecasts is

$$[I_C(\omega) - 0.4]^2 + [I_{C^c}(\omega) - 0.6]^2 = \begin{cases} 0.72 & \text{if } \omega \in C, \\ 0.32 & \text{if } \omega \in C^c, \end{cases}$$

which is uniformly smaller than (2). Hence, the original forecasts are incoherent$_2$.

De Finetti (1974, pp. 88–89, 188–190) gave a geometric argument to show that a set of marginal forecasts is coherent$_1$ if and only if it is coherent$_2$. Examples 1 and 2 illustrate this equivalence. Natural questions arise as to whether other loss functions might be appropriate replacements for squared error in Definition 2, and whether the equivalence to Definition 1 continues to hold. We address these questions in detail in this paper.

The definition of coherence$_1$ is very weak. That is, many forecasts are coherent$_1$, including some that might seem undesirable. For example, suppose that one makes only two forecasts, $P(A_1 \mid B) = 0.9$ and $P(A_2 \mid B) = 0.7$. Suppose also that $B \neq \Omega$ and $A_1 \cap A_2 = \varnothing$. These forecasts are coherent$_1$ because no book can be made against them. In particular, every combination of the form (1) takes the value 0 for all $\omega \in B^C$. Because $B \neq \Omega$, it is possible to coherently assign $P(B \mid \Omega) = 0$ (see Theorem 5). However, one might be uncomfortable giving conditional forecasts (conditional on the same event) to disjoint events that add up to more than 1. Some authors have suggested strengthening the definition of coherence to make it less liberal concerning conditional forecasts given events whose forecasts either are 0 or might coherently be assigned the value 0. Cozman and Seidenfeld (2009) provided a critical survey of some suggested strengthenings. Krauss (1968) and Dubins (1975) showed that for every finitely additive probability there exists a collection of coherent$_1$ conditional forecasts that also satisfies the axioms of probability conditional on events with 0 probability. That is,

$$\begin{aligned} &P(\cdot \mid B) \quad \text{is a probability measure for every } B \neq \varnothing, \\ &P(B \mid B) = 1 \text{ for every } B \neq \varnothing, \\ &P(A \cap C \mid B) = P(C \mid B)P(A \mid B \cap C) \text{ for all } A, B, \text{ and } C \text{ such that } B \cap C \neq \varnothing. \end{aligned} \tag{3}$$

Coherence$_1$ implies the conditions in (3) for all B with $P(B \mid \Omega) > 0$. Cozman and Seidenfeld (2009) showed that if one tries to impose conditions like (3), even when $P(B \mid \Omega) = 0$, then other undesirable consequences will follow. In this paper, we have not required that conditional forecasts satisfy additional requirements such as (3) when $P(B \mid \Omega) = 0$ is either stated or allowed (by extension via Theorem 5). The reason is that our results do not depend on whether the additional requirements hold. Whether different results hold if one further restricted the concept of coherence to require such additional properties is an open question.

Regarding the problem of strategic play in eliciting a decision maker's forecasts, which affects coherence$_1$, that concern is mitigated using coherence$_2$ because, under squared-error loss, the decision maker uniquely minimizes her/his expected loss for a set of forecasts by announcing the (subjective) probability for each event as the forecast. The Brier score is one of many strictly proper scoring rules (defined immediately below) that have this same feature of minimizing the expected score at the probability of each event.

Definition 3 (Scoring Rules). A *scoring rule* for scoring the conditional forecast $P(A \mid B)$ of an event A given another event B is a pair of extended real-valued functions (g_0, g_1) defined on the interval $[0, 1]$ with the following understanding. If A occurs, the forecaster suffers a loss of $I_B g_1(P(A \mid B))$, and if A^C occurs, the forecaster suffers a loss of $I_B g_0(P(A \mid B))$. The scoring rule (g_0, g_1) is *proper* if, for all events A and B, the forecaster's subjective conditional probability of A given B minimizes the expected score. That is, (g_0, g_1) is proper if $x = p$ minimizes $(1-p)g_0(x) + pg_1(x)$ for each $0 \leq p \leq 1$. A proper scoring rule (g_0, g_1) is *strictly proper* if, for all $0 \leq p \leq 1$, $x = p$ is the only value of x that minimizes $(1-p)g_0(x) + pg_1(x)$. For convenience, if a proper scoring rule is not strictly proper, we call it *merely proper*.

When $B \neq \Omega$, the expected score mentioned in Definition 3 is $P(B)$ times the conditional expected score given B. As mentioned earlier, some issues arise if $P(B) = 0$, but most of them do not affect the results of this paper. To avoid the one issue that does affect our results, we assume that $0 \leq P(A \mid B) \leq 1$ for every pair (A, B) of events with $B \neq \varnothing$. We make this assumption because many proper scoring rules do not extend naturally to forecasts outside of the unit interval. Although it is part of our goal to deal with incoherent forecasts, we cannot allow them to be quite so incoherent if we wish to apply proper scoring rules.

1.3. History and a Preview of Results

Savage (1971) generalized de Finetti's (1974) use of squared error loss in Definition 2 and characterized a general class of (strictly) proper scoring rules for eliciting personal probabilities for events. Murphy (1972) and Murphy and Epstein (1967) applied scoring rules to the evaluation of weather forecasters, whereas Gneiting and Raftery (2007) give an overview of scoring rules in a more general setting together with an application.

Savage (1971) did not devote much detail to the argument that the other proper scoring rules also agree with the criterion in Definition 1 in demarcating coherent from incoherent sets of forecasts. Predd et al. (2009) established that, for the case of finitely many marginal forecasts, de Finetti's (1974) geometric argument extends to all continuous strictly proper scoring rules. That is, with each continuous strictly proper scoring rule, if a finite set of marginal forecasts is incoherent$_1$, then it is dominated in score by some coherent$_1$ set of forecasts. And no finite coherent$_1$ set of marginal forecasts can be so dominated.

Definition 4 (Dominance). Let $(A_1, B_1), \ldots, (A_n, B_n)$ be a finite collection of pairs of events such that each A_i is to be forecast conditional on B_i. Suppose that the conditional forecast of A_i given B_i is to be scored by the scoring rule $(g_{0, A_i, B_i}, g_{1, A_i, B_i})$ for $i = 1, \ldots, n$. Define $g'_{A_i, B_i}(x, \omega) = I_{A_i \cap B_i}(\omega) g_{1, A_i, B_i}(x) + I_{A_i^C \cap B_i}(\omega) g_{0, A_i, B_i}(x)$. Let $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ be two different sets of conditional forecasts for the n events. We say that $\mathbf{q}$ *weakly dominates* $\mathbf{p}$ if, for all $\omega \in \Omega$,

$$\sum_{i=1}^{n} g'_{A_i, B_i}(q_i, \omega) \leq \sum_{i=1}^{n} g'_{A_i, B_i}(p_i, \omega), \tag{4}$$

with strict inequality for at least one ω. We say that $\mathbf{q}$ *strictly dominates* $\mathbf{p}$ if (4) holds for all ω with strict inequality for all ω.

In this paper, we extend the result of Predd et al. (2009) to conditional forecasts, we relax the assumption that the scoring rules be strictly proper, we allow more than finitely many forecasts, and we relax the assumption that the scoring rules be continuous. Although our results apply to arbitrarily sized collections of forecasts, the concept of dominance from Definition 4 is applied to each finite subcollection. This is a direct analogy to Definition 1 in which the coherence$_1$ or incoherence$_1$ of an arbitrary collection of forecasts is assessed by examining each finite subcollection. We also allow each forecast to be scored by a different scoring rule. Predd et al. (2009) make a passing remark that this is possible. They also make a passing remark concerning merely proper scoring rules and weak dominance. We prove a stronger version of that remark (Theorem 3).

The need for discontinuous scoring rules would arise in a situation like the following.

Example 3. An agent faces a decision problem in which the best action depends to a large extent on whether the probability of some event A is greater than or less than $1/2$. The agent consults an expert before determining $P(A \mid \Omega)$. The agent wants to hear the expert's subjective probability of A and so ties the expert's fee to a strictly proper scoring rule. To get the expert to be careful about on which side of $1/2$ the probability of A lies, she chooses a scoring rule with a jump discontinuity at $1/2$ such as

$$g_{0, A, \Omega}(x) = \begin{cases} x^2 & \text{if } x \leq 1/2, \\ 1/2 + x^2 & \text{if } x > 1/2; \end{cases}$$

$$g_{1, A, \Omega}(x) = \begin{cases} 1/2 + (1-x)^2 & \text{if } x \leq 1/2, \\ (1-x)^2 & \text{if } x > 1/2. \end{cases}$$

This is just the Brier score with a jump of $1/2$ at $x = 1/2$. (We prove that it is strictly proper in Example 8.) Every strictly proper scoring rule should give the expert incentive to provide her subjective probability of A. But if an expert has limited attention to pay to the task, the scoring rule in this example will focus the expert's attention on the important range of possible answers.

There is a second, decision-theoretic line of argument relating to the equivalence between the two senses of coherence, however, that was anticipated by de Finetti (1972, pp. 181–182) and sketched without much detail by Savage (1971, §§8 and 9.1). Consider a decision problem comprising a set of options $\mathcal{O}$ and subject to a loss function (bounded below) defined with respect to the finite partition Θ. If an option O^* fails to be a Bayes solution to the problem, i.e., if for each probability on Θ, O^* fails to minimize the expected loss with respect to the options in $\mathcal{O}$, then some randomized rule with support in $\mathcal{O}$ strictly dominates O^*. This result was established by Pearce (1984, Lemma 3, p. 1049) for the case in which $\mathcal{O}$ is finite. Here we extend this reasoning to statistical problems where Θ is finite, i.e., where "Nature" has only finitely many nonrandomized options, but the "Statistician" has a continuum of nonrandomized options. We apply the extension to decision problems in which the options are sets of forecasts and the loss function is the sum of the scoring rules.

In the remainder of this article we identify those cases in which an incoherent collection of forecasts is weakly or strictly dominated by a coherent collection of forecasts or by something else. The cases depend on whether the scoring rules are proper or strictly proper, and/or continuous or discontinuous. The cases also depend on whether or not the incoherent forecasts are Bayes decisions in the decision problems described in the previous paragraph. One consequence of what we prove is that Definitions 1 and 2 remain equivalent even if one replaces the Brier score by any collection of strictly proper scoring rules (one for each event being forecast) satisfying some mild assumptions. When the scoring rules are continuous, there is a coherent$_1$ set of forecasts that strictly dominates each finite incoherent$_1$ set of forecasts, just as with the Brier score. However, we illustrate that when the scoring rules are discontinuous (and even though strictly proper), there may fail to be a coherent$_1$ set of forecasts that weakly dominates a particular incoherent$_1$ set of forecasts.

2. Summary of Results

As we noted earlier, scoring rules are like loss functions. Loss functions play a key role in statistical decision theory that was introduced by Wald (1950). For this reason, we can make use of theorems from statistical decision theory to prove our results about forecasts and scoring rules. For an overview of statistical decision theory, see Chapter 3 of Schervish (1995). For each finite collection $\mathcal{A} = \{(A_1, B_1), \ldots, (A_k, B_k)\}$ of pairs of subsets of Ω with $B_i \neq \varnothing$, we construct a decision problem whose action space is the set of vectors of forecasts for the A_i given B_i. In each such decision problem, we check whether a particular set of forecasts is a Bayes decision with respect to some prior distribution, that is, whether the forecasts minimize the expected score under some probability on Ω.

The standard theorems of statistical decision theory usually assume that the loss function is bounded below. In our setting, this would correspond to assuming that the scoring rules are all bounded below. There are several good reasons for making such an assumption, one of which is illustrated in Example 5 in §4. Adding a finite constant to either or both branches of a proper scoring rule does not affect any of the properties that we are studying, i.e., merely proper versus strictly proper, continuity, dominance, or forecasts being Bayes. For this reason, assuming that scoring rules are bounded below is equivalent, for our purposes, to assuming that the greatest lower bound is 0. It is trivial from Definition 3 that every proper scoring rule (g_0, g_1) satisfies the following: $g_0(x)$ is minimized at $x = 0$, and $g_1(x)$ is minimized at $x = 1$. Hence, we lose no generality in assuming that $g_0(0) = g_1(1) = 0$. To summarize, we assume the following of proper scoring rules:

Assumption 1. *For $k = 0, 1$, g_k is bounded below, and $g_0(0) = g_1(1) = 0$.*

There are two other assumptions that play roles in some of our results, and we state them here for completeness.

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

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

Predd et al. (2009) included all three assumptions in the definition of a proper scoring rule. Although there do exist proper scoring rules that fail each of the assumptions, we give examples in §4 to illustrate which parts of our results rely on each assumption. Assumption 3 is satisfied by all strictly proper scoring rules (see Lemma 3).

Example 4 (Logarithmic Score). A popular alternative to the Brier score is the logarithmic scoring rule defined as

$$g_0(x) = -\log(1-x) \quad \text{and} \quad g_1(x) = -\log(x).$$

It is straightforward to verify that this scoring rule is strictly proper and that it satisfies all three assumptions above. A decision maker might choose this scoring rule if forecasts that are almost opposite to what occurs are nearly catastrophic. For example, this might be the case if a forecast close to 0 for an event that occurs is many orders of magnitude worse than a forecast of 1/2 for that same event.

If every conditional forecast has a corresponding scoring rule, and the loss is the sum of all the scores, we can generalize Definition 2 for event forecasts.

Definition 5 (Coherence$_3$). A collection of conditional forecasts for events is *incoherent$_3$* if for some finite subcollection of those forecasts there exists an alternative set of forecasts that strictly dominates in the sense of Definition 4. The collection of forecasts is called *coherent$_3$* otherwise.

The question then arises as to whether or not coherence$_3$ is equivalent to coherence$_1$ in all forecasting problems. The answer depends on which scoring rules one allows. It also depends on what one allows for an "alternative set of forecasts." It turns out that allowing a randomized forecast to serve as the alternative expands the collection of scoring rules that make coherence$_1$ and coherence$_3$ equivalent. (We are more explicit about what we mean by a randomized forecast in Definition 8.) For example, we prove the following:

Theorem 1. *Assume that each conditional forecast is scored by a strictly proper scoring rule that satisfies Assumptions 1 and 2.*

- *If randomized forecasts are allowed, then coherence$_1$ and coherence$_3$ are equivalent in every forecasting problem.*
- *If only nonrandomized forecasts are allowed and if all scoring rules are also continuous, then coherence$_1$ and coherence$_3$ are equivalent in every forecasting problem.*

The proof of Theorem 1 appears in the appendix. We provide an example (Example 8) to illustrate that one cannot, without further assumptions, obtain a dominating nonrandomized forecast when using discontinuous strictly proper scoring rules. We also provide an example (Example 7) to illustrate that incoherent$_1$ forecasts may not be strictly dominated, without further assumptions, when using strictly proper scoring rules that violate Assumption 2.

We prove additional results about the possibility of one set of forecasts being dominated by another. For example, Theorem 2 includes conditions under which dominance occurs with the use of merely proper scoring rules. Theorem 2 also includes conditions under which a weakly dominating, possibly randomized forecast exists. Theorem 3 gives conditions under which an incoherent$_3$ set of forecasts gets scored identically to a coherent$_1$ set of forecasts.

3. Mathematical Framework for Results

As mentioned earlier, we make use of some standard results from statistical decision theory, and hence we want to express the problem of comparing forecasts using proper scoring rules as a statistical decision problem.

3.1. Decision-Theoretic Framework

Each decision problem is indexed by a finite collection $\mathcal{A} = \{(A_1, B_1), \ldots, (A_n, B_n)\}$ of pairs of subsets of a set Ω with all B_i nonempty. The parameter space for each decision problem is the collection of constituent events determined by $\mathcal{A}$, as defined here.

Definition 6 (Constituents). Let $A_1, \ldots, A_n$, $B_1, \ldots, B_n$ be events. Construct the (at most 3^n) events $C_j = E_{1,j} \cap \cdots \cap E_{n,j}$, where each $E_{i,j} \in \{A_i \cap B_i, A_i^C \cap B_i, B_i^C\}$ for $i = 1, \ldots, n$. Let $a_i(j) = 1$ if $E_{i,j} = A_i \cap B_i$, and let $a_i(j) = 0$ if not. Also, let $b_i(j) = 1$ if $E_{i,j} \subseteq B_i$ and $b_i(j) = 0$ if $E_{i,j} = B_i^C$. The distinct nonempty sets $C_1, \ldots, C_m$ of this form are called the *constituents*.

The action space for the decision problem indexed by $\mathcal{A}$ is the set $[0, 1]^n$, where the ith coordinate is interpreted as the conditional forecast for A_i given B_i. The loss function is the total score from a collection of scoring rules, as in Definition 4, and which we make more explicit here.

Definition 7 (Total Scores). Let $\mathcal{A} = \{(A_1, B_1), \ldots, (A_n, B_n)\}$ be a finite collection of pairs of events with all B_i nonempty. Suppose that, for each i, the conditional forecast for A_i given B_i is scored by a proper scoring rule $(g_{0,A_i,B_i}, g_{1,A_i,B_i})$. Let $\mathbf{p} = (p_1, \ldots, p_n)$ be a vector of conditional forecasts. The *total scores* for these forecasts are defined as follows. For each constituent C_j, the total score is constant on C_j and equals

$$d_j = \sum_{i=1}^{n} b_i(j) g_{a_i(j), A_i, B_i}(p_i) = I_{C_j}(\omega) \sum_{i=1}^{n} g'_{A_i, B_i}(p_i, \omega), \quad (5)$$

where $a_i(j)$ and $b_i(j)$ are defined in Definition 6, and g'_{A_i,B_i} is defined in Definition 4.

To avoid ambiguity, we are explicit about what we mean by randomized rules (which we call randomized forecasts) in these decision problems.

Definition 8 (Randomized Forecast). Let $\{(A_1, B_1), \ldots, (A_n, B_n)\}$ be a finite collection of pairs of events. A *randomized forecast* is a probability measure δ on $[0,1]^n$ to be understood as the joint distribution of a random vector of conditional forecasts for $(A_1, \ldots, A_n)$ given $(B_1, \ldots, B_n)$, respectively. The *total scores* for the randomized forecast δ are, for $j = 1, \ldots, m$,

$$d_j = \int_{[0,1]^n} \sum_{i=1}^{n} b_i(j) g_{a_i(j), A_i, B_i}(p_i)\, \delta(d\mathbf{p}). \tag{6}$$

The definition of total scores in Definition 8 matches the definition of the loss function for a randomized rule in statistical decision theory. Of course, a nonrandomized forecast $\mathbf{p}$ can be interpreted as a randomized forecast δ by letting $\delta(\{\mathbf{p}\}) = 1$. In this case, (5) and (6) are the same.

Definition 9 summarizes the above construction of a decision problem.

Definition 9 (Problem $\mathcal{A}$). Let Ω be a space. For each pair (A, B) of subsets of Ω with $B \neq \varnothing$, let $(g_{0,A,B}, g_{1,A,B})$ be a proper scoring rule. For each finite collection $\mathcal{A} = \{(A_1, B_1), \ldots, (A_n, B_n)\}$ of pairs of subsets of Ω with all B_i nonempty, define the following decision problem (called *problem* $\mathcal{A}$). The parameter space is $\Theta = \{C_1, \ldots, C_m\}$, the constituents from Definition 6; the action space $\mathcal{O}$ is $[0,1]^n$, the set of all vectors $\mathbf{p}$ of possible conditional forecasts for $A_1, \ldots, A_n$ given $B_1, \ldots, B_n$, respectively; and the loss function is the total score. A randomized forecast δ is *Bayes in problem* $\mathcal{A}$ if there exists a probability distribution $\mathbf{q} = (q_1, \ldots, q_m)$ over Θ such that δ minimizes the expected loss, i.e.,

$$\sum_{j=1}^{m} q_j d_j = \inf_{\delta^*} \sum_{j=1}^{m} q_j \int_{[0,1]^n} \sum_{i=1}^{n} b_i(j) g_{a_i(j), A_i, B_i}(p_i)\, \delta^*(d\mathbf{p}), \tag{7}$$

where the inf is over all randomized forecasts δ^*.

The $\mathcal{A}$ in Definition 9 can be any finite collection of pairs of events. If $\mathcal{C}$ is another finite collection, problem $\mathcal{C}$ can be defined analogously.

Some simplification of the expression in (7) is possible. First, use the standard notation for the loss function of a randomized rule to denote, for each randomized rule δ,

$$g_{a_i(j), A_i, B_i}(\delta) = \int_{[0,1]^n} g_{a_i(j), A_i, B_i}(p_i)\, \delta(d\mathbf{p}).$$

Next, let R be the probability that extends $\mathbf{q}$ to the algebra of events generated by Θ. In particular,

$$R(A_i \cap B_i) = \sum_{j=1}^{m} a_i(j) b_i(j) q_j \quad \text{and}$$

$$R(A_i^C \cap B_i) = \sum_{j=1}^{m} [1 - a_i(j)] b_i(j) q_j.$$

Then (7) becomes

$$\begin{aligned} &\sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(\delta) + R(A_i^C \cap B_i) g_{0,A_i,B_i}(\delta)\} \\ &\quad = \inf_{\delta^*} \sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(\delta^*) \\ &\qquad\qquad + R(A_i^C \cap B_i) g_{0,A_i,B_i}(\delta^*)\}, \end{aligned} \tag{8}$$

a more familiar formula indicating that δ minimizes the expected total score. If δ is the nonrandomized forecast $\mathbf{p}$, then (8) becomes

$$\begin{aligned} &\sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(p_i) + R(A_i^C \cap B_i) g_{0,A_i,B_i}(p_i)\} \\ &\quad = \inf_{q_1, \ldots, q_n} \sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(q_i) \\ &\qquad\qquad + R(A_i^C \cap B_i) g_{0,A_i,B_i}(q_i)\}. \end{aligned} \tag{9}$$

The reader should note that Definition 9 assumes that the scoring rule used to score the conditional forecast of A given B is the same every time that (A, B) appears in a finite subcollection $\mathcal{A}$.

3.2. Equivalence of Definitions of Coherence

Because we deal with arbitrarily sized collections of forecasts, we want to be able to classify each such collection as Bayes or not in a manner similar to how an arbitrary collection of forecasts is classified as coherent$_1$ or not.

Definition 10 (Bayes Forecasts). Suppose that an agent must produce a conditional forecast for A given B for each pair of events (A, B) in the collection $\mathcal{C}$. Suppose that, for each $(A, B) \in \mathcal{C}$, the conditional

forecast for A given B is scored by a proper scoring rule $(g_{0,A,B}, g_{1,A,B})$. We say that a randomized forecast δ is *weakly Bayes* if, for every finite subcollection $\mathcal{A} = \{(A_1, B_1), \ldots, (A_n, B_n)\} \subseteq \mathcal{C}$, δ is Bayes in problem $\mathcal{A}$. A weakly Bayes forecast δ is *strongly Bayes* if there exists a finitely additive probability R on $(\Omega, 2^\Omega)$ that satisfies

$$\begin{aligned} &\sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(\delta) + R(A_i^C \cap B_i) g_{0,A_i,B_i}(\delta)\} \\ &\quad = \sum_{i=1}^{n} \{R(A_i \cap B_i) g_{1,A_i,B_i}(R(A_i \mid B_i)) \\ &\qquad\qquad + R(A_i^C \cap B_i) g_{0,A_i,B_i}(R(A_i \mid B_i))\} \end{aligned} \tag{10}$$

for every finite subcollection $\mathcal{A}$.

Weakly Bayes forecasts turn out to be the ones that are coherent$_3$.

Lemma 1. *If a collection of forecasts is weakly Bayes, then the forecasts in no finite subcollection are strictly dominated.*

The proof of Lemma 1 and the proofs of all other results stated in this section appear in the appendix.

Table 1 summarizes our results about the existence of dominating forecasts depending on what we assume about the scoring rules.

Lemma 1, together with the results in the second and third rows of Table 1, allow us to derive the following two results:

Corollary 1. *Assume that all scoring rules satisfy Assumptions 1–3 and that randomized forecasts are allowed. Then, a collection of forecasts is weakly Bayes if and only if it is coherent$_3$.*

Corollary 2. *Assume that all scoring rules satisfy Assumptions 1 and 3 and are continuous. Also assume that only nonrandomized forecasts are allowed. Then, a collection of forecasts is weakly Bayes if and only if it is coherent$_3$.*

What remains, to establish Theorem 1, is to show that weakly Bayes is equivalent to coherent$_1$. Our result assumes that all scoring rules are strictly proper.

Lemma 2. *Assume that all scoring rules satisfy Assumption 1. A coherent$_1$ set of forecasts is strongly Bayes. If all of the scoring rules are strictly proper, every collection of weakly Bayes forecasts is coherent$_1$.*

Table 1 **Summary of Assumptions and Conclusions of Results Providing Dominating or Equivalent Forecasts**

Conclusions	Assumptions	Examples to justify assumptions
A possibly randomized forecast weakly dominates	All scoring rules satisfy Assumptions 1 and 3, and all merely proper scoring rules satisfy Assumption 2	Example 5 and Lemma 3
A possibly randomized forecast strictly dominates	All scoring rules satisfy Assumptions 1–3	Example 6
A coherent$_1$ forecast strictly dominates	All scoring rules are continuous and satisfy Assumptions 1 and 3	Example 8
A coherent$_1$ forecast with the same total scores	All scoring rules are continuous and satisfy Assumption 1	Example 9

Notes. The first three results comprise Theorem 2 and assume that a collection of forecasts is given that is not weakly Bayes. The fourth result is Theorem 3 and assumes that a collection of weakly Bayes forecasts is given.

3.3. Weak and Strict Dominance in General

Theorem 2 is our general result containing conditions for the existence of dominating forecasts of various sorts when the initial forecasts are not weakly Bayes. Theorem 3 gives conditions for the existence of a set of coherent$_1$ forecasts with the same total scores when the initial set of forecasts is weakly Bayes. The various conclusions of these theorems are listed below:

Conclusion 1. There exists a finite subcollection $\{(A_1, B_1), \ldots, (A_n, B_n)\} \subseteq \mathcal{C}$ whose forecasts $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$ are not weakly Bayes, and for every such subcollection there exists a possibly randomized forecast that *weakly dominates* $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$.

Conclusion 2. There exists a finite subcollection $\{(A_1, B_1), \ldots, (A_n, B_n)\} \subseteq \mathcal{C}$ whose forecasts $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$ are not weakly Bayes, and for every such subcollection there exists a possibly randomized forecast that *strictly dominates* $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$.

Conclusion 3. There exists a finite subcollection $\{(A_1, B_1), \ldots, (A_n, B_n)\} \subseteq \mathcal{C}$ whose forecasts $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$ are not weakly Bayes, and for every such subcollection there exists a coherent forecast that *strictly dominates* $P(A_1 \mid B_1), \ldots, P(A_n \mid B_n)$.

Conclusion 4. There exists a coherent collection of forecasts that has the *same total scores* as

$\{P(A \mid B)\colon (A, B) \in \mathscr{C}\}$ for every finite subcollection $\{(A_1, B_1), \ldots, (A_n, B_n)\} \subseteq \mathscr{C}$.

THEOREM 2. *Suppose that an agent must provide a conditional forecast for each pair of events in the collection $\mathscr{C}$. Suppose that the agent chooses forecasts that are not weakly Bayes in the sense of Definition 10. Assume that all of the proper scoring rules (mentioned in Definition 10) satisfy Assumptions 1 and 3, and that all of the merely proper scoring rules satisfy Assumption 2. Then, Conclusion 1 holds. If, in addition, Assumption 2 holds for every scoring rule, then Conclusion 2 holds. If, in addition, all of the scoring rules are continuous, then Conclusion 3 holds.*

THEOREM 3. *Suppose that an agent must provide a conditional forecast for each pair of events in the collection $\mathscr{C}$. Suppose that the agent chooses forecasts that are weakly Bayes in the sense of Definition 10. Assume that all of the proper scoring rules (mentioned in Definition 10) are continuous and satisfy Assumptions 1 and 3. Then, Conclusion 4 holds.*

The proofs of Theorems 2 and 3 appear in the appendix.

4. Examples

In this section, we provide results and examples to illustrate why we make each of the assumptions in our various theorems. The examples involve only marginal forecasts. None of the assumptions that we make is needed solely because we allow both conditional and marginal forecasts. That is, even if we were to restrict attention solely to marginal forecasts, as Predd et al. (2009) did, our proofs would still use all of the assumptions to deal with the examples in this section. To simplify notation, we do not write all of these marginal forecasts as conditional on Ω. Instead, we leave off the "$|\Omega$" from each forecast in these examples. Similarly, the scoring rule for scoring the forecast of each event A is denoted $(g_{0,A}, g_{1,A})$ instead of $(g_{0,A,\Omega}, g_{1,A,\Omega})$. And the collections of pairs of events are written as collections of individual events because the second coordinate of each pair is implicitly Ω.

4.1. Assumption 1

If scoring rules are allowed to be unbounded both above and below, one runs the risk of encountering $\infty - \infty$ in even the most elementary calculations, such as total scores. The possibility of $\infty - \infty$ also makes the definition of Bayes' rule problematic.

Example 5 illustrates that incoherent$_1$ forecasts may not be even weakly dominated without Assumption 1. It also shows how a finite collection of forecasts can be weakly Bayes without being strongly Bayes.

EXAMPLE 5. Let $\mathscr{C} = \{A_1, A_2\}$ where $A_1 \subset A_2$. Suppose also that none of $C_1 = A_1$, $C_2 = A_2 \cap A_1^C$, and $C_3 = A_2^C$ is empty. The constituents are then C_1, C_2, and C_3. Let

$$(g_{0,A_1}(x), g_{1,A_1}(x)) = (x^2, (1-x)^2) \quad \text{and}$$
$$(g_{0,A_2}(x), g_{1,A_2}(x)) = (\log(x), \log(x) + 1/x).$$

The first is the Brier score, whereas the second is peculiar. To see that the second scoring rule is strictly proper, note that the expected score (when $\Pr(A_2) = p$) is

$$(1-p)\log(x) + p\left[\log(x) + \frac{1}{x}\right] = \log(x) + \frac{p}{x}. \tag{11}$$

The expression in (11) is smooth as a function of x for $x > 0$, and its derivative with respect to x is $1/x - p/x^2$. For $p > 0$, the derivative equals 0 if and only if $x = p$. Also, the second derivative is $-1/x^2 + 2p/x^3$, which is positive at $x = p$, so $x = p$ provides the unique minimum. For $p = 0$, the expression in (11) is also minimized uniquely at $x = 0$.

Now, suppose that an incoherent$_1$ agent assigns $P(A_1) = 1$ and $P(A_2) = 0$. The total scores are $d_1 = \infty$, $d_2 = \infty$, and $d_3 = -\infty$. No forecast can do better than $-\infty$ on the third constituent, so there is no set of forecasts that strictly dominates these incoherent$_1$ forecasts. The only way to match $-\infty$ on the third constituent is to forecast 0 for A_2. (For a randomized forecast, there must be positive probability of forecasting 0 for A_2.) No matter what one then forecasts for A_1, the total scores are now the same as those of the incoherent$_1$ forecast. So the incoherent$_1$ forecasts cannot be weakly dominated by another forecast, coherent$_1$ or otherwise.

The forecasts in this example are Bayes in problem $\mathscr{C}$ with respect to a prior R if and only if $R(C_3) = 1$. For the subcollection $\mathscr{A} = \{A_1\}$, the forecasts are Bayes in problem $\mathscr{A}$ with respect to R if and only if $R(A_1) = 1$. No prior gives probability 1 to both A_1 and C_3 because they are disjoint. Hence, although the choices $P(A_1) = 1$ and $P(A_2) = 0$ are weakly Bayes, they are not strongly Bayes.

4.2. Assumption 2

This is the assumption that each branch of the scoring rule is continuous at the point where it achieves its minimum. Discontinuity at this low end of the scoring rule can have curious consequences for the existence of dominating forecasts. Example 6 illustrates why we distinguish between Conclusions 1 and 2 depending on whether Assumption 2 holds for all scoring rules.

EXAMPLE 6. Let $\mathscr{C} = \{A_1, A_2\}$ with $A_2 = A_1^C$. Suppose also that neither A_1 nor A_2 is empty. The constituents are $C_1 = A_1$ and $C_2 = A_1^C$. Let $(g_{0,A_1}(x), g_{1,A_1}(x)) = (x^2, (1-x)^2)$. The other scoring rule is $g_{1,A_2}(x) = x - \log(x)$ and

$$g_{0,A_2}(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1 + x & \text{if } x > 0. \end{cases}$$

To see that the second scoring rule is strictly proper, note that the expected score (when $\Pr(A_2) = p$) is

$$\begin{cases} (1-p)(1+x) + p[x - \log(x)] & \text{if } x > 0, \\ \infty & \text{if } x = 0 \text{ and } p > 0, \\ 0 & \text{if } x = 0 \text{ and } p = 0. \end{cases}$$

Clearly, $x = 0$ minimizes this function if and only if $p = 0$. If $p > 0$, the function is smooth for $x > 0$ with derivative (with respect to x) equal to $1 - p/x$ and second derivative p/x^2. The derivative equals 0 if and only if $x = p$. Also, the second derivative is positive at $x = p$, so $x = p$ provides the unique minimum.

Now, suppose that an incoherent$_1$ agent assigns $P(A_1) = 1/2$ and $P(A_2) = 0$. The total scores are $d_1 = 1/4$ and $d_2 = \infty$. Every randomized forecast that has a score on the first constituent of less than 1/4 must choose a forecast of 0 for A_2 with probability greater than 3/4. Every randomized forecast that assigns positive probability to a 0 forecast for A_2 produces a score on the second constituent of ∞. So the incoherent forecasts cannot be strictly dominated. The forecasts $P'(A_1) = 1$ and $P'(A_2) = 0$ weakly dominate.

Example 7 shows that a collection of forecasts can be weakly Bayes without being strongly Bayes if some merely proper scoring rules fail to satisfy Assumption 2, even when all scoring rules satisfy Assumption 1. This is why Lemma 10 and Conclusion 1 in Theorem 2 assume that merely proper scoring rules satisfy Assumption 2.

EXAMPLE 7. Let $\mathscr{C} = \{A_0, A_1, A_2, \ldots\}$, where each event is nonempty, $A_1 \supset A_2 \supset \cdots$, and $A_0 \subseteq A_i$ for all $i > 0$. Suppose that an agent gives forecasts such that $0 < P(A_i) < 1$ for all i, but for $i > 0$, $P(A_i) \downarrow 0$. Suppose that (g_{0,A_i}, g_{1,A_i}) is the Brier score for all $i > 0$ but

$$g_{0,A_0}(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1/4 & \text{if } 0 < x < 1/2, \\ x^2 & \text{if } 1/2 \le x \le 1, \end{cases}$$

$$g_{1,A_0}(x) = \begin{cases} \infty & \text{if } x = 0, \\ 1/4 & \text{if } 0 < x < 1/2, \\ (1-x)^2 & \text{if } 1/2 \le x \le 1. \end{cases}$$

To see that this last scoring rule is proper, argue by cases. If $\Pr(A_0) = 0$, then the expected score is 0 if $x = 0$, and strictly positive otherwise. If $0 < \Pr(A_0) \le 1/2$, then the expected score is ∞ if $x = 0$, 1/4 if $0 < x \le 1/2$, and strictly greater than 1/4 otherwise. If $1/2 < \Pr(A_0) = p \le 1$, then the expected score is ∞ if $x = 0$, 1/4 if $0 < x \le 1/2$, and $(1-p)x^2 + p(1-x)^2$ if $1/2 < x \le 1$. The last quantity is uniquely minimized at $x = p$ with a value $p(1-p)$ that is strictly less than 1/4.

For every finite subcollection $\mathscr{A}$ of $\mathscr{C}$ that does not include A_0, the corresponding forecasts are Bayes in problem $\mathscr{A}$ because they are coherent$_1$. For each finite subcollection $\mathscr{A}$ that includes A_0, the corresponding forecasts are still Bayes in problem $\mathscr{A}$, because the score from $P(A_0)$ is identical to the score one would get by replacing $P(A_0)$ by any number strictly between 0 and the smallest of the $P(A_i)$ for $A_i \in \mathscr{A}$. Hence, the forecasts might as well be coherent$_1$ as far as the scores are concerned. So, the collection of forecasts is weakly Bayes.

However, the entire collection of forecasts is not strongly Bayes. The reason is that every finitely additive probability R that makes (10) true for all of the finite subcollections that do not include A_0 has $R(A_0) = 0$. But (10) does not hold with $R(A_0) = 0$ when the finite subcollection includes A_0.

4.3. Assumption 3

Scoring rules that violate Assumption 3 are pathological as the following result shows. Lemma 3 characterizes proper scoring rules that satisfy Assumption 1 but violate Assumption 3.

LEMMA 3. *Let g_0 and g_1 be functions that are bounded below. Assume that either $g_0(x)$ or $g_1(x)$ is infinite for at least one value of $x \in (0,1)$. Then (g_0, g_1) is a proper scoring rule if and only if the following conditions hold:*
- *For $k=0,1$, $g_k(x)$ is minimized at $x=k$.*
- *For all $0 \le x \le 1$, $\max\{g_0(x), g_1(x)\} = \infty$.*

Such a scoring rule is not strictly proper.

PROOF. The expected score for forecasting x when the probability of the event is p equals

$$(1-p)g_0(x) + pg_1(x). \tag{12}$$

It is now clear that the first condition is necessary and sufficient for (12) to be minimized at $x=p$ if $p \in \{0,1\}$. It is also clear that the second condition is sufficient for (12) to be minimized at $x=p$ for all $p \in (0,1)$. To see that the second condition is necessary, assume that $g_0(x_0) = \infty$ for some $x_0 \in (0,1)$. (A similar argument works if $g_1(x_1) = \infty$ for some $x_1 \in (0,1)$.) The only way that (12) can be minimized at $x=p$ when $p=x_0$ is for (12) to be infinite for all $0 \le x \le 1$. That is, the second condition must hold. Because the expected score is minimized at all x when $0<p<1$, such scoring rules are not strictly proper. □

Notice that the scoring rules in Lemma 3 all have an infinite expected score whenever a forecast is strictly between 0 and 1 and/or the probability of the event is strictly between 0 and 1.

4.4. Continuity

In our decision-theoretic framework, a discontinuous scoring rule corresponds to a loss function that will not satisfy the conditions of classical minimax and complete class theorems. (See §§3.2.4 and 3.2.5 of Schervish (1995) for more details.) In particular, when such loss functions are used, there may be dominated forecasts such that the only forecasts that dominate them are themselves dominated. Example 8 illustrates why we cannot expect a dominating coherent$_1$ set of forecasts with discontinuous scoring rules even though there are other incoherent$_1$ dominating forecasts.

EXAMPLE 8. Consider the scoring rule used in Example 3. If $p \in \{0,1\}$, the expected score is clearly minimized uniquely by forecasting $x=p$. If $0<p<1$, the expected score from forecasting x is

$$\begin{cases} p[1/2+(1-x)^2]+(1-p)x^2 & \text{if } x \le 1/2, \\ p(1-x)^2+(1-p)(1/2+x^2) & \text{if } x > 1/2. \end{cases}$$

If $p \le 1/2$, the first branch has a unique minimum at $x=p$, the second branch is strictly increasing, and there is a jump up immediately after $x=1/2$, so the expected score is minimized uniquely at $x=p$. (If $p=1/2$, the jump is of size 0.) Similarly, if $p>1/2$, the second branch has a unique minimum at $x=p$, the first branch is strictly decreasing, and there is a jump down immediately after $x=1/2$, so the expected score is minimized uniquely at $x=p$. Hence, the scoring rule is strictly proper.

Next, consider a case with $n=2$, $A_2 = A_1^C$, with neither event empty. Use the same scoring rule for both events. There are two constituents, $C_1 = A_1$ and $C_2 = A_1^C$. Let the incoherent$_1$ forecasts be $p_1 = 0.6$ and $p_2 = 0.7$. Then, $d_1 = 1.15$ and $d_2 = 0.95$. A forecast (r_1, r_2) is coherent$_1$ if and only if $r_1 + r_2 = 1$. For each coherent$_1$ forecast with $r_1 < 1/2$, the score on C_1 is $1+(1-r_1)^2+r_2^2$, which is always strictly greater than 1.5. For each coherent$_1$ forecast with $r_1 > 1/2$, the score on C_2 is $1+r_1^2+(1-r_2)^2$, which is always strictly greater than 1.5. For the coherent$_1$ forecast $r_1 = r_2 = 0.5$, the scores on both C_1 and C_2 are equal to 1.0. Hence, no coherent$_1$ forecast can weakly dominate the incoherent$_1$ forecast $p_1 = 0.6$ and $p_2 = 0.7$. On the other hand, there are other incoherent$_1$ forecasts that dominate (p_1, p_2). For example, $p_1' = 0.55$ and $p_2' = 0.65$ have total scores of $d_1' = 1.125$ and $d_2' = 0.925$, respectively.

When a scoring rule is merely proper, it is possible for incoherent$_1$ forecasts to be Bayes. If the probability with respect to which incoherent$_1$ forecasts are Bayes happens to assign probability 0 to at least one constituent, then there may be some strongly Bayes forecasts that are both coherent$_1$ and weakly dominated. (No Bayes forecast can be strictly dominated.) Example 9 illustrates why Theorem 3 does not deal with the case in which an incoherent$_1$ set of forecasts is Bayes but the scoring rules are merely proper and discontinuous.

EXAMPLE 9. Consider the collection $\mathscr{C} = \{A_1, A_2, A_3\}$ with $A_3 = (A_1 \cup A_2)^C$ and $A_1 \cap A_2 \neq \varnothing$.

The constituents are $C_1 = A_1 \cap A_2$, $C_2 = A_1 \cap A_2^C$, $C_3 = A_1^C \cap A_2$, and $C_4 = A_3$. The forecasts are $P(A_1) = P(A_2) = P(A_3) = 1/2$. The scoring rules (g_{0,A_i}, g_{1,A_i}) for $i=1,2$ are as follows:

$$g_0(x) = \begin{cases} 0 & \text{if } x < 1/2, \\ x^2 & \text{if } x \ge 1/2, \end{cases}$$

$$g_1(x) = \begin{cases} 1/2 & \text{if } x < 1/2, \\ (1-x)^2 & \text{if } x \ge 1/2. \end{cases}$$

To see that this is proper, let $0 \le p \le 1$. The expected score for forecasting x is

$$\begin{cases} p/2 & \text{if } x < 1/2, \\ p(1-x)^2+(1-p)x^2 & \text{if } x \ge 1/2. \end{cases}$$

This is minimized at $x=p$ for all p. Of course, if $p<1/2$, it is minimized at all $x<1/2$ also. The scoring rule is merely proper. Let (g_{0,A_3}, g_{1,A_3}) be $(2g_0, 2g_1)$. The forecasts are Bayes with respect to every probability R that satisfies $R(C_1) = 1/2$, $R(C_2) = 0$, $R(C_3) = 0$, and $R(C_4) = 1/2$. The total scores are $d_j = 1$ for all j. The alternative forecasts $q_1 = q_2 = q_3 = 0$ have total scores of $d_1 = d_4 = 1$ and $d_2 = d_3 = 1/2$, which weakly dominate the original strongly Bayes forecasts.

4.5. Conclusion 3

Example 10 illustrates why Theorem 2 has the weak Conclusion 3 rather than the stronger claim that the same set of coherent$_1$ forecasts dominates every incoherent$_1$ subcollection.

EXAMPLE 10. Let $\mathscr{C} = \{A_1, A_2, A_3, A_4, A_5\}$, where A_1, A_2, A_3 form a partition of Ω into three nonempty events, $A_4 = A_1 \cup A_2$, and $A_5 = A_2 \cup A_3$. Suppose that the forecast for each event is scored using the Brier score. Consider the following incoherent$_1$ forecasts: $P(A_1) = 1.0$, $P(A_2) = 0.3$, $P(A_3) = 1.0$, $P(A_4) = 0.1$, and $P(A_5) = 0.1$. The subcollection $\{A_3, A_4\}$ is not Bayes. The constituents are $C_1 = A_4$ and $C_2 = A_3$ with total scores of $d_1 = 1.81$ and $d_2 = 0.01$, respectively. Every set of forecasts that dominates these must give a forecast less than 0.1 to A_4. The subcollection $\{A_1, A_5\}$ is also incoherent$_1$. The constituents now are $B_1' = A_1$ and $B_2' = A_5$ with total scores of $d_1' = 0.01$ and $d_2' = 1.81$, respectively. Every set of forecasts that dominates these must give a forecast less than 0.1 to A_5. If a single set of forecasts was to dominate both of the two finite subcollections above, it would have to give forecasts less than 0.1 to each of A_4 and A_5. But $A_4 \cup A_5 = \Omega$, and hence no coherent$_1$ set of forecasts can dominate both of the incoherent$_1$ subcollections above.

5. Discussion

We have given sufficient conditions for a set of incoherent$_1$ (Definition 1) forecasts to be weakly or strictly dominated (according to proper scoring rules) by either a coherent$_1$ set of forecasts or by something else. Our conditions are not necessary. On the other hand, for each of our conditions, we have provided a counterexample to show that the condition cannot be eliminated without replacing it by some other condition that would rule out the counterexample. For instance, the condition in Theorem 2 that all of the merely proper scoring rules satisfy Assumption 2 is stronger than needed. With some extra work, one could prove that the only merely proper scoring rules that need to satisfy Assumption 2 are the ones that are flat either on the interval $(0, \epsilon)$ or on the interval $(1-\epsilon, 1)$ for some $\epsilon > 0$. (See Example 7 to see what can go wrong if Assumption 2 fails for such a scoring rule.) The basic idea is that so long as the set D in (18) does not contain a sequence of points $\{(e_{1,n}, e_{2,n})\}_{n=1}^{\infty}$ with $\lim_n (e_{1,n}+e_{2,n}) > 0$ and $e_{1,n}/(e_{1,n}+e_{2,n})$ arbitrarily close to one of the endpoints, the function $l_i^*(r)$ defined in the proof of Lemma 10 will be continuous at all r that matter. As another instance where our conditions are not necessary, the reader will note that our theorems do not make any assumptions on the types of events being forecast, or on the particular not-weakly-Bayes forecasts (aside from them not being weakly Bayes). Hence, we find that, in the proof of Lemma 13, we obtain the strongest Conclusion 3 without assuming continuity of scoring rules so long as the original forecasts have an infinite score in every constituent. Similarly, in Lemma 12, we obtain Conclusion 2 without assuming that any of our scoring rules satisfy Assumption 2. It is only in Lemma 14 that we make use of which scoring rules satisfy Assumption 2. But the distinction among these three lemmas is based on the forecasts, and we wanted the conclusions to Theorem 2 to hold for all forecasts simultaneously.

One somewhat surprising result that we found is the distinction between the conditions under which Conclusions 1 and 2 hold. Whether or not we can guarantee weak dominance does not depend on whether or not we are using merely proper scoring rules, but rather on a continuity property of the strictly proper scoring rules (Assumption 2). Predd et al. (2009) claimed that if one uses continuous merely proper scoring rules, one can guarantee a weakly dominating coherent$_1$ set of forecasts. Our results show that one gets either a strictly dominating coherent$_1$ set of forecasts (via the third part of Theorem 2) or a coherent$_1$ set of forecasts with identical scores (via Theorem 3). There is no middle ground in which one achieves weakly dominating but not identical scores.

Despite not having a complete characterization of all cases in which each of the four conclusions holds, we believe that we have delineated the cases very thoroughly. Our theorems apply regardless of which events are being forecast, regardless of what forecasts are given, and regardless of which proper scoring rule is used to score each event (so long as every scoring rule satisfies the conditions of the relevant theorem).

Our main results are formulated for forecasting events, where events are identified with their indicator functions. However, we noted in the introduction that de Finetti (1974) used the Brier score to establish the equivalence between coherence$_1$ and coherence$_2$ of a set of forecasts over the class of bounded variables, measurable with respect to some common measurable space. Our Theorems 6 and 7, on which the proof of our main result relies, apply to bounded random variables and general loss functions, not merely indicators for events. Thus, we have reason to explore generalizations of our principal results for forecasting bounded variables with proper scoring rules. The first thing that we would need is a general definition of a proper scoring rule for bounded random variables. If we are interested only in scoring forecasts, suppose that X is a bounded random variable and x is a forecast. The score could be some function $g(X,x)$. We could call g proper if, for every bounded random variable X, $E[g(X,x)]$ is minimized by $x=P(X)$. Some guidance in this direction is provided by Savage (1971). We conjecture that a collection of forecasts for bounded random variables is coherent$_1$ if and only if it is impossible to find an alternative collection of forecasts that lead to a uniformly smaller total score.

Acknowledgments
The authors thank Dan Osherson for bringing Predd et al. (2009) to their attention. They also thank the referees and associate editor for being exceptionally careful in reading earlier drafts of this manuscript and making suggestions.

Appendix. Proofs of Results

Some General Results About Scoring Rules

Some of our results rely on an understanding of the structure of general scoring rules. We make use of the following two results from Schervish (1989).

Theorem 4 (Schervish 1989, Theorem 4.2). *Let (g_0, g_1) be a left-continuous scoring rule that satisfies Assumptions 1–3 such that $g_1(x)$ does not jump to ∞ at $x=0$. The scoring rule is proper if and only if there exists a σ-finite measure λ on $[0,1)$ such that for all x,*

$$g_1(x)=\int_{[x,1)}(1-q)\lambda(dq) \quad \textit{and} \quad g_0(x)=\int_{[0,x)}q\lambda(dq). \tag{13}$$

The scoring rule is strictly proper if and only if, in addition, λ gives positive measure to every nondegenerate interval.

Lemma 4 (Schervish 1989, Lemma A.2). *Let (g_0, g_1) be a proper scoring rule. Let $0\le p\le 1$, and consider $m_p(x)=pg_1(x)+(1-p)g_0(x)$ as a function of x for fixed p. If g_1 and g_0 are bounded in a neighborhood of p, then m_p is continuous at $x=p$.*

We also need a few additional general results about scoring rules.

Lemma 5. *Suppose that a left-continuous merely proper scoring rule (g_0, g_1) satisfies Assumptions 1–3. Let $0\le p\le 1$, and suppose that $x=p'\neq p$ also minimizes $(1-p)g_0(x)+pg_1(x)$. Then, both g_0 and g_1 are constant on the open interval from p to p'. If g_0 and g_1 are continuous, then they are constant on the closed interval from p to p'.*

Proof. Assume that $p'>p$. The other case is similar. By (13), we have

$$pg_1(p')+(1-p)g_0(p')-[pg_1(p)+(1-p)g_0(p)] = \int_{[p,p')}(p'-q)\lambda(dq). \tag{14}$$

Because $p'-q>0$ for $q\in(p,p')$, the fact that the left-hand side of (14) equals 0 implies that $\lambda((p,p'))=0$. From the representation in Theorem 4, we see that both g_0 and g_1 are constant on each interval to which λ assigns 0 mass. In the continuous case, all such intervals are closed. □

Lemma 6. *Let (g_0, g_1) be a (strictly) proper scoring rule that satisfies Assumption 3. For each $0<x<1$, define $h_i(x)=\lim_{y\uparrow x} g_i(y)$. Then (h_0, h_1) is (strictly) proper.*

Proof. Let $m_p(x)$ be as in Lemma 4, and define $l_p(x)=ph_1(x)+(1-p)h_0(x)$. If there were an $x\in(0,1)$ at which one of g_0 or g_1 were discontinuous but not the other, then m_x would be discontinuous at x, which contradicts Lemma 4. It follows that g_0 and g_1 are discontinuous at the same set of points in $(0,1)$. From the definition of (h_0,h_1) we see that all four functions g_0, g_1, h_0, and h_1 share the same set of discontinuities. Let $0<p<1$ be a discontinuity point of (g_0,g_1) (if any). Then, by Lemma 4,

$$\begin{aligned} ph_1(p)+(1-p)h_0(p) &= p\lim_{x\uparrow p} g_1(x)+(1-p)\lim_{x\uparrow p} g_0(x) \\ &= \lim_{x\uparrow p}[pg_1(x)+(1-p)g_0(x) \\ &= pg_1(p)+(1-p)g_0(p). \end{aligned}$$

Hence, $l_p(p)=m_p(p)$ for all $0<p<1$. For $p\in\{0,1\}$, we also have $l_p(p)=m_p(p)$, because $g_i=h_i$ at both endpoints for $i=0,1$. For $x\neq p$ and $x\in\{0,1\}$, we have

$$\begin{aligned} ph_1(p)+(1-p)h_0(p) &= pg_1(p)+(1-p)g_0(p) \\ &\le pg_1(x)+(1-p)g_0(x). \end{aligned} \tag{15}$$

For $x\neq p$ and $0<x<1$, we have

$$\begin{aligned} ph_1(p)+(1-p)h_0(p) &\le \lim_{y\uparrow x}[pg_1(y)+(1-p)g_0(y)] \\ &= ph_1(x)+(1-p)h_0(x). \end{aligned} \tag{16}$$

Together, (15) and (16) imply that (h_0,h_1) is proper. If (g_0,g_1) is strictly proper, then the inequality is strict in (15). Assume by way of contradiction that the inequality is equality in (16). Apply Lemma 5 to (h_0,h_1) to conclude that h_0 and h_1 are both flat on the open interval between p and x. Theorem 4 implies that h_0 and h_1 are monotone, so they have at most countably many discontinuities. Hence, g_0 and g_1 have at most countably many discontinuities, and g_0 and g_1 are also both flat on the interval between p and x. This contradicts the fact that (g_0,g_1) is strictly proper. □

Lemma 7 extends one direction of Theorem 4 to general scoring rules.

Lemma 7. *Let (g_0,g_1) be a proper scoring rule that satisfies Assumptions 1–3. Then there exists a σ-finite measure λ on $[0,1)$ such that for all continuity points $x\in(0,1)$*

$$g_1(x)=\int_{(x,1)}(1-q)\lambda(dq) \quad \textit{and} \quad g_0(x)=\int_{(0,x)}q\lambda(dq).$$

Proof. Let (g_0,g_1) be a proper scoring rule, and create the left-continuous proper scoring rule (h_0,h_1) in Lemma 6. First, assume that $g_1(x)$ does not jump to ∞ at $x=0$. Note that $g_i(x)=h_i(x)$ for all continuity points x and $i=0,1$. The conclusion now follows from Theorem 4 applied to (h_0,h_1). Finally, if $g_1(x)$ jumps to ∞ at $x=0$, let $h'_1(0)=\lim_{x\downarrow 0} g_1(x)$ and let $h'_i(x)=h_i(x)$ for all other i and x. If we can show that (h'_0,h'_1) is proper, the above reasoning will finish the proof. The only way that (h'_0,h'_1) could fail to be proper is if there exists $p>0$ such that $ph'_1(0)+(1-p)h'_0(0)<ph'_1(p)+(1-p)h'_0(p)$. But both h'_0 and h'_1 are continuous at 0; hence, this inequality would imply that $ph_1(x)+(1-p)h_0(x)<ph_1(p)+(1-p)h_0(p)$ for some $0<x<p$, which contradicts (h_0,h_1) being proper. □

Lemma 8. *Suppose that (g_0,g_1) is a proper scoring rule that satisfies Assumptions 1–3. Define $m(p)=(1-p)g_0(p)+pg_1(p)$. Then,*

$$\lim_{p\to 0} m(p)=\lim_{p\to 1} m(p)=0.$$

Proof. We prove the limit at 0, because the limit at 1 is similar. Because $g_0(p)$ goes to 0, we need only prove that $pg_1(p)$ goes to 0. Suppose, to the contrary, that it does not go to 0. For $0<p<1$, $m(p)$ is the pointwise minimum of a collection of linear functions, and hence is concave and continuous on the open interval. It follows that $\lim_{p\to 0} pg_1(p)$ exists. Let the limit be $c>0$. From Lemma 7, for every continuity point p of g_0 and every continuity point $t\in(p,1)$,

$$pg_1(p)=p\int_{(p,t)}(1-q)\lambda(dq)+pg_1(t).$$

Hence, for every continuity point $t\in(0,1)$, $\lim_{p\to 0} p\int_{(p,t)}(1-q)\lambda(dq)=c$. Let $\{p_n\}_{n=1}^{\infty}$ be a sequence of continuity points of g_0 that converges to 0. In the integral above, $1-q>1-t$ for all $q\in(p,t)$; hence, for every continuity point $t\in(0,1)$, $p_n\lambda((p_n,t))$ eventually gets larger than $c/2$. Let $t>0$ be a continuity point small enough so that $g_0(t)<c/3$. It follows from Lemma 7 that, for all but finitely many n,

$$\frac{c}{3}>g_0(t)\ge\int_{(p_n,t)}q\lambda(dq)\ge p_n\lambda([p_n,t))>\frac{c}{2},$$

a contradiction. □

Equivalence of Definitions of Coherence

In this section, we prove Lemmas 1 and 2.

Proof of Lemma 1. Let $\mathscr{A}=\{(A_1,B_1),\ldots,(A_n,B_n)\}$ be a finite subcollection, and let $p_1,\ldots,p_n$ be the conditional forecasts. The left-hand side of (9) and the left-hand side of (10) are the expected total score under the probability R. If the forecasts in this subcollection were strictly dominated, then the dominating forecasts would have a strictly smaller score for every ω. Because there are only finitely many different total scores, the expected total score would be strictly smaller for every finitely additive probability; hence, the dominated forecasts could not satisfy (8) and they could not satisfy (10), and hence they would not be weakly Bayes. □

The proof of Lemma 2 is broken into a series of intermediate results.

Lemma 9. *Suppose that* (10) *holds and that the right-hand side of* (10) *is finite. Then, for each $i=1,\ldots,n$,*

$$\begin{aligned} &R(A_i\cap B_i)g_{1,A_i,B_i}(\delta)+R(A_i^C\cap B_i)g_{0,A_i,B_i}(\delta) \\ &\quad = R(A_i\cap B_i)g_{1,A_i,B_i}(R(A_i\mid B_i)) \\ &\qquad + R(A_i^C\cap B_i)g_{0,A_i,B_i}(R(A_i\mid B_i)). \end{aligned} \tag{17}$$

Proof. If $R(B_i)=0$, the result is trivial, so assume that $R(B_i)>0$. Because the scoring rule is proper, we know that for each i

$$\begin{aligned}R(A_i\mid B_i)g_{1,A_i,B_i}(\delta)+[1-R(A_i\mid B_i)]g_{0,A_i,B_i}(\delta)\\ \geq R(A_i\mid B_i)g_{1,A_i,B_i}(R(A_i\mid B_i))\\ +[1-R(A_i\mid B_i)]g_{0,A_i}(R(A_i\mid B_i)).\end{aligned}$$

If the inequality above were strict for some i with $R(B_i)>0$, then the left-hand side of (10) would be strictly larger than the right-hand side of (10). □

Lemma 10. *Assume that all scoring rules satisfy Assumption 1 and that all merely proper scoring rules satisfy Assumptions 2 and 3. If a collection of forecasts is weakly Bayes, then it is strongly Bayes.*

Proof. For every finite subcollection of $\mathscr{C}$, there exists a finitely additive probability R such that (10) holds. We show that there is an R that works for all finite subsets. Let $\mathscr{A}=\{(A_1,B_1),\ldots,(A_n,B_n)\}$ be an arbitrary finite subcollection of $\mathscr{C}$. Let $\mathscr{P}_{\mathscr{A}}=\{R\colon (10)\text{ holds}\}$. We would like to show first that $P_{\mathscr{A}}$ is closed in the topology of pointwise convergence, which is also the product topology on the function space $\mathscr{P}=[0,1]^{2^{\Omega}}$, which includes all finitely additive probabilities on Ω. All strictly proper scoring rules satisfy Assumption 3. Assumptions 1 and 3 guarantee that the right-hand side of (10) is always finite. Lemma 9 then says that for each $R\in\mathscr{P}_{\mathscr{A}}$ and each i, (17) holds. Hence, $\mathscr{P}_{\mathscr{A}}=\bigcap_{i=1}^{n}\mathscr{P}_{\{(A_i,B_i)\}}$. Next, we write each $\mathscr{P}_{\{(A_i,B_i)\}}$ as the inverse image of a closed set under a continuous function. For each $A\in 2^{\Omega}$, the coordinate projection function $f_A\colon \mathscr{P}\to[0,1]$, defined by $f_A(R)=R(A)$, is continuous. For each i, define the function

$$l_i(e_1,e_2)=\begin{cases}e_1g_{1,A_i,B_i}\left(\dfrac{e_1}{e_1+e_2}\right)+e_2g_{0,A_i,B_i}\left(\dfrac{e_1}{e_1+e_2}\right) & \text{if } e_1+e_2>0,\\ 0 & \text{otherwise,}\end{cases}$$

for $0\leq e_1,\ e_2\leq 1$. We can write

$$\mathscr{P}_{\{(A_i,B_i)\}}=(f_{A_i\cap B_i},f_{A_i^C\cap B_i})^{-1}(D_i),$$

where

$$\begin{aligned}D_i=\{(e_1,e_2)\colon e_1g_{1,A_i,B_i}(\delta)+e_2g_{0,A_i,B_i}(\delta)\\ =l_i(e_1,e_2),\text{ and } e_1+e_2\leq 1\}.\end{aligned}\qquad(18)$$

If D_i is closed, then so is $\mathscr{P}_{\{(A_i,B_i)\}}$. The argument that D_i is closed differs depending on whether or not the scoring rule $(g_{0,A_i,B_i},g_{1,A_i,B_i})$ is strictly proper. If the scoring rule is strictly proper, then the ith coordinate of the randomized forecast δ must be nonrandomized because only nonrandomized rules can be weakly Bayes with strictly proper scoring rules. Let δ assign probability 1 to the ith coordinate being p. In this case, $D_i=\{(e_1,e_2)\colon e_1(1-p)=e_2p\}$, which is a closed set.

If the scoring rule is merely proper, we argue as follows. Because D_i is a subset of $\mathbb{R}^2$, it is closed if it contains the limit of every convergent sequence. Let $\{(e_{1,n},e_{2,n})\}_{n=1}^{\infty}$ be a convergent sequence in D_i. Let the limit be $(e_{1,0},e_{2,0})$. We need to consider two cases. First, if $e_{1,0}+e_{2,0}=0$, then $(e_{1,0},e_{2,0})\in D_i$ trivially. For the rest of this part of the proof, assume that $e_{1,0}+e_{2,0}>0$. Define $h(e_1,e_2)=e_1/(e_1+e_2)$, and let $r_n=h(e_{1,n},e_{2,n})$ for $n=0,1,\ldots$. Then h is continuous at $(e_{1,0},e_{2,0})$ and r_n converges to r_0. We can write $l_i(e_1,e_2)=(e_1+e_2)l_i^*(h(e_1,e_2))$, where $l_i^*(r)=rg_{1,A_i,B_i}(r)+(1-r)g_{0,A_i,B_i}(r)$. $r_n=r_0$ for all n, and $(e_{1,0},e_{2,0})\in D_i$. Lemma 8 establishes that each l_i^* is continuous on the closed interval $[0,1]$. So, l_i is continuous at $(e_{1,0},e_{2,0})$, which is then in D_i. Hence, D_i is closed and so is $\mathscr{P}_{\{(A_i,B_i)\}}$. It follows that each $\mathscr{P}_{\mathscr{A}}$ is closed.

Finally, we show that the intersection of all $\mathscr{P}_{\mathscr{A}}$ is nonempty. It is easy to see that if $\mathscr{B}$ is a finite subcollection such that $\mathscr{B}\subseteq\mathscr{A}$, then $\mathscr{P}_{\mathscr{A}}\subseteq\mathscr{P}_{\mathscr{B}}$. It follows that the collection of all $\mathscr{P}_{\mathscr{A}}$ has the finite intersection property. Because the set of finitely additive probabilities is compact in the product topology, it follows that the intersection of all $\mathscr{P}_{\mathscr{A}}$ is nonempty. That is, there is at least one finitely additive probability R such that (10) holds for all finite subcollections. Hence, the forecasts are strongly Bayes. □

The connection between coherent$_1$ forecasts and strongly Bayes forecasts needed for Lemma 2 relies on the ability to extend a collection of coherent$_1$ forecasts into a linear functional on a linear space. The next two theorems extend de Finetti's (1974, §3.10) fundamental theorem of probability to deal explicitly with conditional forecasts. The fundamental theorem of probability applies to more general random variables than indicators of events. In keeping with de Finetti's (1974) original presentation, we state and prove the next two results and their corollary for bounded random variables rather than merely for indicators of events. In this paper, we use the results only for indicators of events. The word *prevision* in the statements and proofs of these results can be understood as "elicited expected value" in the same way that forecast stands for elicited probability.

Theorem 5 (Fundamental Theorem of Prevision). *Let $\mathscr{C}$ be a set of pairs where the first element of each pair is a bounded random variable and the second is a nonempty event. For each $(X,B)\in\mathscr{C}$, let $P(X\mid B)$ be a conditional prevision. Assume that the conditional previsions are coherent$_1$. Let $(X,\Omega)\notin\mathscr{C}$. Then there exists a closed interval $[c,d]$ such that $P(X\mid\Omega)=x$ is coherent$_1$ with all the other conditional previsions if and only if $c\leq x\leq d$.*

Proof. Define the linear space

$$\mathscr{Y}=\left\{\sum_{i=1}^{n}\alpha_iI_{B_i}[X_i-P(X_i\mid B_i)]+f\colon (X_i,B_i)\in\mathscr{C}\text{ and } \alpha_i\in\mathbb{R}\text{ for } i=1,\ldots,n\text{ and } f\in\mathbb{R}\right\}.\qquad(19)$$

For each $Y\in\mathscr{Y}$ expressed as in (19), let $L(Y)=f$. The coherence$_1$ of the previsions makes it clear that L is well defined. (If the same Y could be expressed two different ways with different values of f, then book could be made by trading the two different representations of Y against each other.) It is also easy to see that L is a linear functional defined on $\mathscr{Y}$. Define the following two sets

$$\underline{P}=\{Y\in\mathscr{Y}\colon Y\leq X\},$$

$$\bar{P}=\{Y\in\mathscr{Y}\colon Y\geq X\}.$$

Set $c=\sup_{Y\in\underline{P}}L(Y)$ and $d=\inf_{Y\in\bar{P}}L(Y)$. For the "if" direction, suppose that $c\leq x\leq d$. Suppose, to the contrary, that there exist $(X_1,B_1),\ldots,(X_n,B_n)$ in $\mathscr{C}$ and real numbers $\alpha_1,\ldots,\alpha_n$, and $\beta\neq 0$ and $\epsilon>0$ such that

$$\beta(X-x)+\sum_{i=1}^{n}\alpha_iI_{B_i}[X_i-P(X_i\mid B_i)]<-\epsilon.\qquad(20)$$

If $\beta>0$, then (20) implies

$$X<x-\frac{\epsilon}{\beta}-\sum_{i=1}^{n}\frac{\alpha_i}{\beta}I_{B_i}[X_i-P(X_i\mid B_i)].\qquad(21)$$

The right side of (21) is an element of $\bar{P}$, and hence $d\leq x-\epsilon/\beta$, which contradicts $x\leq d$. Similarly, if $\beta<0$, we arrive at a contradiction to $c\leq x$. For the "only if" direction, suppose that $P(X\mid\Omega)=x$ is coherent$_1$ with the other previsions. Suppose, to the contrary, that $x<c$. Let $Y=\sum_{i=1}^{n}\alpha_iI_{B_i}[X_i-P(X_i\mid B_i)]+f\in\underline{P}$ be such that $Y\leq X$ and $f>(c+x)/2$. The following gambles make book against these previsions:

$$-(X-x)+(Y-f)\leq x-f<\frac{x-c}{2}<0.$$

Similarly, if $x>d$, we can find a $Y\in\bar{P}$ that allows us to make book. □

Theorem 6 (Fundamental Theorem of Conditional Prevision). *Let $\mathscr{C}$ be a set of pairs where the first element of each pair is a bounded random variable and the second is a nonempty event. For each $(X,B)\in\mathscr{C}$, let $P(X\mid B)$ be a conditional prevision. Assume that the conditional previsions are coherent$_1$. Let $(X,D)\notin\mathscr{C}$ with $D\neq\varnothing$. Then there exists a set E of real numbers such that $P(X\mid D)=x$ is coherent$_1$ with all the other conditional previsions if and only if $x\in E$.*

Proof. First, suppose that both (D,Ω) and (XI_D,Ω) are in $\mathscr{C}$. De Finetti (1974) proves that a necessary and sufficient condition for $P(X\mid D)$ to be coherent$_1$ is $P(XI_D\mid\Omega)=P(X\mid D)P(D\mid\Omega)$. If $P(D\mid\Omega)>0$, then $E=\{P(XI_D\mid\Omega)/P(D\mid\Omega)\}$. If $P(D\mid\Omega)=0$, then $E=\mathbb{R}$. Next, suppose that $(D,\Omega)\in\mathscr{C}$ but $(XI_D,\Omega)\notin\mathscr{C}$. If $P(D\mid\Omega)=0$, then $E=\mathbb{R}$. If $P(D\mid\Omega)\neq 0$, apply Theorem 5 to find an interval $[c,d]$ of possible coherent$_1$ values for $P(XI_D\mid\Omega)$. Then $E=\{x/P(D\mid\Omega)\colon c\leq x\leq d\}$. Next, assume that $(D,\Omega)\notin\mathscr{C}$ but $(XI_D,\Omega)\in\mathscr{C}$. Apply Theorem 5 to find an interval $[c,d]$ of possible coherent$_1$ values for $P(D\mid\Omega)$. If $c=0$, then $E=\mathbb{R}$. If $c>0$, then $E=\{P(XI_D\mid\Omega)/x\colon c\leq x\leq d\}$. Finally, assume that neither (D,Ω) nor (XI_D,Ω) is in $\mathscr{C}$. Apply Theorem 5 to find an interval $[c_1,d_1]$ of possible coherent$_1$ values of $P(D\mid\Omega)$. For each $x\in[c_1,d_1]$, apply the argument above for the case in which $(D,\Omega)\in\mathscr{C}$ but $(XI_D,\Omega)\notin\mathscr{C}$ to find a set E_x of possible coherent$_1$ values of $P(X\mid D)$. Then $E=\bigcup_{x\in[c_1,d_1]}E_x$. □

Corollary 3. *Let $\mathscr{C}_1$ and $\mathscr{C}_2$ be two disjoint sets of pairs where the first element of each pair is a bounded random variable and the second is a nonempty event. For each $(X,B)\in\mathscr{C}_1$, let $P(X\mid B)$ be a conditional prevision. Assume that the conditional previsions are coherent$_1$. For each $(X,B)\in\mathscr{C}_2$, there exists a conditional prevision $P(X\mid B)$ such that $\{P(X\mid B)\colon (X,B)\in\mathscr{C}_1\cup\mathscr{C}_2\}$ are coherent$_1$.*

Proof. Use Zermelo's lemma to well order the elements of $\mathscr{C}_2$. Let D be the corresponding set of ordinals. We use transfinite induction to finish the proof. For each successor ordinal $\alpha\in D$, apply Theorem 6 to find a conditional prevision $P(X_\alpha\mid B_\alpha)$ that is coherent$_1$ with all earlier previsions. For each limit ordinal $\beta\in D$, it is easy to see that the previsions $\{P(X_\alpha\mid B_\alpha)\colon \alpha<\beta\}\cup\{P(X\mid B)\colon (X,B)\in\mathscr{C}_1\}$ are coherent$_1$ because every finite subcollection was verified as coherent$_1$ at an earlier stage in the induction. □

Lemma 11. *Let $\mathscr{C}$ be a collection of pairs of events. A collection of conditional forecasts $\{P(A\mid B)\colon (A,B)\in\mathscr{C}\}$ is coherent$_1$ if and only if there exists a finitely additive probability R on $(\Omega,2^{\Omega})$ that agrees with P on $\mathscr{C}$ in the following sense: For each $(A,B)\in\mathscr{C}$, $R(B)P(A\mid B)=R(A\cap B)$.*

Proof. For the "if" part, assume that such an R exists. Let

$$\mathscr{A}=\{(A_1,B_1),\ldots,(A_n,B_n)\}\subseteq\mathscr{C}$$

be a finite subcollection. It is trivial to extend R to a positive linear functional L on the linear span $\mathscr{L}$ of constants and the indicators $I_{A_i\cap B_i}$ and I_{B_i} for $i=1,\ldots,n$ by

$$L\left(c+\sum_{i=1}^{n}a_iI_{A_i\cap B_i}+\sum_{i=1}^{n}b_iI_{B_i}\right)=c+\sum_{i=1}^{n}a_iR(A_i\cap B_i)+\sum_{i=1}^{n}b_iR(B_i).$$

Suppose, to the contrary, that there exist $\alpha_1,\ldots,\alpha_n$ and $\epsilon>0$ such that

$$X=\sum_{i=1}^{n}\alpha_iI_{B_i}[I_{A_i}-P(A_i\mid B_i)]<-\epsilon.\qquad(22)$$

Note that X in (22) is an element of $\mathscr{L}$. Because R agrees with the conditional previsions, X equals

$$\sum_{i=1}^{n}\alpha_i[I_{A_i\cap B_i}-R(A_i\cap B_i)]-\sum_{i=1}^{n}\alpha_iP(A_i\mid B_i)[I_{B_i}-R(B_i)].\qquad(23)$$

It follows from (23) that $L(X)=0$, but (22) implies that $L(X)<-\epsilon$, a contradiction. Hence, no book can be made, and the conditional previsions are coherent$_1$. For the "only if" part, assume that the conditional previsions are coherent$_1$. Extend the collection of conditional previsions to include all pairs (B, Ω) and $(A\cap B, \Omega)$ for each $(A, B)\in \mathscr{C}$ using Corollary 3. Let $\mathscr{L}'$ be the linear span of all constants and indicators $I_{A\cap B}$ and I_B for $(A, B)\in\mathscr{C}$. On $\mathscr{L}'$, define

$$L'\left(c+\sum_{i=1}^{n} a_i I_{A_i\cap B_i}+\sum_{i=1}^{n} b_i I_{B_i}\right)=c+\sum_{i=1}^{n} a_i P(A_i\cap B_i\mid\Omega)+\sum_{i=1}^{n} b_i P(B_i\mid\Omega).$$

Note that L' satisfies $L'(X)\le\|X\|_\infty$ and $L'(1)=1$. According to the Hahn-Banach theorem, L' can be extended to a linear functional on the linear span of all indicators of subsets of Ω. This extension, when restricted to the indicators of events, is a finitely additive probability R that agrees with P on $\mathscr{C}$. □

Proof of Lemma 2. Let $\{P(A\mid B)\colon (A, B)\in\mathscr{C}\}$ be a collection of conditional forecasts. To prove the first claim, assume that the forecasts are coherent$_1$. Let R be as in Lemma 11. To prove the second claim, assume that all of the scoring rules are strictly proper and that the forecasts are weakly Bayes. Because all strictly proper scoring rules satisfy Assumption 3, the forecasts are strongly Bayes by Lemma 10. Let R be as in Definition 10, and let $\{(A_1, B_1), \ldots, (A_n, B_n)\}\subseteq\mathscr{C}$ be a finite subcollection. The right-hand side of (10) is always finite. Lemma 9 says that (17) holds for all $i=1,\ldots,n$. Because the scoring rules are strictly proper, $P(A_i\mid B_i)=R(A_i\mid B_i)$ for all i such that $R(B_i)>0$. Because this is true for every finite subset, R agrees with P on all of $\mathscr{C}$. Lemma 11 completes the proof. □

Theorem 3

Proof of Theorem 3. We have enough assumptions to apply Lemma 10 so that the forecasts are strongly Bayes. Let R be as in Definition 10. Let $\{(A_1, B_1), \ldots, (A_n, B_n)\}\subseteq\mathscr{C}$ be a finite subcollection. We can also apply Lemma 9. For each i such that $P(A_i\mid B_i)\neq R(A_i\mid B_i)$ (if there are any), apply Lemma 5 to (g_{0,A_i}, g_{1,A_i}) to conclude that $g_{k,A_i}(P(A_i)) = g_{k,A_i}(R(A_i))$ for $k=0,1$. For all i such that $P(A_i)=R(A_i)$ (if there are any), we already have $g_{k,A_i}(P(A_i))=g_{k,A_i}(R(A_i))$ for $k=0,1$. Because the finite subcollection was arbitrary, Conclusion 4 now follows. □

Theorem 2 and Its Corollaries

The proof of Theorem 2 relies on a general result from decision theory, a strengthening of the standard minimax theorem based on a construction of Pearce (1984).

Definition 11 (Lower Boundary). Let $\Omega=\{\theta_1,\ldots,\theta_m\}$ be a finite parameter space and let $\aleph$ be an action space. The *Risk set* is

$$R=\{(R(C_1,\delta),\ldots,R(C_m,\delta))\colon \delta \text{ is a randomized rule}\}.$$

The *lower boundary* of the risk set is

$$\partial_L=\{(x_1,\ldots,x_m)\in\bar{R}\colon y_i\le x_i \text{ for all } i \text{ and } y_i<x_i \text{ for some } i \text{ implies } (y_1,\ldots,y_m)\notin R\}.$$

The risk set is *closed from below* if $\partial_L\subseteq R$.

Theorem 7. *Let $\Theta=\{\theta_1,\ldots,\theta_m\}$ be a finite parameter space. Let $\mathscr{A}$ be an action space. Let $\mathrm{L}\colon\Theta\times\mathscr{A}\to\mathbb{R}$ be a loss function that is bounded below. Let $a_0\in\mathscr{A}$ be an action that is not Bayes for even a single prior distribution and such that $\mathrm{L}(\theta_j, a_0)<\infty$ for all j. Then, there exists a randomized rule that strictly dominates a_0. If the risk set for the decision problem is closed from below, then there is a dominating rule that is a Bayes rule with respect to some prior.*

Proof. Replace L by $L'(\theta_j, a)=L(\theta_j, a)-L(\theta_j, a_0)$. Then L' is still bounded below, and the risk set is closed from below if and only if the original risk set was closed from below. The risk function of a randomized rule δ is

$$R(\theta_j,\delta)=\int_{\mathscr{A}} L'(\theta_j, a)\delta(da).$$

The Bayes risk of δ with respect to a prior $\mathbf{s}=(s_1,\ldots,s_m)$ is

$$r(\mathbf{s},\delta)=\sum_{j=1}^{m} s_j R(\theta_j,\delta).$$

The minimax theorem (for example, Schervish 1995, Theorem 3.77) says that the decision problem has a least favorable distribution $\mathbf{u}=(u_1,\ldots,u_m)$ and a minimax value

$$\inf_\delta\sup_j R(\theta_j,\delta)=\inf_\delta r(\mathbf{u},\delta).$$

By construction, the nonrandomized rule a_0 is an equalizer with $L'(\theta_j, a_0)=0$ for each j. Because a_0 is not a Bayes rule with respect to the least favorable distribution, its expected loss (namely 0) is strictly greater than the minimax value. Hence, there exists a rule δ such that $R(\theta_j,\delta)<0$ for all j, and so

$$\int_{\mathscr{A}} L(\theta_j, a)\delta(da)<L(\theta_j, a_0)\quad\text{for all } j.$$

This completes the proof of the first claim.

If the risk set is closed from below, it follows from Theorem 3.77 of Schervish (1995) that there is a minimax rule δ_0 that is also a Bayes rule with respect to $\mathbf{u}$. □

The proof of Theorem 2 begins by noting that we have enough assumptions to apply Lemma 10; hence, there is a finite subcollection $\mathscr{A}=\{(A_1, B_1),\ldots,(A_n, B_n)\}$ such that (10) fails for every finitely additive probability R. It follows that the forecasts $p_i=P(A_i\mid B_i)$ for $i=1,\ldots,n$ for the pairs in $\mathscr{A}$ are not weakly Bayes. Let $C_1,\ldots,C_m$ be the distinct nonempty constituents from Definition 6, and let $d_1,\ldots,d_m$ be the total scores from Definition 7.

The remainder of the proof is split into three cases depending on whether the set

$$J=\{j\colon d_j=\infty\}$$

and/or its complement is empty. Together, Lemmas 12, 13, and 14 establish Theorem 2.

Lemma 12. *Under the conditions of Theorem 2, if $J=\varnothing$, the conclusions to Theorem 2 hold.*

Proof. For each finite collection $\mathscr{A}=\{(A_1, B_1),\ldots,(A_n, B_n)\}$ whose forecasts are not weakly Bayes, construct problem $\mathscr{A}$ as in Definition 9. The loss function (and risk function) is

$$L(C_j,\mathbf{q})=\sum_{i=1}^{n} b_i(j) g_{a_i(j), A_i}(q_i).$$

The action $\mathbf{p}=(p_1,\ldots,p_n)\in\mathscr{O}$ is not Bayes. Apply Theorem 7 to achieve Conclusion 2, which implies Conclusion 1.

If, in addition, all of the scoring rules are continuous, then the risk set is closed from below. Apply the last part of Theorem 7 to obtain a dominating rule δ_0 that is also a Bayes rule with respect to a prior $\mathbf{u}=(u_1,\ldots,u_m)$. Because $\mathbf{u}$ is a probability vector, it corresponds to an essentially unique set of forecasts $(r_1,\ldots,r_n)$, where r_i is the conditional probability of A_i given B_i inferred from the probabilities of the constituents. Specifically, let $R(A_i\cap B_i)=\sum_{j=1}^{m} a_i(j)b_i(j)u_j$ and $R(A_i^C\cap B_i)=\sum_{j=1}^{m}[1-a_i(j)]b_i(j)u_j$ so that $R(B_i)=\sum_{j=1}^{m} b_i(j)u_j$ and

$$r_i=\begin{cases}\dfrac{R(A_i\cap B_i)}{R(B_i)} & \text{if } R(B_i)>0,\\ \text{arbitrary} & \text{if } R(B_i)=0.\end{cases}$$

The Bayes risk of δ_0 is

$$\inf_\delta\sum_{j=1}^{m} b_i(j)u_j\int_{\mathscr{O}}\sum_{i=1}^{n} g_{a_i(j), A_i, B_i}(q_i)\delta(d\mathbf{q})$$
$$=\inf_\delta\int_{\mathscr{O}}\sum_{i=1}^{n} R(B_i)[r_i g_{1,A_i,B_i}(q_i)+(1-r_i)g_{0,A_i,B_i}(q_i)]\delta(d\mathbf{q}).$$

Each summand inside the integral can be minimized separately by $q_i=r_i$; Hence, δ_0 has the same Bayes risk as the nonrandomized rule $\mathbf{r}=(r_1,\ldots,r_n)$. So, we assume that δ_0 is the nonrandomized rule. The risk function for this dominating nonrandomized rule is $\sum_{i=1}^{n} b_i(j)g_{a_i(j), A_i, B_i}(r_i)<d_j$ for each j. Hence, we have Conclusion 3. □

Lemma 13. *Under the conditions of Theorem 2, if $J^C=\varnothing$, the conclusions to Theorem 2 hold.*

Proof. Let $\mathscr{A}=\{(A_1, B_1),\ldots,(A_n, B_n)\}$ be a finite collection whose forecasts are not weakly Bayes. Let $s_1=\cdots=s_m=1/m$, and define, for $i=1,\ldots,n$, $Q(A_i\cap B_i)=\sum_{j=1}^{m} a_i(j)b_i(j)s_j$ and $Q(A_i^C\cap B_i)=\sum_{j=1}^{m}[1-a_i]b_i(j)s_j$ so that $Q(B_i)=\sum_{j=1}^{m} b_i(j)s_j$. Then,

$$q_n=\begin{cases}\dfrac{Q(A_i\cap B_i)}{Q(B_i)} & \text{if } Q(B_i)>0,\\ \text{arbitrary} & \text{if } Q(B_i)=0.\end{cases}$$

Because $q_1,\ldots,q_n$ correspond to a probability, they are coherent$_1$ conditional forecasts and they have finite total scores in all constituents. Hence, Conclusion 3 holds, which implies Conclusions 1 and 2. □

Lemma 14. *Under the conditions of Theorem 2, if neither J nor J^C is empty, the conclusions to Theorem 2 hold.*

Proof. Let $\mathscr{A}=\{(A_1, B_1),\ldots,(A_n, B_n)\}$ be a finite collection whose forecasts are not weakly Bayes. Because the scoring rules are all finite, except possibly at the endpoints, the only way to get an infinite score is for one of the events A_i to get an extreme forecast that is not always correct. That is, either $p_i=0$ but $A_i\neq\varnothing$, or $p_i=1$ but $A_i\not\supseteq B_i$. (And, of course, the scoring rule corresponding to (A_i, B_i) has to be unbounded at the appropriate endpoint.) Define

$$I=\{i\colon \text{ either } (g_{1,A_i,B_i}(p_i)=\infty \text{ and } A_i\neq\varnothing) \text{ or } (g_{0,A_i,B_i}(p_i)=\infty \text{ and } A_i\not\supseteq B_i)\}.$$

Because $J\neq\varnothing$, it follows that $I\neq\varnothing$. For each $i\in I^C$ (if any), let

$$(h_{0,A_i,B_i}, h_{1,A_i,B_i})=(g_{0,A_i,B_i}, g_{1,A_i,B_i}).$$

For each $i\in I$, replace $(g_{0,A_i,B_i}, g_{1,A_i,B_i})$ by the Brier score $(h_{0,A_i,B_i}(x), h_{1,A_i,B_i}(x))=(x^2,(1-x)^2)$. In the remainder of the proof, when we wish to refer to scores under the original scoring rules, we call them the "g-scores." When we wish to refer to scores under the modified scoring rules, we call them the "h-scores."

Next, we show that the original forecasts $\mathbf{p}=(p_1,\ldots,p_n)$ are not weakly Bayes under the h-scores. Suppose, to the contrary, that they are weakly Bayes under the h-scores. We have enough assumptions to apply Lemma 10 so that the forecasts are strongly Bayes under the h-scores. Let R be as in Definition 10, and let $r_{1,i}=R(A_i\cap B_i)$, $r_{2,i}=R(A_i^C\cap B_i)$, and $r_i=r_{1,i}/(r_{1,i}+r_{2,i})$ for $i=1,\ldots,n$. (If $r_{1,i}+r_{2,i}=0$ for some i, then $r_{1,i}/(r_{1,i}+r_{2,i})$ should be interpreted as some arbitrary number in [0, 1].) For each $i=1,\ldots,n$, $x=p_i$ must minimize $r_{1,i}h_{1,A_i}(x)+r_{2,i}h_{0,A_i}(x)$. Because the h-scoring rules are strictly proper for $i\in I$, we must have $p_i=r_i$ for all $i\in I$ such that $r_{1,i}+r_{2,i}>0$. Because all of the p_i for $i\in I$ are extreme, it follows that $r_{1,i}h_{1,A_i}(p_i)+r_{2,i}h_{0,A_i}(p_i)=0$ for each $i\in I$. The expected total h-score for $\mathbf{p}$ is then

$$\sum_{i\in I^C} r_{1,i}h_{1,A_i,B_i}(p_i)+r_{2,i}h_{0,A_i,B_i}(p_i)$$
$$=\sum_{i\in I^C} r_{1,i}g_{1,A_i,B_i}(p_i)+r_{2,i}g_{0,A_i,B_i}(p_i)$$
$$=\sum_{i\in I^C} r_{1,i}g_{1,A_i,B_i}(r_i)+r_{2,i}g_{0,A_i,B_i}(r_i),\qquad(24)$$

where the last equality follows from the fact that r_i for $i=1,\ldots,n$ also minimize the expected g-scores. Because $\mathbf{p}$ is not weakly Bayes under the g-scores, it must be that

$$\sum_{i=1}^{n} r_{1,i}g_{1,A_i,B_i}(p_i)+r_{2,i}g_{0,A_i,B_i}(p_i)>\sum_{i=1}^{n} r_{1,i}g_{1,A_i,B_i}(r_i)+r_{2,i}g_{0,A_i,B_i}(r_i).$$

Because the $p_i = r_i$ for $i \in I$ are extreme, they contribute 0 to the total expected g-score. Hence,

$$\sum_{i \in I^C} r_{1,i} g_{1, A_i, B_i}(p_i) + r_{2,i} g_{0, A_i, B_i}(p_i) > \sum_{i \in I^C} r_{1,i} g_{1, A_i, B_i}(r_i) + r_{2,i} g_{0, A_i, B_i}(r_i).$$

This contradicts (24), and hence **p** is not weakly Bayes under the h-scores.

Next, apply Lemma 12 to find a (possibly randomized) rule δ that dominates **p** under the h-scores. If δ is a nonran-domized rule $\mathbf{q} = (q_1, \ldots, q_n)$, represent it as a randomized rule with $\delta(\{\mathbf{q}\}) = 1$. The total h-scores from δ and the original forecasts are, respectively,

$$d'_j = \sum_{i=1}^{n} \int_{\mathscr{O}} b_i(j) h_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q}) \quad \text{and}$$

$$d''_j = \sum_{i=1}^{n} b_i(j) h_{a_i(j), A_i, B_i}(p_i) < \infty,$$

where $\mathscr{O}$ is the action space in the proof of Lemma 12. Let $w = \min_{j \in J^C}(d''_j - d'_j)$, the minimum amount by which δ dominates the original forecasts amongst those constituents where the original g-scores are finite. Table A.1 summarizes some of what we know about the g-scores of δ and the incoherent$_1$ forecasts. The reason that $E = 0$ is that all of the p_i for $i \in I$ are extreme and they contribute either 0 or ∞ to each total score. For $j \in J^C$, the total score is finite; hence, all p_i for $i \in I$ must contribute 0 to the total score. It follows that $E = 0$ and $F = d_j$. The reason $d_j = d''_j$ is that the g-scores and h-scores are the same for all $i \in I^C$. In addition, we know that $D < d''_j$, hence $D < \infty$. Also, $H \le F - w$. If we knew that $G < w$, then we would know that δ weakly dominates the incoherent$_1$ forecasts under the g-scores. If, in addition, we knew that $C < \infty$, then we would know that δ strictly dominates the incoherent$_1$ forecasts under the g-scores. Even if these two facts are not true, we notice that C and G depend only on the distribution (under δ) of the ith coordinates of **q** for $i \in I^C$. If we change the joint distribution of these coordinates without affecting the joint distributions of the other coordinates, none of the other numbers in Table A.1 is affected. We proceed now to replace δ by another randomized rule δ' to make $G < w$, and if Assumption 2 holds for all scoring rules, $C < \infty$.

Table A.1 Total g-Scores for the Incoherent$_1$ Forecasts and the Dominating Randomized Rule δ

	$i \in I$	$i \in I^C$
	Incoherent forecasts	
$j \in J$	$A = \infty$	$B = \sum_{i \in I} b_i(j) g_{a_i(j), A_i, B_i}(p_i)$
$j \in J^C$	$E = 0$	$F = d''_j = d_j$
	Dominating randomized rule δ	
$j \in J$	$C = \sum_{i \in I} \int_{\mathscr{A}} b_i(j) g_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$	$D = \sum_{i \in I^C} \int_{\mathscr{A}} b_i(j) h_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$
$j \in J^C$	$G = \sum_{i \in I} \int_{\mathscr{A}} b_i(j) g_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$	$H = \sum_{i \in I^C} \int_{\mathscr{A}} b_i(j) h_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$

Notes. The g-scores are expressed in terms of the h-scores when the two agree. The total scores are split into the contributions from $i \in I$ and from $i \in I^C$.

First, consider the case in which we do not assume that every scoring rule satisfies Assumption 2. Define δ' as follows. The joint distribution of $\{q_i: i \in I^C\}$ is the same as that under δ. Under δ', $q_i = p_i$ with probability 1 for all $i \in I$. With this change, we have $C = A = \infty$ and $G = E = 0$. Hence, Conclusion 1 holds because $G + H < E + F$, whereas $A + B = C + D$.

Next, assume that Assumption 2 holds for all scoring rules. Let $v = w/(2n)$. Under Assumption 2, each $g_{k, A_i, B_i}(x)$ is continuous at $x = k$, and $g_{k, A_i, B_i}(k) = 0$ for $k = 0, 1$. For each i and each $k = 0, 1$, let $t_{k,i} \notin \{0, 1\}$ be close enough to k so that $g_{k, A_i, B_i}(t_{k,i}) \le v$. Let δ' be defined as follows. The joint distribution of $\{q_i: i \in I^C\}$ is the same as that under δ. Under δ', $q_i = t_{p_i, i}$ with probability 1 for all $i \in I$. With this change, we have that $C < \infty$ because δ' gives 0 mass to extreme forecasts in the coordinates in I. Also, for $j \in J^C$, $G \le n \max_{i \in I} g_{a_i(j), A_i, B_i}(t_{p_i, i}) \le nv = w/2$. So, Conclusion 2 holds.

Finally, suppose that all of the g-scoring rules are continuous. Then all of the h-scoring rules are continuous, and all of the randomized rules above are nonrandomized. If the nonrandomized rule **q** is coherent, the proof is complete. If the forecasts in **q** are not coherent$_1$ but are weakly Bayes, then Theorem 3 says that there are coherent$_1$ forecasts that have the same total score in every constituent and hence strictly dominate **p**. If **q** are neither coherent$_1$ nor Bayes, at least they produce a finite score in every constituent, and they strictly dominate $p_1, \ldots, p_n$. Now, apply Lemma 12 to **q** producing another nonrandomized rule that is coherent$_1$ and strictly dominates **q**, and hence strictly dominates **p**, so that Conclusion 3 holds. □

Proof of Corollary 1. If the forecasts are coherent$_3$, no randomized forecast dominates the forecasts. The contrapositive of the second part of Theorem 2 says that the forecasts are weakly Bayes.

If the forecasts are weakly Bayes, Lemma 1 says that the forecasts are coherent$_3$. □

Proof of Corollary 2. If the forecasts are coherent$_3$, no nonrandomized forecast dominates the forecasts. The contrapositive of the third part of Theorem 2 says that the forecasts are weakly Bayes.

If the forecasts are weakly Bayes, Lemma 1 says that the forecasts are coherent$_3$. □

Theorem 1

Proof of Theorem 1. Assume that we have conditional forecasts for the pairs of events in a collection $\mathscr{C}$ and that the forecast for each pair $(A.B) \in \mathscr{C}$ is scored by a strictly

proper scoring rule $(g_{0,A,B}, g_{1,A,B})$ that satisfies Assumptions 1 and 2. Every strictly proper scoring rule satisfies Assumption 3.

For the first claim in Theorem 1, apply Corollary 1 to conclude that the forecasts are coherent$_3$ if and only if they are weakly Bayes. For the second claim, apply Corollary 2 to conclude that the forecasts are coherent$_3$ if and only if they are weakly Bayes.

For either claim, apply Lemma 2 to conclude that forecasts are weakly Bayes if and only if they are coherent$_1$. □

References

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Rev.* **78** 1–3.

Cozman, F., T. Seidenfeld. 2009. Independence of full conditional measures, graphoids and independence. B. Löwe, E. Pacuit, J.-W. Romeijn, eds. *Foundations of the Formal Sciences VI: Reasoning About Probabilities and Probabilistic Reasoning Studies in Logic.* College Publications, London.

de Finetti, B. 1972. *Probability, Induction and Statistics.* John Wiley & Sons, New York.

de Finetti, B. 1974. *The Theory of Probability.* John Wiley & Sons, New York.

de Finetti, B. 1981. The role of "Dutch books" and "proper scoring rules." *British J. Philos. Sci.* **32** 55–56.

Dubins, L. 1975. Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probability* **3** 89–99.

Gneiting, T., A. E. Raftery. 2007. Strictly proper scoring rules, prediction and estimation. *J. Amer. Statist. Assoc.* **102** 359–378.

Kadane, J. B., J. M. Dickey, R. L. Winkler, W. Smith, S. C. Peters. 1980. Interactive elicitation of opinion for a normal linear model. *J. Amer. Statist. Assoc.* **75** 845–854.

Krauss, P. H. 1968. Representation of conditional probability measures on Boolean algebras. *Acta Math. Sci. Hungar.* 229–241.

Murphy, A. H. 1972. Scalar and vector partitions of the probability score: Part I. Two-state situation. *J. Appl. Meteorology* **11** 273–282.

Murphy, A. H., E. S. Epstein. 1967. A note on probability forecasts and "hedging." *J. Appl. Meteorology* **6** 1002–1004.

Pearce, D. 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* **52** 1029–1050.

Predd, J., R. Seiringer, E. H. Lieb, D. Osherson, V. Poor, S. Kulkarni. 2009. Probabilistic coherence and proper scoring rules. *IEEE Trans. Inform. Theory* **55**(10) 4786–4792.

Savage, L. J. 1971. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

Schervish, M. J. 1989. A general method for comparing probability assessors. *Ann. Statist.* **17** 1856–1879.

Schervish, M. J. 1995. *Theory of Statistics.* Springer-Verlag, New York.

Wald, A. 1950. *Statistical Decision Functions.* Wiley, New York.