

7-2008

# Proper Scoring Rules, Dominated Forecasts, and Coherence

Mark J. Schervish

*Carnegie Mellon University*, [mark@stat.cmu.edu](mailto:mark@stat.cmu.edu)


Teddy Seidenfeld

*Carnegie Mellon University*, [teddy@stat.cmu.edu](mailto:teddy@stat.cmu.edu)

Joseph B. Kadane

*Carnegie Mellon University*, [kadane@stat.cmu.edu](mailto:kadane@stat.cmu.edu)

Follow this and additional works at: <http://repository.cmu.edu/statistics>

 Part of the [Philosophy Commons](#), and the [Statistics and Probability Commons](#)

---

## Published In

Decision Analysis, 6, 4, 202-221.

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Statistics by an authorized administrator of Research Showcase @ CMU. For more information, please contact [research-showcase@andrew.cmu.edu](mailto:research-showcase@andrew.cmu.edu).

## PROPER SCORING RULES, DOMINATED FORECASTS, AND COHERENCE

BY MARK J. SCHERVISH, TEDDY SEIDENFELD, AND JOSEPH B. KADANE

*Carnegie Mellon University*

De Finetti introduced the concept of coherent previsions and conditional previsions through a gambling argument and through a parallel argument based on a quadratic scoring rule. He shows that the two arguments lead to the same concept of coherence. When dealing with events only, there is a rich class of scoring rules which 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 prevision to deal with conditional previsions.

**1. Introduction.** Chapter 3 of *Theory of Probability* [3, vol. 1] provides two criteria of coherence for previsions (or probabilistic forecasts) over a set of events. The first criterion is formulated under the assumption that previsions serve as fair prices for buying and/or selling contracts of the form  $c(I_A - P(A))$ . Here,  $I_A$  is the indicator of an event  $A$  and  $P(A)$  is the prevision of the event. Each event is a subset of the set  $\Omega$  whose generic element is denoted  $\omega$ . With  $c$  positive, the decision maker will pay the price  $cP(A)$  and receive  $c$  units if  $\omega \in A$  and 0 if  $\omega \notin A$ . With  $c$  negative, the decision maker sells this gamble on  $A$  and receives a payment of  $cP(A)$  units. With  $c = 0$ , the decision maker remains at his status-quo fortune. Furthermore, conditional previsions are also defined as fair prices for buying and/or selling contracts of the form  $cI_B[I_A - P(A|B)]$ . Here,  $B$  is the conditioning event. The contract has 0 value if  $B$  does not occur. It is easy to see that, if  $B = \Omega$ ,  $P(A|B)$  has the same operational meaning as  $P(A)$ . In this way, previsions are special cases of conditional previsions. We refer to conditional previsions for which  $B = \Omega$  as *marginal previsions* when the need arises. When we refer to previsions unqualified, we mean both marginal and conditional previsions.

In order to be able to deal with marginal and conditional previsions simultaneously, we introduce some notation. Let  $\aleph$  be an index set, and let  $\mathcal{X} = \{(A_\alpha, B_\alpha) : \alpha \in \aleph\}$  be a set of pairs of events. For each  $\alpha \in \aleph$ ,  $A_\alpha$  is

---

*AMS 2000 subject classifications:* Primary 62A01; secondary 62C05

*Keywords and phrases:* coherence, finite additivity, forecasts

a subset of  $\Omega$ , and  $B_\alpha$  is a non-empty subset of  $\Omega$ . For each  $(A, B) \in \mathcal{X}$  the decision maker is required to provide a real-valued (conditional) prevision  $P(A|B)$ .

DEFINITION 1 (Coherence<sub>1</sub>). *A set of previsions is incoherent<sub>1</sub> if there exists a finite subset  $\{\alpha_1, \dots, \alpha_k\}$  of  $\mathbb{N}$  and corresponding values  $\{c_{\alpha_1}, \dots, c_{\alpha_k}\}$  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  $\omega \in \Omega$ ,*

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

*The set of previsions is called coherent<sub>1</sub> otherwise. If (1) occurs, we say that book has been made against the decision maker.*

Thus, coherence<sub>1</sub> is the requirement that the decision maker's previsions cannot be (uniformly) dominated by the status-quo, corresponding to the state of neither buying nor selling such gambles.

De Finetti had some misgivings about one aspect of this criterion of coherence. As he explained in [4], he was concerned that, even as an idealization, it failed to satisfy operational requirements for eliciting the decision maker's uncertainties over the pairs in  $\mathcal{X}$ . The problem he recognized is that the decision maker's selection of announced previsions might reflect her or his anticipation of the opinions of the "opponent," whose role in the game is to select the sign and magnitude of the coefficients, the  $c_s$ , in the gambles. In this regard, he disliked the strategic aspects that entered into the prevision-game under the first criterion for coherence.

De Finetti proposed a second criterion of coherence, which he demonstrated is equivalent to the first for classifying sets of previsions as coherent and, since it does not appeal to decisions of an opponent, avoids the concerns with strategic play that affect the first criterion. The second criterion uses a squared-error loss function to score previsions:  $L = I_B(I_A - P(A))^2$ . That is, the decision maker suffers a loss  $L$ , depending upon the events  $A$  and  $B$ . The loss given to a finite set of previsions  $\{P(A_{\alpha_1}|B_{\alpha_1}), \dots, P(A_{\alpha_k}|B_{\alpha_k})\}$  is the sum of the losses for the individual previsions.

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

Thus, coherence<sub>2</sub> is the requirement that no finite subset of the decision maker's previsions can be (uniformly) dominated by a rival set of previsions in terms of squared-error loss.

[3] established that a set of marginal previsions is coherent in the one sense if and only if it is coherent in the other sense. In [3, pp. 88-89 and 188-190] he gave a geometric argument to show that the second criterion distinguishes coherent from incoherent sets of marginal previsions of events identically with the first criterion. His elegant reasoning is simple to state. When an incoherent<sub>1</sub> set of marginal previsions are announced for a finite set of events, collectively those correspond to a point outside the simplex generated by the linear span of the indicator functions for the finitely many events forecasted. Then, projecting onto the simplex from that point produces a coherent<sub>1</sub> set of marginal previsions whose squared-error loss dominates that of the first set of previsions, regardless which state occurs. That is, the Euclidean distances from the projected point on the face of the simplex to each corner of the simplex is strictly less than the corresponding distances from the point outside the simplex to those corners; hence, those incoherent<sub>1</sub> previsions for a finite set of events are also incoherent<sub>2</sub>. Moreover, if the set of previsions are coherent<sub>1</sub> they correspond to a point within the simplex. But no other point within the simplex is simultaneously closer to all the corners. Hence, a coherent<sub>1</sub> set of previsions for events also is coherent<sub>2</sub>.

Regarding the problem of strategic play in eliciting a decision maker's previsions, which affects the first criterion, that concern is mitigated using the second criterion since, under squared-error (Brier) score, the decision maker uniquely minimizes her/his expected loss for a set of previsions by announcing the (subjective) expected value for each variable forecast. When forecast variables are indicator functions for a set of events, the decision maker's previsions then are her/his personal probabilities for those events. That is, as [4] reported he "invented" the rule, known since [1] as Brier score, which is a strictly proper scoring rule for eliciting probabilities for events.

**DEFINITION 3 (Scoring Rules).** *A scoring rule for scoring the conditional forecast  $P(A|B)$  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  $I_B g_1(P(A|B))$ , and if  $A^C$  occurs, the forecaster suffers a loss of  $I_B g_0(P(A|B))$ . 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  $B$ . If  $P(B) = 0$ , there are some issues that arise, but most of them do not affect the results that we present in this paper. We return to these issues in Section 6. To avoid the one issue that does affect our results, we assume that  $0 \leq P(A|B) \leq 1$  for every pair  $(A, B)$  of events with  $B \neq \emptyset$ . We make this assumption because many proper scoring rules are not defined for previsions outside of the unit interval. Although it is part of our goal to deal with incoherent previsions, we cannot allow them to be quite so incoherent if we wish to apply proper scoring rules.

[9] generalized de Finetti's use of squared error loss in Definition 2 and characterized a general class of (strictly) proper scoring rules for eliciting personal probabilities for events. However, Savage 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 previsions. [7] established that, for the case of finitely many marginal previsions, de Finetti's geometric argument extends to all continuous strictly proper scoring rules by generalizing the role played by the Euclidean metric with Brier score to Bregman divergence for continuous scoring rules. That is, with each continuous, strictly proper scoring rule if a finite set of marginal previsions is incoherent<sub>1</sub> then it is dominated in score by some coherent<sub>1</sub> set of forecasts. And no finite coherent<sub>1</sub> set of marginal previsions can be so dominated.

**DEFINITION 4 (Dominance).** *Let  $(A_1, B_1), \dots, (A_n, B_n)$  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  $i = 1, \dots, n$ . 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  $\mathbf{p} = (p_1, \dots, p_n)$  and  $\mathbf{q} = (q_1, \dots, q_n)$  be two different sets of conditional forecasts for the  $n$  events. We say that  $\mathbf{q}$  weakly dominates  $\mathbf{p}$  if, for all  $\omega \in \Omega$ ,*

$$(2) \quad \sum_{i=1}^n g'_{A_i,B_i}(q_i, \omega) \leq \sum_{i=1}^n g'_{A_i,B_i}(p_i, \omega),$$

*with strict inequality for at least one  $\omega$ . We say that  $\mathbf{q}$  strictly dominates  $\mathbf{p}$  if (2) holds for all  $\omega$  with strict inequality for all  $\omega$ .*

There is a second, decision-theoretic line of argument relating to the equivalence between the two senses of coherence, however, that was anticipated by

[2, p. 181-182] and sketched without much detail by [9, sections 8 and 9.1]. Consider a decision problem comprised by a set of options  $\mathcal{O}$  and subject to a loss function (bounded below) defined with respect to the finite partition  $\Theta$ . If an option  $O^*$  fails to be a Bayes solution to the problem, i.e., if for each probability on  $\Theta$ ,  $O^*$  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  $O^*$ . When  $\mathcal{O}$  is finite, this result was established by [8, Lemma 3, p. 1049]. Here we extend this reasoning to statistical problems where  $\Theta$  is finite, i.e., “Nature” has only finitely many nonrandomized options, but where 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 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<sub>1</sub> set of forecasts that strictly dominates each finite incoherent<sub>1</sub> set of forecasts, just as with Brier score. However, we illustrate that when the scoring rules are discontinuous (and even though strictly proper) there may fail to be a coherent<sub>1</sub> set of forecasts that weakly dominate a particular incoherent<sub>1</sub> set of forecasts. In other words, although there is equivalence between the two senses of coherence even with discontinuous strictly proper scoring rules, it is no longer the case that incoherence<sub>2</sub> is characterized by the existence of a dominating coherent<sub>2</sub> set of forecasts.

**2. Summary of Results.** The concept of dominance is similar to that of inadmissibility, and scoring rules are just like loss functions. For these reasons, we make use of results from statistical decision theory to prove our results about when a given set of forecasts can be dominated. For each finite collection  $\mathcal{A} = \{(A_1, B_1), \dots, (A_k, B_k)\}$  of pairs of subsets of  $\Omega$  with  $B_i \neq \emptyset$ , 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.

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 1 in Section 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 vs. strictly proper, continuity, dominance, 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  $0 < x < 1$ .*

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<sub>3</sub>). *A collection of conditional forecasts for events is incoherent<sub>3</sub> 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<sub>3</sub> otherwise.*

The question then arises as to whether or not coherence<sub>3</sub> is equivalent to coherence<sub>1</sub> 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<sub>1</sub> and coherence<sub>3</sub> 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<sub>1</sub> and coherence<sub>3</sub> are equivalent in every forecasting problem.*
- *If only nonrandomized forecasts are allowed and if all scoring rules are also continuous, then coherence<sub>1</sub> and coherence<sub>3</sub> are equivalent in every forecasting problem.*

The proof of Theorem 1 appears in Section 5. We provide an example (Example 4) 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 3) to illustrate that incoherent<sub>1</sub> 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<sub>3</sub> set of forecasts gets scored identically to a coherent<sub>1</sub> 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  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\}$  of pairs of subsets of a set  $\Omega$  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  $A_1, \dots, A_n, B_1, \dots, B_n$  be events. Construct the (at most  $3^n$ ) events  $C_j = E_{1,j} \cap \dots \cap E_{n,j}$  where each  $E_{i,j} \in \{A_i \cap B_i, A_i^C \cap B_i, B_i^C\}$  for  $i = 1, \dots, n$ . 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  $C_1, \dots, C_m$  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  $i$ th 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  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\}$  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  $\mathbf{p} = (p_1, \dots, p_n)$  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

$$(3) \quad d_j = \sum_{i=1}^n b_i(j) g_{a_i(j), A_i, B_i}(p_i) = I_{C_j}(\omega) \sum_{i=1}^n g'_{A_i, B_i}(p_i, \omega),$$

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

$$(4) \quad d_j = \int_{[0,1]^n} \sum_{i=1}^n b_i(j) g_{a_i(j), A_i, B_i}(p_i) \delta(d\mathbf{p}).$$

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  $\delta$  by letting  $\delta(\{\mathbf{p}\}) = 1$ . In this case, (3) and (4) are the same.

Definition 9 summarizes the above construction of a decision problem.

DEFINITION 9 (Problem  $\mathcal{A}$ ). Let  $\Omega$  be a space. For each pair  $(A, B)$  of subsets of  $\Omega$  with  $B \neq \emptyset$ , let  $(g_{0,A,B}, g_{1,A,B})$  be a proper scoring rule. For each finite collection  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\}$  of pairs of subsets of  $\Omega$  with all  $B_i$  nonempty, define the following decision problem (called problem  $\mathcal{A}$ ). The parameter space is  $\Theta = \{C_1, \dots, C_m\}$  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  $A_1, \dots, A_n$  given  $B_1, \dots, B_n$  respectively, and the loss function is the total score. A randomized forecast  $\delta$  is Bayes in problem  $\mathcal{A}$  if there exists a probability distribution  $\mathbf{q} = (q_1, \dots, q_m)$  over  $\Theta$  such that

$\delta$  minimizes the expected loss, i.e.,

$$(5) \quad \sum_{j=1}^m q_j d_j = \inf_{\delta^*} \sum_{j=1}^m q_j \int_{[0,1]^n} \sum_{i=1}^n b_i(j) g_{a_i(j), A_i, B_i}(p_i) \delta^*(d\mathbf{p}),$$

where the inf is over all randomized forecasts  $\delta^*$ .

Some simplification of the expression in (5) is possible. First, use the standard notation for the loss function of a randomized rule to denote, for each randomized rule  $\delta$ ,

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

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

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

Then (5) becomes

$$(6) \quad \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1, A_i, B_i}(\delta) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(\delta) \right\} \\ = \inf_{\delta^*} \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1, A_i, B_i}(\delta^*) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(\delta^*) \right\},$$

a more familiar formula indicating that  $\delta$  minimizes the expected total score. If  $\delta$  is the nonrandomized forecast  $\mathbf{p}$ , then (6) becomes

$$(7) \quad \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1, A_i, B_i}(p_i) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(p_i) \right\} \\ = \inf_{q_1, \dots, q_n} \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1, A_i, B_i}(q_i) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(q_i) \right\}.$$

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 previsions is classified as coherent 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  $\delta$  is weakly Bayes if, for every finite subcollection  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$ ,  $\delta$  is Bayes in problem  $\mathcal{A}$ . A weakly Bayes forecast  $\delta$  is strongly Bayes if there exists a finitely additive probability  $R$  on  $(\Omega, 2^\Omega)$  that satisfies*

$$(8) \quad \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1,A_i,B_i}(\delta) + R(A_i^C \cap B_i) g_{0,A_i,B_i}(\delta) \right\} \\ = \sum_{i=1}^n \left\{ R(A_i \cap B_i) g_{1,A_i,B_i}(R(A_i|B_i)) + R(A_i^C \cap B_i) g_{0,A_i,B_i}(R(A_i|B_i)) \right\},$$

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

Weakly Bayes forecasts turn out to be the ones that are coherent<sub>3</sub>,

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 Section 5.

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<sub>3</sub>.*

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<sub>3</sub>.*

TABLE 1

*Summary of assumptions and conclusions of results providing dominating or equivalent forecasts. 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.*

Conclusions	Assumptions	Examples to Justify Assumptions
A possibly randomized forecast weakly dominates.	All scoring rules satisfy Assumption 1 and all merely proper scoring rules satisfy Assumptions 2 and 3.	Example 1 and Lemma 3.
A possibly randomized forecast strictly dominates.	All scoring rules satisfy Assumptions 1-3.	Example 2.
A coherent forecast strictly dominates	All scoring rules are continuous and satisfy Assumptions 1 and 3.	Example 4.
A coherent forecast with the same total scores.	All scoring rules are continuous and satisfy Assumption 1.	Example 5.

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

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

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 forecasts with the same total scores when the initial set of forecasts is weakly Bayes. The various conclusions of these theorems are listed here:

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

3. There exists a finite subcollection  $\{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$  whose forecasts  $P(A_1|B_1), \dots, P(A_n|B_n)$  are not weakly Bayes, and for every such subcollection there exists a coherent forecast that strictly dominates  $P(A_1|B_1), \dots, P(A_n|B_n)$ .
4. There exists a coherent collection of forecasts that has the same total scores as  $\{P(A|B) : (A, B) \in \mathcal{C}\}$  for every finite subcollection  $\{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$ .

We refer to each of these as Conclusion  $i$  for  $i = 1, 2, 3, 4$  in the remainder of the paper.

**THEOREM 2.** *Suppose that an agent must provide a conditional forecast for each pair of events in the collection  $\mathcal{C}$ . 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  $\mathcal{C}$ . 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 Theorem 2 and 3 appear in Section 5.

**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 previsions. None of the assumptions that we make is needed solely because we allow both conditional and marginal previsions. That is, even if we were to restrict attention solely to marginal previsions, as [7] do, our proofs would still use all of the assumptions in order to deal with the examples in this section. To simplify notation, we do not write all of these marginal previsions as conditional on  $\Omega$ . Instead, we leave off the “ $|\Omega$ ” from each prevision in these examples. Similarly, the scoring rule for scoring the prevision 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  $\Omega$ .

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 1 illustrates that incoherent<sub>1</sub> 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 1. Let  $\mathcal{C} = \{A_1, A_2\}$  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

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

The first is Brier score, while 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

$$(9) \quad (1-p)\log(x) + p \left[ \log(x) + \frac{1}{x} \right] = \log(x) + \frac{p}{x}.$$

The expression in (9) 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 (9) is also minimized uniquely at  $x = 0$ .

Now, suppose that an incoherent<sub>1</sub> 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<sub>1</sub> 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<sub>1</sub> forecast. So the incoherent<sub>1</sub> forecasts cannot be weakly dominated by another forecast, coherent<sub>1</sub> or otherwise.

The forecasts in this example are Bayes in problem  $\mathcal{C}$  with respect to a prior  $R$  if and only if  $R(C_3) = 1$ . For the subcollection  $\mathcal{A} = \{A_1\}$ , the forecasts are Bayes in problem  $\mathcal{A}$  with respect to  $R$  if and only if  $R(A_1) = 1$ . No prior gives probability 1 to both  $A_1$  and  $C_3$  as 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 2 illustrates why we distinguish between Conclusions 1 and 2 depending on whether Assumption 2 holds for all scoring rules.

EXAMPLE 2. Let  $\mathcal{C} = \{A_1, A_2\}$  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<sub>1</sub> 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  $\infty$ . So the incoherent forecasts cannot be strictly dominated. The forecasts  $P'(A_1) = 1$  and  $P'(A_2) = 0$  weakly dominate.

Example 3 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 3. Let  $\mathcal{C} = \{A_0, A_1, A_2, \dots\}$  where each event is nonempty,  $A_1 \supset A_2 \supset \dots$  and  $A_0 \subseteq A_i$  for all  $i > 0$ . Suppose that an agent assigns

previsions 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 Brier score for all  $i > 0$  but

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

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

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  $0 < \Pr(A_0) \leq 1/2$ , then the expected score is  $\infty$  if  $x = 0$ ,  $1/4$  if  $0 < x \leq 1/2$ , and strictly greater than  $1/4$  otherwise. If  $1/2 < \Pr(A_0) = p \leq 1$ , then the expected score is  $\infty$  if  $x = 0$ ,  $1/4$  if  $0 < x \leq 1/2$ , and  $(1-p)x^2 + p(1-x)^2$  if  $1/2 < x \leq 1$ . 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  $\mathcal{A}$  of  $\mathcal{C}$  that does not include  $A_0$ , the corresponding forecasts are Bayes in problem  $\mathcal{A}$  because they are coherent<sub>1</sub>. For each finite subcollection  $\mathcal{A}$  that includes  $A_0$ , the corresponding forecasts are still Bayes in problem  $\mathcal{A}$ , 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  $A_i \in \mathcal{A}$ . Hence, the forecasts might as well be coherent<sub>1</sub> 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 (8) true for all of the finite subcollections that do not include  $A_0$  has  $R(A_0) = 0$ . But (8) 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 scoring rules that satisfy Assumption 1 but violate the 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  $0 \leq x \leq 1$ ,  $\max\{g_0(x), g_1(x)\} = \infty$ .

Such a scoring rule is not strictly proper.



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

$$(10) \quad (1 - p)g_0(x) + pg_1(x).$$

It is now clear that the first condition is necessary and sufficient for (10) to be minimized at  $x = p$  if  $p \in \{0, 1\}$ . It is also clear that the second condition is sufficient for (10) 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 (10) can be minimized at  $x = p$  when  $p = x_0$  is for (10) to be infinite for all  $0 \leq x \leq 1$ . That is, the second condition must hold. Since the expected score is minimized at all  $x$  when  $0 < p < 1$ , such scoring rules are not strictly proper.  $\square$

Notice that the scoring rules in Lemma 3 all have 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.* Discontinuous scoring rules tend to have risk sets that do not contain all of the lower boundary. Hence, the admissible rules will not always form a complete class. That is, there may be inadmissible rules that are dominated only by other inadmissible rules but not by admissible rules. Example 4 illustrates why we cannot expect a dominating coherent<sub>1</sub> set of forecasts with discontinuous scoring rules even though there are other incoherent<sub>1</sub> dominating strategies.

EXAMPLE 4. *Consider the scoring rule*

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

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

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

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

If  $p \leq 1/2$ , 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<sub>1</sub> 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<sub>1</sub> if and only if  $r_1 + r_2 = 1$ . For each coherent<sub>1</sub> 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<sub>1</sub> 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<sub>1</sub> forecast  $r_1 = r_2 = 0.5$ , the scores on both  $C_1$  and  $C_2$  are equal to 1.0. Hence, no coherent<sub>1</sub> forecast can weakly dominate the incoherent<sub>1</sub> forecast  $p_1 = 0.6$  and  $p_2 = 0.7$ . On the other hand, there are other incoherent<sub>1</sub> forecasts that dominate  $(p_1, p_2)$ . For example  $p'_1 = 0.55$  and  $p'_2 = 0.65$  has total scores of  $d'_1 = 1.125$  and  $d'_2 = 0.925$ .

Example 5 illustrates why Theorem 3 doesn't deal with the case in which an incoherent<sub>1</sub> set of forecasts is Bayes but the scoring rules are not continuous.

EXAMPLE 5. Consider the scoring rule

$$\begin{aligned} g_0(x) &= \begin{cases} x^2 & \text{if } x \leq 1/2, \\ 1 & \text{if } x > 1/2, \end{cases} \\ g_1(x) &= \begin{cases} 1/2 + (1 - x)^2 & \text{if } x \leq 1/2, \\ 0 & \text{if } x > 1/2. \end{cases} \end{aligned}$$

Let  $0 \leq p \leq 1$ . The expected score for forecasting  $x$  is

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

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.

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 forecasts be  $p_1 = 0.5$  and  $p_2 = 0.7$ . Then  $d_1 = 1/4$  and  $d_2 = 7/4$ . If the probability of  $A_1$  is  $1/2$ , the Bayes rule

assigns both events probability  $1/2$  and the total scores are both equal to 1. The expected total score is 1, which is the same as the total expected score of the incoherent<sub>1</sub> forecasts. Hence, the incoherent<sub>1</sub> forecasts are Bayes with all probabilities strictly positive and hence cannot be dominated. In order for forecasts  $(r_1, r_2)$  to have total score equal to  $1/4$  on  $C_1$  one needs  $r_1 = 1/2$  and  $r_2 > 1/2$ , which is not coherent<sub>1</sub>.

Example 6 illustrates how strongly Bayes forecasts can be weakly dominated if some merely proper scoring rules are discontinuous.

EXAMPLE 6. Consider the collection  $\mathcal{C} = \{A_1, A_2, A_3\}$  with  $A_3 = (A_1 \cup A_2)^C$  and  $A_1 \cap A_2 \neq \emptyset$ . 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 both the following:

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

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

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

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

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 7 illustrates why Theorem 2 has the weak Conclusion 3 rather than the stronger claim that the same set of coherent<sub>1</sub> forecasts dominates every incoherent<sub>1</sub> subcollection.

EXAMPLE 7. Let  $\mathcal{C} = \{A_1, A_2, A_3, A_4, A_5\}$  where  $A_1, A_2, A_3$  form a partition of  $\Omega$  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 Brier score. Consider the following incoherent<sub>1</sub> 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  $d_2 = 0.01$ . Every set of forecasts that dominates these must assign probability less than 0.1 to  $A_4$ . The subcollection  $\{A_1, A_5\}$  is also incoherent<sub>1</sub>. The constituents now are  $B'_1 = A_1$  and  $B'_2 = A_5$  with total scores of  $d'_1 = 0.01$  and  $d'_2 = 1.81$ . Every set of forecasts that dominates these must assign probability less than 0.1 to  $A_5$ . If a single set of forecasts were to dominate both of the two finite subcollections above, it would have to assign probability less than 0.1 to each of  $A_4$  and  $A_5$ . But  $A_4 \cup A_5 = \Omega$ , hence no coherent<sub>1</sub> set of forecasts can dominate both of the incoherent<sub>1</sub> subcollections above.

## 5. Proofs of Results.

5.1. *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 [10].

THEOREM 4 ([10, Theorem 4.2]). *Let  $(g_0, g_1)$  be a left-continuous scoring rule that satisfies Assumptions 1–3 and such that  $g_1(x)$  doesn't jump to  $\infty$  at  $x = 0$ . The scoring rule is proper if and only if there exists a  $\sigma$ -finite measure  $\lambda$  on  $[0, 1)$  such that for all  $x$*

$$(11) \quad g_1(x) = \int_{[x,1)} (1-q)\lambda(dq), \quad \text{and} \quad g_0(x) = \int_{[0,x)} q\lambda(dq).$$

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

LEMMA 4 ([10, Lemma A.2]). *Let  $(g_0, g_1)$  be a proper scoring rule. Let  $0 \leq p \leq 1$ , 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  $0 \leq p \leq 1$ , 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 (11), we have

$$(12) \quad pg_1(p') + (1-p)g_0(p') - [pg_1(p) + (1-p)g_0(p)] = \int_{[p,p')} (p' - q)\lambda(dq).$$

Since  $p' - q > 0$  for  $q \in (p, p')$ , the fact that the left-hand side of (12) 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  $\lambda$  assigns 0 mass. In the continuous case, all such intervals are closed.  $\square$

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  $\ell_p(x) = ph_1(x) + (1-p)h_0(x)$ . 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  $g_0, g_1, h_0, h_1$  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,  $\ell_p(p) = m_p(p)$  for all  $0 < p < 1$ . For  $p \in \{0, 1\}$ , we also have  $\ell_p(p) = m_p(p)$ , because  $g_i = h_i$  at both endpoints for  $i = 0, 1$ . For  $x \neq p$  and  $x \in \{0, 1\}$ , we have

$$(13) \quad ph_1(p) + (1-p)h_0(p) = pg_1(p) + (1-p)g_0(p) \leq pg_1(x) + (1-p)g_0(x).$$

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

$$(14) \quad ph_1(p) + (1-p)h_0(p) \leq \lim_{y \uparrow x} [pg_1(y) + (1-p)g_0(y)] = ph_1(x) + (1-p)h_0(x).$$

Together (13) and (14) imply that  $(h_0, h_1)$  is proper. If  $(g_0, g_1)$  is strictly proper, then the inequality is strict in (13). Assume by way of contradiction that the inequality is equality in (14). 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.  $\square$

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  $\sigma$ -finite measure  $\lambda$  on  $[0, 1)$  such that for all continuity points  $x \in (0, 1)$*

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

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  $\infty$  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  $\infty$  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.  $\square$

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*

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

PROOF. We prove the limit at 0, as the limit at 1 is similar. Since  $g_0(p)$  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  $\lim_{p \rightarrow 0} pg_1(p)$  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  $t \in (0, 1)$ ,  $\lim_{p \rightarrow 0} p \int_{(p,t)} (1-q)\lambda(dq) = c$ . 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  $n$ ,

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

a contradiction.  $\square$

5.2. *Equivalence of Definitions of Coherence.* In this section, we prove Lemma 1 and Lemma 2.

LEMMA 1. Let  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\}$  be a finite subcollection, and let  $p_1, \dots, p_n$  be the conditional forecasts. The left-hand side of (7) and the left-hand side of (8) are the expected total score under the probability  $R$ . If the forecasts in this subcollection were strictly dominated, then the dominating forecasts would have strictly smaller score for every  $\omega$ . Since 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 (6) and they could not satisfy (8) and hence they would not be weakly Bayes.  $\square$

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

LEMMA 9. *Suppose that (8) holds and that the right-hand side of (8) is finite. Then for each  $i = 1, \dots, n$ ,*

$$(15) \quad \begin{aligned} & R(A_i \cap B_i) g_{1, A_i, B_i}(\delta) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(\delta) \\ &= R(A_i \cap B_i) g_{1, A_i, B_i}(R(A_i|B_i)) + R(A_i^C \cap B_i) g_{0, A_i, B_i}(R(A_i|B_i)). \end{aligned}$$

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  $i$

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

If the inequality above were strict for some  $i$  with  $R(B_i) > 0$ , then the left-hand side of (8) would be strictly larger than the right-hand side of (8).  $\square$

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  $\mathcal{C}$ , there exists a finitely additive probability  $R$  such that (8) holds. We show that there is an  $R$  that works for all finite subsets. Let  $\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\}$  be an arbitrary finite subcollection of  $\mathcal{C}$ . Let  $\mathcal{P}_{\mathcal{A}} = \{R : (8) \text{ holds}\}$ . We would like to show first that  $\mathcal{P}_{\mathcal{A}}$  is closed in the topology of pointwise convergence, which is also the product topology on the function space  $\mathcal{P} = [0, 1]^{2^\Omega}$  which includes all finitely additive probabilities on  $\Omega$ . All strictly proper scoring rules satisfy Assumption 3. Assumptions 1 and 3 guarantee that the right-hand side of (8) is always finite. Lemma 9 then says that for each  $R \in \mathcal{P}_{\mathcal{A}}$  and each  $i$ , (15) holds. Hence,  $\mathcal{P}_{\mathcal{A}} = \bigcap_{i=1}^n \mathcal{P}_{\{(A_i, B_i)\}}$ . Next, we write each  $\mathcal{P}_{\{(A_i, B_i)\}}$  as the inverse image of a closed set under a continuous function. For each  $A \in 2^\Omega$ , the coordinate projection function  $f_A : \mathcal{P} \rightarrow [0, 1]$ , defined by  $f_A(R) = R(A)$ , is continuous. For each  $i$ , define the function

$$\ell_i(e_1, e_2) = \begin{cases} e_1 g_{1, A_i, B_i} \left( \frac{e_1}{e_1 + e_2} \right) + e_2 g_{0, A_i, B_i} \left( \frac{e_1}{e_1 + e_2} \right) & \text{if } e_1 + e_2 > 0, \\ 0 & \text{otherwise,} \end{cases}$$

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

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

where

(16)

$$D_i = \{(e_1, e_2) : e_1 g_{1, A_i, B_i}(\delta) + e_2 g_{0, A_i, B_i}(\delta) = \ell_i(e_1, e_2), \text{ and } e_1 + e_2 \leq 1\}.$$

If  $D_i$  is closed, then so is  $\mathcal{P}_{\{(A_i, B_i)\}}$ . 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  $i$ th coordinate of the randomized forecast  $\delta$  must be nonrandomized since only nonrandomized rules can be weakly Bayes with strictly proper scoring rules. Let the  $\delta$  assign probability 1 to the  $i$ th coordinate being  $p$ . In this case  $D_i = \{(e_1, e_2) : e_1(1-p) = e_2 p\}$ , which is a closed set.

If the scoring rule is merely proper, we argue as follows. Since  $D_i$  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  $n = 0, 1, \dots$ . Then  $h$  is continuous at  $(e_{1,0}, e_{2,0})$  and  $r_n$  converges to  $r_0$ . We can write  $\ell_i(e_1, e_2) = (e_1 + e_2) \ell_i^*(h(e_1, e_2))$ , where  $\ell_i^*(r) = r g_{1, A_i, B_i}(r) + (1-r) g_{0, A_i, B_i}(r)$ .  $r_n = r_0$  for all  $n$  and  $(e_{1,0}, e_{2,0}) \in D_i$ . Lemma 8 establishes that each  $\ell_i^*$  is continuous on the closed interval  $[0, 1]$ . So,  $\ell_i$  is continuous at  $(e_{1,0}, e_{2,0})$  which is then



in  $D_i$ . Hence,  $D_i$  is closed and so is  $\mathcal{P}_{\{(A_i, B_i)\}}$ . It follows that each  $\mathcal{P}_{\mathcal{A}}$  is closed.

Finally, we show that the intersection of all  $\mathcal{P}_{\mathcal{A}}$  is nonempty. It is easy to see that if  $\mathcal{B}$  is a finite subcollection such that  $\mathcal{B} \subseteq \mathcal{A}$ , then  $\mathcal{P}_{\mathcal{A}} \subseteq \mathcal{P}_{\mathcal{B}}$ . It follows that the collection of all  $\mathcal{P}_{\mathcal{A}}$  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  $\mathcal{P}_{\mathcal{A}}$  is nonempty. That is, there is at least one finitely additive probability  $R$  such that (8) holds for all finite subcollections. Hence the forecasts are strongly Bayes.  $\square$

The connection between coherent<sub>1</sub> forecasts and strongly Bayes forecasts needed for Lemma 2 relies on the ability to extend a collection of coherent<sub>1</sub> previsions into a linear functional on a linear space. The next two theorems extend de Finetti's fundamental theorem of prevision [3, Section 3.10] to deal explicitly with conditional previsions. The fundamental theorem of prevision applies to more general random variables than indicators of events. In keeping with de Finetti's 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.

**THEOREM 5 (Fundamental Theorem of Prevision).** *Let  $\mathcal{C}$  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  $(X, B) \in \mathcal{C}$ , let  $P(X|B)$  be a conditional prevision. Assume that the conditional previsions are coherent<sub>1</sub>. Let  $(X, \Omega) \notin \mathcal{C}$ . Then there exists a closed interval  $[c, d]$  such that  $P(X|\Omega) = x$  is coherent<sub>1</sub> with all the other conditional previsions if and only if  $c \leq x \leq d$ .*

**PROOF.** Define the linear space

$$(17) \quad \mathcal{Y} = \left\{ \sum_{i=1}^n \alpha_i I_{B_i} [X_i - P(X_i|B_i)] + f : (X_i, B_i) \in \mathcal{C} \text{ and } \alpha_i \in \mathbb{R} \text{ for } i = 1, \dots, n \text{ and } f \in \mathbb{R} \right\}.$$

For each  $Y \in \mathcal{Y}$  expressed as in (17), let  $L(Y) = f$ . The coherence<sub>1</sub> 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  $\mathcal{Y}$ . Define the

following two sets

$$\begin{aligned}\underline{P} &= \{Y \in \mathcal{Y} : Y \leq X\}, \\ \overline{P} &= \{Y \in \mathcal{Y} : Y \geq X\}.\end{aligned}$$

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

$$(18) \quad \beta(X - x) + \sum_{i=1}^n \alpha_i I_{B_i} [X_i - P(X_i|B_i)] < -\epsilon.$$

If  $\beta > 0$ , then (18) implies

$$(19) \quad X < x - \frac{\epsilon}{\beta} - \sum_{i=1}^n \frac{\alpha_i}{\beta} I_{B_i} [X_i - P(X_i|B_i)].$$

The right side of (19) is an element of  $\overline{P}$ , hence  $d \leq x - \frac{\epsilon}{\beta}$  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  $P(X|\Omega) = x$  is coherent<sub>1</sub> with the other previsions. Suppose, to the contrary, that  $x < c$ . Let  $Y = \sum_{i=1}^n \alpha_i I_{B_i} [X_i - P(X_i|B_i)] + f \in \underline{P}$  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  $Y \in \overline{P}$  that allows us to make book.  $\square$

**THEOREM 6** (Fundamental Theorem of Conditional Prevision). *Let  $\mathcal{C}$  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  $(X, B) \in \mathcal{C}$ , let  $P(X|B)$  be a conditional prevision. Assume that the conditional previsions are coherent<sub>1</sub>. Let  $(X, D) \notin \mathcal{C}$  with  $D \neq \emptyset$ . Then there exists a set  $E$  of real numbers such that  $P(X|D) = x$  is coherent<sub>1</sub> with all the other conditional previsions if and only if  $x \in E$ .*

**PROOF.** First, suppose that both  $(D, \Omega)$  and  $(XI_D, \Omega)$  are in  $\mathcal{C}$ . de Finetti [3] proves that a necessary and sufficient condition for  $P(X|D)$  to be coherent<sub>1</sub> is  $P(XI_D|\Omega) = P(X|D)P(D|\Omega)$ . If  $P(D|\Omega) > 0$ , then  $E = \{P(XI_D|\Omega)/P(D|\Omega)\}$ . If  $P(D|\Omega) = 0$ , then  $E = \mathbb{R}$ . Next, suppose that  $(D, \Omega) \in \mathcal{C}$  but  $(XI_D, \Omega) \notin \mathcal{C}$ . If  $P(D|\Omega) = 0$ , then  $E = \mathbb{R}$ . If  $P(D|\Omega) \neq 0$ ,

apply Theorem 5 to find an interval  $[c, d]$  of possible coherent<sub>1</sub> values for  $P(XI_D|\Omega)$ . Then  $E = \{x/P(D|\Omega) : c \leq x \leq d\}$ . Next, assume that  $(D, \Omega) \notin \mathcal{C}$  but  $(XI_D, \Omega) \in \mathcal{C}$ . Apply Theorem 5 to find an interval  $[c, d]$  of possible coherent<sub>1</sub> values for  $P(D|\Omega)$ . If  $c = 0$ , then  $E = \mathbb{R}$ . If  $c > 0$ , then  $E = \{P(XI_D|\Omega)/x : c \leq x \leq d\}$ . Finally, assume that neither  $(D, \Omega)$  nor  $(XI_D, \Omega)$  is in  $\mathcal{C}$ . Apply Theorem 5 to find an interval  $[c_1, d_1]$  of possible coherent<sub>1</sub> values of  $P(D|\Omega)$ . For each  $x \in [c_1, d_1]$ , apply the argument above for the case in which  $(D, \Omega) \in \mathcal{C}$  but  $(XI_D, \Omega) \notin \mathcal{C}$  to find a set  $E_x$  of possible coherent<sub>1</sub> values of  $P(X|D)$ . Then  $E = \cup_{x \in [c_1, d_1]} E_x$ .  $\square$

**COROLLARY 3.** *Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  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  $(X, B) \in \mathcal{C}_1$ , let  $P(X|B)$  be a conditional prevision. Assume that the conditional previsions are coherent<sub>1</sub>. For each  $(X, B) \in \mathcal{C}_2$ , there exists a conditional prevision  $P(X|B)$  such that  $\{P(X|B) : (X, B) \in \mathcal{C}_1 \cup \mathcal{C}_2\}$  are coherent<sub>1</sub>.*

**PROOF.** Use Zermelo's lemma to well-order the elements of  $\mathcal{C}_2$ . 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  $P(X_\alpha|B_\alpha)$  that is coherent<sub>1</sub> with all earlier previsions. For each limit ordinal  $\beta \in D$ , it is easy to see that the previsions  $\{P(X_\alpha|B_\alpha) : \alpha < \beta\} \cup \{P(X|B) : (X, B) \in \mathcal{C}_1\}$  are coherent<sub>1</sub> because every finite subcollection was verified as coherent<sub>1</sub> at an earlier stage in the induction.  $\square$

**LEMMA 11.** *Let  $\mathcal{C}$  be a collection of pairs of events. A collection of conditional previsions  $\{P(A|B) : (A, B) \in \mathcal{C}\}$  is coherent<sub>1</sub> if and only if there exists a finitely additive probability  $R$  on  $(\Omega, 2^\Omega)$  that agrees with  $P$  on  $\mathcal{C}$  in the following sense: For each  $(A, B) \in \mathcal{C}$ ,  $R(B)P(A|B) = R(A \cap B)$ .*

**PROOF.** For the "if" part, assume that such an  $R$  exists. Let

$$\mathcal{A} = \{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$$

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

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

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

$$(20) \quad X = \sum_{i=1}^n \alpha_i I_{B_i} [I_{A_i} - P(A_i|B_i)] < -\epsilon.$$

Note that  $X$  in (20) is an element of  $\mathcal{L}$ . Because  $R$  agrees with the conditional previsions,  $X$  equals

$$(21) \quad \sum_{i=1}^n \alpha_i [I_{A_i \cap B_i} - R(A_i \cap B_i)] - \sum_{i=1}^n \alpha_i P(A_i|B_i) [I_{B_i} - R(B_i)].$$

It follows from (21) that  $L(X) = 0$ , but (20) implies that  $L(X) < -\epsilon$ , a contradiction. Hence, no book can be made and the conditional previsions are coherent<sub>1</sub>. For the “only if” part, assume that the conditional previsions are coherent<sub>1</sub>. Extend the collection of conditional previsions to include all pairs  $(B, \Omega)$  and  $(A \cap B, \Omega)$  for each  $(A, B) \in \mathcal{C}$  using Corollary 3. Let  $\mathcal{L}'$  be the linear span of all constants and indicators  $I_{A \cap B}$  and  $I_B$  for  $(A, B) \in \mathcal{C}$ . On  $\mathcal{L}'$ , define

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

Note that  $L'$  satisfies  $L'(X) \leq \|X\|_\infty$  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  $\Omega$ . This extension, when restricted to the indicators of events, is a finitely additive probability  $R$  that agrees with  $P$  on  $\mathcal{C}$ .  $\square$

LEMMA 2. Let  $\{P(A|B) : (A, B) \in \mathcal{C}\}$  be a collection of conditional forecasts. To prove the first claim, assume that the forecasts are coherent<sub>1</sub>. 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. Since 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  $\{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$  be a finite subcollection. The right-hand side of (8) is always finite. Lemma 9 says that (15) holds for all  $i = 1, \dots, n$ . Because the scoring rules are strictly proper,  $P(A_i|B_i) = R(A_i|B_i)$  for all  $i$  such that  $R(B_i) > 0$ . Since this is true for every finite subset,  $R$  agrees with  $P$  on all of  $\mathcal{C}$ . Lemma 11 completes the proof.  $\square$

5.3. *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  $\{(A_1, B_1), \dots, (A_n, B_n)\} \subseteq \mathcal{C}$  be a finite subcollection. We can also apply Lemma 9. For each  $i$  such that  $P(A_i|B_i) \neq R(A_i|B_i)$  (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  $k = 0, 1$ . Since the finite subcollection was arbitrary, Conclusion 4 now follows.

5.4. *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 [8].

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

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

*The lower boundary of the risk set is*

$$\begin{aligned} \partial_L = \{ & (x_1, \dots, x_m) \in \overline{R} : y_i \leq x_i \text{ for all } i \text{ and } y_i < x_i \text{ for some } i \\ & \text{implies } (y_1, \dots, y_m) \notin R\}. \end{aligned}$$

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

THEOREM 7. *Let  $\Theta = \{\theta_1, \dots, \theta_m\}$  be a finite parameter space. Let  $\mathcal{A}$  be an action space. Let  $L : \Theta \times \mathcal{A} \rightarrow \mathbb{R}$  be a loss function that is bounded below. Let  $a_0 \in \mathcal{A}$  be an action that is not Bayes for even a single prior distribution and such that  $L(\theta_j, a_0) < \infty$  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  $\delta$  is

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

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

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

The minimax theorem (for example, [11, Theorem 3.77]) says that the decision problem has a least favorable distribution  $\mathbf{u} = (u_1, \dots, u_m)$  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  $j$ . Since  $a_0$  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  $\delta$  such that  $R(\theta_j, \delta) < 0$  for all  $j$ , and so

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

This completes the proof of the first claim.

If the risk set is closed from below, it follows from [11, Theorem 3.77] that there is a minimax rule  $\delta_0$  that is also a Bayes rule with respect to  $\mathbf{u}$ .  $\square$

The proof of Theorem 2 begins by noting that we have enough assumptions to apply Lemma 10, hence there is a finite subcollection  $A_1, \dots, A_n$  such that (8) fails for every finitely additive probability  $R$ . Let  $p_i = P(A_i|B_i)$  for  $i = 1, \dots, n$ , and let  $C_1, \dots, C_m$  be the distinct nonempty constituents from Definition 6. Let  $d_1, \dots, d_m$  be the total scores from Definition 7.

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

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

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

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

PROOF. Construct problem  $\mathcal{A}$  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  $\mathbf{p} = (p_1, \dots, p_n) \in \mathcal{O}$  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  $\delta_0$  that is also a Bayes rule with respect to a prior  $\mathbf{u} = (u_1, \dots, u_m)$ . Since  $\mathbf{u}$  is a probability vector, it corresponds to an essentially unique set of

forecasts  $(r_1, \dots, r_n)$  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

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

The Bayes risk of  $\delta_0$  is

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

Each summand inside the integral can be minimized separately by  $q_i = r_i$ , hence  $\delta_0$  has the same Bayes risk as the nonrandomized rule  $\mathbf{r} = (r_1, \dots, r_n)$ . So, we assume that  $\delta_0$  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.  $\square$

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

PROOF. Let  $s_1 = \dots = s_m = 1/m$ , and define, for  $i = 1, \dots, n$ ,  $Q(A_i \cap B_i) = \sum_{j=1}^m a_i(j)b_i(j)s_j$  and  $Q(A_i^C \cap B_i) = \sum_{j=1}^m [1 - a_i(j)]b_i(j)s_j$ , so that  $Q(B_i) = \sum_{j=1}^m b_i(j)s_j$ . Then

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

Because  $q_1, \dots, q_n$  correspond to a probability, they are coherent<sub>1</sub> conditional forecasts and they have finite total scores in all constituents. Hence, Conclusion 3 holds, which implies Conclusions 1 and 2.  $\square$

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

PROOF. 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

$A_i \neq \emptyset$  or  $p_i = 1$  but  $A_i \neq B_i$ . (And, of course, the scoring rule corresponding to  $(A_i, B_i)$  has to be unbounded at the appropriate endpoint.) Define

$$I = \{i : \text{either } (g_{1,A_i,B_i}(p_i) = \infty \text{ and } A_i \neq \emptyset) \\ \text{or } (g_{0,A_i,B_i}(p_i) = \infty \text{ and } A_i \neq B_i)\}.$$

Because  $J \neq \emptyset$ , it follows that  $I \neq \emptyset$ . 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 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  $\mathbf{p} = (p_1, \dots, p_n)$  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  $i = 1, \dots, n$ . (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  $i = 1, \dots, n$ ,  $x = p_i$  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

$$\begin{aligned} \sum_{i \in I^C} r_{1,i}h_{1,A_i,B_i}(p_i) + r_{2,i}h_{0,A_i,B_i}(p_i) &= \sum_{i \in I^C} r_{1,i}g_{1,A_i,B_i}(p_i) + r_{2,i}g_{0,A_i,B_i}(p_i) \\ (22) \qquad \qquad \qquad &= \sum_{i \in I^C} r_{1,i}g_{1,A_i,B_i}(r_i) + r_{2,i}g_{0,A_i,B_i}(r_i), \end{aligned}$$

where the last equality follows from the fact that  $r_i$  for  $i = 1, \dots, n$  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 (22), hence  $\mathbf{p}$  is not weakly Bayes under the  $h$ -scores.

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

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

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

where  $\mathcal{O}$  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  $\delta$  dominates the original forecasts amongst those constituents where the original  $g$ -scores are finite. Table 2 summarizes some of what we know about the  $g$ -scores of  $\delta$  and the incoherent<sub>1</sub> forecasts. The reason that  $E = 0$  is that all of the  $p_i$  for  $i \in I$  are extreme and they

TABLE 2

*Total  $g$ -scores for the incoherent<sub>1</sub> forecasts and the dominating randomized rule  $\delta$ . 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  $i \in I^C$ .*

	$i \in I$	$i \in I^C$
	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 $\delta$	
$j \in J$	$C = \sum_{i \in I} \int_{\mathcal{A}} b_i(j) g_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$	$D = \sum_{i \in I^C} \int_{\mathcal{A}} b_i(j) h_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$
$j \in J^C$	$G = \sum_{i \in I} \int_{\mathcal{A}} b_i(j) g_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$	$H = \sum_{i \in I^C} \int_{\mathcal{A}} b_i(j) h_{a_i(j), A_i, B_i}(q_i) \delta(d\mathbf{q})$

contribute either 0 or  $\infty$  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,  $H \leq F - w$ . If we knew that  $G < w$ , then we would know that  $\delta$  weakly dominates the incoherent<sub>1</sub> forecasts under the  $g$ -scores. If, in addition, we knew that  $C < \infty$ , then we would know that  $\delta$  strictly dominates the incoherent<sub>1</sub> 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  $\delta$ ) of the  $i$ th coordinates of  $\mathbf{q}$  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 2 is affected. We proceed now to replace  $\delta$  by another randomized rule  $\delta'$  in order to make  $G < w$ , and if Assumption 2 holds for all scoring rules,  $C < \infty$ .

First, consider the case in which we do not assume that every scoring rule satisfies Assumption 2. Define  $\delta'$  as follows. The joint distribution of  $\{q_i : i \in I^C\}$  is the same as that of  $\delta$ . Under  $\delta'$ ,  $q_i = p_i$  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  $G + H < E + F$  while  $A + B = C + D$ .

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  $g_{k,A_i,B_i}(t_{k,i}) \leq v$ . Let  $\delta'$  be defined as follows. The joint distribution of  $\{q_i : i \in I^C\}$  is the same as that of  $\delta$ . Under  $\delta'$ ,  $q_i = t_{p_i,i}$  with probability 1 for all  $i \in I$ . With this change, we have that  $C < \infty$  because  $\delta'$  gives 0 mass to extreme forecasts in the coordinates in  $I$ . Also, for  $j \in J^C$ ,  $G \leq n \max_{i \in I} g_{a_i(j),A_i,B_i}(t_{p_i,i}) \leq nv = w/2$ . 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  $\mathbf{q}$  is coherent, the proof is complete. If the forecasts in  $\mathbf{q}$  are not coherent<sub>1</sub> but are weakly Bayes, then Theorem 3 says that there are coherent<sub>1</sub> forecasts that have the same total score in every constituent and hence strictly dominate  $\mathbf{p}$ . If  $\mathbf{q}$  are neither coherent<sub>1</sub> nor Bayes, at least they produce finite score in every constituent, and they strictly dominate  $p_1, \dots, p_n$ . Now, apply Lemma 12 to  $\mathbf{q}$  producing another nonrandomized rule that is coherent<sub>1</sub> and strictly dominates  $\mathbf{q}$  and hence strictly dominates  $\mathbf{p}$ , so that Conclusion 3 holds.  $\square$

**COROLLARY 1.** If the forecasts are coherent<sub>3</sub>, 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<sub>3</sub>.  $\square$

**COROLLARY 2.** If the forecasts are coherent<sub>3</sub>, 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<sub>3</sub>.  $\square$

5.5. *Theorem 1.* Assume that we have conditional forecasts for the pairs of events in a collection  $\mathcal{C}$  and that the forecast for each pair  $(A, B) \in \mathcal{C}$  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<sub>3</sub> if and only if they are weakly Bayes. For the second claim, apply Corollary 2 to conclude that the forecasts are coherent<sub>3</sub> 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<sub>1</sub>.

**6. Discussion.** We have given sufficient conditions for a set of incoherent (Definition 1) forecasts to be weakly or strictly dominated (according to proper scoring rules) by either a coherent<sub>1</sub> set of forecasts or by something else. Our conditions are not necessary. On the other hand, for each of our conditions, we have provided an example to show that the condition cannot be eliminated without replacing it by some other condition that would rule out the example. For example, 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 3 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 (16) doesn't 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  $\ell_i^*(r)$  defined in the proof of Lemma 10 will be continuous at all  $r$  that matter. As another example, 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 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 between 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). [7] claim that, if one uses continuous merely proper scoring rules, one can guarantee a weakly dominating coherent<sub>1</sub> set of forecasts. In reality, one gets either a strictly dominating coherent<sub>1</sub> set of forecasts (via the third part of Theorem 2) or a coherent<sub>1</sub> set of forecasts with identical scores (via Theorem 3). There is no middle ground in which one can only achieve 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 used Brier score to establish the equivalence between coherence<sub>1</sub> and coherence<sub>2</sub> of a set of previsions over the class of bounded variables, measurable with respect to some common measurable space. Our Theorems 6 and 7, on which relies the proof of our main result, 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 proper scoring rule for bounded random variables. If we are interested only in scoring previsions, suppose that  $X$  is a bounded random variable and  $x$  is a proposed prevision. 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 [9]. We conjecture that a collection of previsions for bounded random variables is coherent<sub>1</sub> if and only if it is impossible to find an alternative collection of possible previsions which lead to uniformly smaller total score.

On a final note, we should say something about the impact of conditional previsions given events whose probabilities are or could be coherently assigned as 0. For example, if one assigns only two previsions,  $P(A_1|B) = 0.9$  and  $P(A_2|B) = 0.7$ . Suppose also that  $B \neq \Omega$  and  $A_1 \cap A_2 = \emptyset$ . These previsions are coherent<sub>1</sub> because no book can be made against them. Because  $B \neq \Omega$ , Theorem 5 allows  $P(B|\Omega) = 0$ . However, one might be uncom-

portable giving conditional probabilities (conditional on the same event) to disjoint events that add up to more than 1. [5, 6] show that for every finitely additive probability there exists a collection of coherent<sub>1</sub> conditional previsions that also satisfies the axioms of probability conditional on events with 0 prevision. That is,

$$\begin{aligned} P(\cdot|B) & \text{ is a probability for every } B \neq \emptyset, \\ P(B|B) & = 1, \text{ for every } B \neq \emptyset, \\ P(A \cap C|B) & = P(C|B)P(A|B \cap C) \text{ for all } A, B, \text{ and } C \\ & \text{ such that } B \cap C \neq \emptyset. \end{aligned}$$

In this paper, we have not required that conditional previsions satisfy these additional requirements when  $P(B|\Omega) = 0$  is either stated or allowed (by extension). The reason is that the equivalence between coherence<sub>1</sub> and coherence<sub>3</sub> does not depend on whether the additional requirements hold. We conjecture that, if one further restricted the concept of coherence<sub>1</sub> to require such additional properties, then one could modify the concept of coherence<sub>3</sub> to remain equivalent. The modification would need to prevent strict dominance on certain events related to those on which various previsions were conditioned.

## References.

- [1] BRIER, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78** 1–3.
- [2] DE FINETTI, B. (1972). *Probability, Induction and Statistics*, Wiley, New York.
- [3] DE FINETTI, B. (1974). *The Theory of Probability*, Wiley, New York.
- [4] DE FINETTI, B. (1981). The role of “dutch books” and “proper scoring rules”. *British J. Phil. Sci.* **32** 55–56.
- [5] DUBINS, L. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Annals of Probability* **3** 89–99.
- [6] KRAUSS, P. H. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Math. Sci. Hungar.* 229–241.
- [7] LIEB, E. H., OSHERSON, D., PREDD, J. and SEIRINGER, R. (2007). Probabilistic coherence and proper scoring rules. Technical Report ???, Princeton University.
- [8] PEARCE, D. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica* **52** 1029–1050.
- [9] SAVAGE, L. J. (1971). Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.
- [10] SCHERVISH, M. J. (1989). A general method for comparing probability assessors. *Annals of Statistics* **17** 1856–1879.
- [11] SCHERVISH, M. J. (1995). *Theory of Statistics*, Springer-Verlag, New York.

MARK J. SCHERVISH AND JOSEPH B. KADANE  
DEPARTMENT OF STATISTICS  
CARNEGIE MELLON UNIVERSITY  
PITTSBURGH, PA 15213  
E-MAIL: mark@stat.cmu.edu  
kadane@stat.cmu.edu

TEDDY SEIDENFELD  
DEPARTMENT OF STATISTICS  
AND DEPARTMENT OF PHILOSOPHY  
CARNEGIE MELLON UNIVERSITY  
PITTSBURGH, PA 15213  
E-MAIL: teddy@stat.cmu.edu