The Annals of Statistics
1985, Vol. 13, No. 3, 1198–1212

MODELING EXPERT JUDGMENTS FOR BAYESIAN UPDATING[1]

BY CHRISTIAN GENEST AND MARK J. SCHERVISH

University of Waterloo and Carnegie-Mellon University

This paper examines how a Bayesian decision maker would update his/her probability p for the occurrence of an event A in the light of a number of expert opinions expressed as probabilities $q_1, \cdots, q_n$ of A. It is seen, among other things, that the linear opinion pool, $\lambda_0 p + \sum_{i=1}^n \lambda_i q_i$, corresponds to an application of Bayes' Theorem when the decision maker has specified only the mean of the marginal distribution for $(q_1, \cdots, q_n)$ and requires his/her formula for the posterior probability of A to satisfy a certain consistency condition. A product formula similar to that of Bordley (1982) is also derived in the case where the experts are deemed to be conditionally independent given A (and given its complement).

1. Introduction. This paper examines the situation in which a decision maker (DM) uses the opinions of $n \geq 1$ expert individuals to revise his/her own probability for the occurrence of A, an event of interest. These subjective opinions are represented by probabilities $q_1, \cdots, q_n$ of A, and it is desired to use this information to form the DM's posterior probability for A. The approach here is that of Morris (1974, 1977), French (1980, 1981), Winkler (1968, 1981), and Lindley (1985). The experts' opinions are treated as random variables Q_i whose values q_i, $1 \leq i \leq n$, are to be revealed to the DM. Using Bayes' Theorem, he/she can then update p, his/her prior probability of A, by forming the posterior probability p^* given $\mathbf{Q} = (Q_1, \cdots, Q_n)$ as follows

$$p^* = p \Pr(\mathbf{Q} = \mathbf{q} \mid A)/\Pr(\mathbf{Q} = \mathbf{q}). \tag{1.1}$$

In order to do this, however, the DM must proceed to elicit his/her beliefs about how good the experts are and how much common information they provide. This assessment can be modeled in different ways.

French (1981), in extending the work of Lindley, Tversky and Brown (1979) and French (1980), assumes that $\lambda = (l_1, \cdots, l_n)$ has a multivariate normal distribution in the DM's mind, where $l_i = \log[Q_i/(1 - Q_i)]$ represents the log-odds for A versus its complement $\bar{A}$ as assessed by the ith expert. These distributions for λ are conditional on whether or not A occurs and on the DM's prior information. When Bayes' rule is applied, French finds that $\log[p^*/(1-p^*)]$ equals $\log[p/(1-p)]$ plus a linear function of λ. Naturally, this solution depends on the choice of distributions for λ which implicitly model the DM's beliefs regarding the behavior of $\mathbf{Q}$. Different modeling assumptions for $\mathbf{Q}$

Received October 1983; revised October 1984.

[1] This research was supported in part by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

AMS 1980 *subject classifications*. Primary 62C10; secondary 62A15.

Key words and phrases. Bayesian inference, consensus, expert opinions, linear opinion pool, logarithmic opinion pool.

will produce different answers for p^*; compare French (1980, 1981) with Winkler (1981) or Lindley (1985). Moreover, the choice of a model is often complicated by the lack of empirical evidence concerning the experts' performance as predictors, as well as a legitimate desire for tractable solutions. For these reasons, simpler direct ways of pooling opinions have been sought (see, for example, Section 1 in French, 1985).

Our purpose here is to illustrate how a DM might still exploit the Bayesian scheme described above, even though he/she might be unable or unwilling to assess his/her beliefs about the experts' opinions thoroughly. More specifically, we will assume throughout that the DM has specified his/her prior probability p for the occurrence of A, whose indicator will also be denoted A. We will also assume, however, that the DM does not feel confident in specifying the probability distributions $\Pr(\mathbf{Q} = \mathbf{q} \mid A)$ and $\Pr(\mathbf{Q} = \mathbf{q} \mid \bar{A})$ which are required for computing p^* in (1.1). (One reason for this could be that the DM does not feel comfortable assessing the distribution of $\mathbf{Q}$ conditional on an event which has not yet occurred.) Rather, we will suppose that initially, the DM is willing to specify either certain aspects of his/her *marginal* distribution $dF(\mathbf{q})$ of $\mathbf{Q}$, or perhaps certain features of the joint distribution of A and $\mathbf{Q}$, in addition to p. For example, the DM could specify some moments of the marginal distributions of the Q_i's (as in Sections 2 and 3), or he/she could determine that the experts are conditionally independent given A and $\bar{A}$ (see Section 4). Such specifications will be referred to as *initial specifications*, and will always include, in order to model reality more closely, the assumption that the support of dF is the entire cube $[0, 1]^n$.

Because the DM has not assessed the joint distribution of A and $\mathbf{Q}$ completely, he/she realizes that p^* in (1.1) cannot be determined uniquely. As an alternative, he/she might decide to choose a formula $p^*(\mathbf{q})$ from those which satisfy the following intuitively reasonable requirement.

CONSISTENCY CONDITION. No matter what the unspecified marginal distribution dF for $\mathbf{Q}$ is, there exists a joint distribution for A and $\mathbf{Q}$ which is compatible with dF, satisfies the initial specifications and is such that $p^*(\mathbf{q}) = \Pr(A \mid \mathbf{Q} = \mathbf{q})$.

To choose a *specific* formula $p^*(\mathbf{q})$, however, it turns out that the DM will need to specify at least some *further* aspects of the joint distribution of A and $\mathbf{Q}$ in addition to his/her initial specifications. In the one expert case, for instance, this could be accomplished by specifying the correlation between A and Q. The hope here is that while remaining entirely consistent with the Bayesian paradigm, this procedure will help the DM to produce a legitimate posterior probability for A given $\mathbf{Q} = \mathbf{q}$ without forcing him/her to endorse an entire joint distribution for A and $\mathbf{Q}$ with which he/she may not feel comfortable.

Suppose, for example, that the initial specifications consist of the DM's prior probability p and the mean vector of $\mathbf{Q}$, say $E(\mathbf{Q}) = (\mu_1, \cdots, \mu_n)$. We will show in Sections 2 and 3 that, in this case, the only formulas $p^*(\mathbf{q})$ which satisfy the above Consistency Condition are those linear combinations of the q_i's and p of the form

$$p^*(\mathbf{q}) = p + \textstyle\sum_{i=1}^n \lambda_i(q_i - \mu_i) \tag{1.2}$$

with possibly negative weights, λ_i, expressing the amount of correlation between each Q_i and A, as estimated by the DM. When $\mu_1 = \cdots = \mu_n = p$ and $\lambda_i \geq 0$ for all i, (1.2) reduces to the *linear opinion pool* of Stone (1961),

$$p^*(\mathbf{q}) = \lambda_0 p + \textstyle\sum_{i=1}^n \lambda_i q_i, \tag{1.3}$$

thereby lending support to a supposition of French (1981, page 335) to the effect that (1.3) sometimes corresponds to an application of Bayes' Theorem.

Although this work provides some additional support for the linear opinion pool, it should be emphasized that it is not written in defense of any particular pooling formula, including (1.3). Rather, the authors want to promote an approach to expert resolution which is

- not fully axiomatic, and
- Bayesian without requiring the DM to perform a very demanding assessment task.

To illustrate this point, formulas other that (1.3) are also derived in the single expert case. When the initial specification consists of the mean of some bounded function g of Q, we show that the formula for $p^*(q)$ is linear in $g(q)$. Also, when the initial specification consists of $l \geq 2$ moments for the marginal distribution of Q, a polynomial formula of degree at most l then results for $p^*(q)$. If, on the other hand, the initial specifications consist of the mean vector of the marginal distribution of $\mathbf{Q}$ as well as the requirement that the experts be independent given A and $\bar{A}$, a product formula then emerges which bears some resemblance to that obtained by Bordley (1982). As a consequence, Bordley's prescription can be interpreted as a posterior probability in a truly Bayesian fashion, even though his own derivation is *not* Bayesian (see French, 1985, Section 1.3). To prove this result, however, it will be necessary to introduce a slightly modified version of the Consistency Condition which is given in Section 4.

2. A single expert. Consider first the case in which the DM will receive the opinion of only one expert. Let Q denote the probability which the expert will assign to the event A, let p be the DM's probability of A prior to learning Q, and let $p^*(q)$ represent the DM's probability of A after learning $Q = q$. Given the joint distribution of A and Q, it is a simple matter, in theory, to calculate $\Pr(A \mid Q = q)$. Suppose, however, that the DM feels comfortable specifying μ, the marginal mean of Q, but not the remainder of the joint distribution of A and Q. Given the initial specifications $p = \Pr(A)$ and $\mu = E(Q)$, what are the possible functions $p^*(q)$ which satisfy the Consistency Condition of Section 1?

To answer this question, let $dF(q)$ denote the DM's prior marginal probability measure for Q. We assume that this measure exists, although the DM has not specified it completely. In accordance with Cromwell's Rule (Lindley, 1982), we restrict our attention to those measures dF whose support is the entire interval $[0, 1]$. In particular, this requires μ to lie *strictly* between 0 and 1. (If μ were 0, for example, this would mean that the DM is certain that the expert will say $Q = 0$.) By definition, the joint probability measure for A and Q is

$$(2.1)\quad \Pr(A, Q = q) = \begin{cases} p^*(q)\ dF(q) & \text{for } A = 1 \text{ and } q \text{ in } [0, 1], \\ [1 - p^*(q)]\ dF(q) & \text{for } A = 0 \text{ and } q \text{ in } [0, 1]. \end{cases}$$

Let $0 \le p \le 1$ be fixed. To find those functions $p^*(q)$ which satisfy the Consistency Condition in Section 1, one must guarantee that $\int p^*(q)\ dF(q) = p$ for every dF that has mean μ and the closed interval [0, 1] as its support. If such a function p^* exists, then no matter what the DM's prior distribution dF with mean μ really is, (2.1) is a joint distribution for A and Q with the property that the conditional probability of A given $Q = q$ is $p^*(q)$. Theorem 2.1 below shows that the only functions p^* which have the above property are linear functions of q.

THEOREM 2.1. *Let the initial specifications consist of* $p = \Pr(A)$, $\mu = E(Q)$, *and that the support of dF is* [0, 1]. *Then* $p^*(q)$ *satisfies the Consistency Condition of Section* 1 *if and only if*

$$(2.2)\qquad p^*(q) = p + \lambda(q - \mu)$$

for some λ *satisfying*

$$(2.3)\qquad \max\{p/(\mu - 1), (p - 1)/\mu\} \le \lambda \le \min\{p/\mu, (1 - p)/(1 - \mu)\}.$$

The proof of this theorem requires the following lemma.

LEMMA 2.2. *Let* μ *be some fixed number in the open interval* (0, 1), *and let* Δ_μ *denote the collection of all distribution functions F with support* [0, 1] *for which*

$$\int_{[0,1]} t\ dF(t) = \mu.$$

If k is a real-valued Lebesgue measurable function on [0, 1] *such that*

$$(2.4)\qquad \int_{[0,1]} k(t)\ dF(t) = p \text{ for all } F \text{ in } \Delta_\mu,$$

then $k(t) = \lambda(t - \mu) + p$ *for some real number* λ.

PROOF. Let $F_* \in \Delta_\mu$ be arbitrary. First consider the set of all $0 \le x < \mu$. For each such x, let F_x be the distribution that has mass $(1 - \mu)/(1 - x)$ at x and mass $(\mu - x)/(1 - x)$ at 1. Then $F = \frac{1}{2}F_* + \frac{1}{2}F_x$ is in Δ_μ. Now assume that k satisfies (2.4). It follows that

$$(2.5)\qquad p = \int_{[0,1]} k(t)\ dF(t) = \tfrac{1}{2}p + \tfrac{1}{2}k(x)\,\frac{1 - \mu}{1 - x} + \tfrac{1}{2}k(1)\,\frac{\mu - x}{1 - x}.$$

Setting the far left and right members of (2.5) equal yields

$$(2.6)\qquad k(x) = \frac{k(1) - p}{1 - \mu}(x - \mu) + p, \quad \text{for all} \quad x < \mu.$$

A similar argument for $1 \geq x > \mu$ yields

$$k(x) = \frac{p - k(0)}{\mu}(x - \mu) + p, \quad \text{for all} \quad x > \mu. \tag{2.7}$$

Plugging $x = 0$ into (2.6) and $x = 1$ into (2.7) shows that the two coefficients of $(x - \mu)$ are equal. Call their common value λ. It is trivial to see that $k(\mu)$ must equal p, and this completes the proof of the lemma. □

PROOF OF THEOREM 2.1. It is straightforward to show that if $p^*(q) = p + \lambda(q - \mu)$ for some λ satisfying (2.3), then (2.1) is a joint distribution for A and Q for all marginal probability measures dF for Q with mean μ. To prove the other implication, note that Lemma 2.2 implies that $p^*(q) = p + \lambda(q - \mu)$. Since $p^*(q)$ is a probability, the constant λ must satisfy (2.3). □

If, instead of specifying the mean of Q, the DM had specified the mean of $g(Q)$ for some measurable function g, an argument nearly identical to the proof of Theorem 2.1 would show that the only functions p^* satisfying the Consistency Condition are of the form $p^*(q) = \lambda(g(q) - \mu) + p$, where λ satisfies

$$\begin{aligned} &\max\{p/(\mu - \bar{g}), (p - 1)/(\mu - \underline{g})\} \\ &\qquad\qquad \leq \lambda \leq \min\{p/(\mu - \underline{g}), (1 - p)/(\bar{g} - \mu)\}, \end{aligned} \tag{2.8}$$

with $\underline{g} = \inf_{q\in[0,1]} g(q)$ and $\bar{g} = \sup_{q\in[0,1]} g(q)$. Although this result applies to an arbitrary function g, it has serious implications only when g is bounded. Indeed, the upper and lower bounds in (2.8) are both zero if g is unbounded. For the remainder of this paper, we will assume that the bounded function of interest is the identity function $g(x) = x$. In this case, note that if $\mu = p$ and $\lambda \geq 0$, then p^* is a convex combination of p and q. This combination rule $\lambda q + (1 - \lambda)p$ was derived axiomatically by Morris (1983, Section 6), but Schervish (1983) has pointed out that the supporting system of axioms is flawed.

Consider next how a DM could use Theorem 2.1 to choose a formula $p^*(q)$ for his/her posterior probability of A. His/her ability to model the relationship between A and Q is drastically limited by this approach. Namely, the conditional probability measure of Q given A (and $\bar{A}$) must now be chosen from the class of measures of the form

$$\begin{aligned} &\Pr(Q = q \mid A) \\ &\qquad = \begin{cases} [1 + \lambda(q - \mu)/p]dF(q) & \text{for } A = 1, q \text{ in } [0, 1], \\ [1 - \lambda(q - \mu)/(1 - p)]dF(q) & \text{for } A = 0, q \text{ in } [0, 1], \end{cases} \end{aligned} \tag{2.9}$$

i.e., the marginal probability measure of Q times a linear function of Q. This is unlike the usual Bayesian approach to modeling, in which the DM would assess *directly* the conditional distributions of Q given A and $\bar{A}$. Because of the DM's unwillingness to specify these conditional distributions completely, the Consistency Condition requires him/her to choose a model of the form (2.9). In such a model, the linear functions $1 + \lambda(q - \mu)/p$ and $1 - \lambda(q - \mu)/(1 - p)$ simply represent the ways in which the conditional distributions of Q given A and $\bar{A}$, respectively, differ from the marginal of Q. In making his/her final choice of a formula $p^*(q)$, therefore, these differences are the further aspect of the joint distribution of A and Q which the DM must assess. In principle, this task should be simpler than that of completely specifying two conditional distributions.

EXAMPLE 2.3. Suppose that the DM's initial specifications are $p = 0.5$ and $\mu = 0.6$. From (2.9), we see that the ratio of the conditional probability measure of Q given A to that given $\bar{A}$ is

$$\begin{aligned} R(q) &= \Pr(Q = q \mid A)/\Pr(Q = q \mid \bar{A}) \\ &= [1 + 2\lambda(q - 0.6)]/[1 - 2\lambda(q - 0.6)] \end{aligned} \tag{2.10}$$

where $0 \leq q \leq 1$. By Equation (2.3), we know that $|\lambda| \leq 5/6$, and so $R(q)$ must satisfy

$$\min\{q/(1.2 - q), (1.2 - q)/q\} \leq R(q) \leq \max\{q/(1.2 - q), (1.2 - q)/q\}$$

for all q. In order to choose his/her specific value of λ, the DM need only assess $R(q)$ for some $q \neq 0.6$. For instance, he/she might estimate that the expert is twice as likely to say $q = 0.85$ if A later occurs than if $\bar{A}$ occurs. In that case, $R(.85) = (2 + \lambda)/(2 - \lambda) = 2$ implies that λ should be $2/3$.

It is worthwhile to observe that λ in Equation (2.2) can also be interpreted as the coefficient of linear regression of A on Q. Indeed, it follows easily from (2.9) or from (2.1) that $E[(A - p)(Q - \mu)] = \lambda \operatorname{Var}(Q)$, so that

$$\lambda = \operatorname{Corr}(A, Q)[\operatorname{Var}(A)/Var(Q)]^{1/2},$$

which is the coefficient of linear regression of A on Q in the DM's opinion. For example, if the DM thinks that Q is more likely to be high (and less likely to be low) when $A = 1$ than when $A = 0$, he/she should choose $\lambda > 0$. In fact, the more closely the DM feels Q varies with A, the higher λ should be. On the other hand, negative values of λ indicate that the DM believes Q varies inversely with A. Incidentally, note that even though $\operatorname{Var}(Q)$ appears in the above formula for λ, it is *not* necessary to specify its value in order to determine λ. Rather, one can either assess λ as a regression coefficient per se, or proceed along the lines of Example 2.3. Either way, the λ in (2.2) is operationally defined in terms of the DM's joint distribution of A and Q, unlike the expert weights in many other derivations of the linear opinion pool. Finally, one should note what the restrictions (2.3) say about the relationship between p and μ. When the DM has a great deal of confidence in the expert and wishes to use a value of λ very near 1, then (2.3) says that μ must be close to p. That is, the DM believes that the mean of the expert's probability Q is near the DM's prior probability p when he/she believes the expert is very good (high λ). Alternatively, (2.3) can be interpreted as saying that if the expert is very good, then the DM's prior probability p will be close to the mean of Q. The converse, however, may not be true. That is, even if the DM believes the mean of Q is near p, there is no need to choose λ near 1.

As was mentioned in the introduction, the DM may sometimes be willing to specify other aspects of his/her marginal distribution of Q beyond its mean. In that case, the class of formulas $p^*(q)$ satisfying our Consistency Condition will be enlarged. The following natural generalization of Lemma 2.2 makes this observation more precise.

LEMMA 2.4. Let D^l denote the set of all l-dimensional vectors whose ith coordinate is the ith moment of a distribution on $[0, 1]$. Let $\mu = (\mu_1, \cdots, \mu_l)$ be some fixed vector in D^l which is the vector of moments of a distribution with support $[0, 1]$. Let Δ_μ denote the collection of all distribution functions F with support $[0, 1]$ for which

$$\int_{[0,1]} t^i \, dF(t) = \mu_i \quad \text{for} \quad i = 1, \cdots, l. \tag{2.11}$$

If k is a real-valued Lebesgue measurable function on $[0, 1]$ such that

$$\int_{[0,1]} k(t) \, dF(t) = p \quad \text{for all} \quad F \text{ in } \Delta_\mu, \tag{2.12}$$

then $k(t) = \sum_{i=1}^{l} \lambda_i (t^i - \mu_i) + p$ for some real numbers λ_i, $i = 1, \cdots, l$.

PROOF. Choose t_0 in $[0, 1]$ and $d \geq 0$ such that $t_0 + (l + 1)d \leq 1$. Define the following two discrete distribution functions:

$$P = 2^{-l} \sum_{k=1, k \text{ odd}}^{l+1} \binom{l+1}{k} I_{[t_0 + kd, 1]},$$

$$R = 2^{-l} \sum_{k=0, k \text{ even}}^{l+1} \binom{l+1}{k} I_{[t_0 + kd, 1]}.$$

By writing the binomial expansion of $(t - 1)^l$ and its first l derivatives at $t = 0$, we can easily show that P and R have the same first l moments. Let $m = (m_1, \cdots, m_l)$ be the vector of these l moments. Using an observation on page 65 of Karlin and Shapley (1953) together with their Theorem 20.1, we can show that μ is in the interior of D^l. It follows that there exists $\varepsilon > 0$ and a distribution function G whose vector of moments is $(1 - \varepsilon)^{-1}(\mu - \varepsilon m)$. Let $H_1 = (1 - \varepsilon)G + \varepsilon P$ and let $H_2 = (1 - \varepsilon)G + \varepsilon R$. Then both H_1 and H_2 are in Δ_μ. Applying (2.12) to both H_1 and H_2 yields

$$\sum_{j=0}^{l+1} (-1)^{l+1-j} \binom{l+1}{j} k(t_0 + jd) = 0 \quad \text{for all} \quad t_0 \text{ and } d.$$

It follows from results of Fréchet (1909) and Popoviciu (1934) that k must be a polynomial of degree at most l. That the polynomial must have the stated form then follows from (2.11). □

It follows from Lemma 2.4 that if $p^*(q)$ is to be a coherent conditional probability of A given $Q = q$ no matter what the prior distribution of Q is (so long as its vector of first l moments is μ and its support is [0, 1]), then p^* must be a polynomial in q of degree at most (but possibly) l, viz.

$$p^*(q) = p + \textstyle\sum_{i=1}^{l} \lambda_i(q^i - \mu_i). \tag{2.13}$$

To illustrate the connection between Theorem 2.1 and the above result, suppose that a DM specifies the mean μ of the marginal distribution of Q and then chooses $p^*(q)$ of the form (2.2). In the event that he/she later decides to specify further moments of Q, say $E(Q^i)$ for $i = 2, \cdots, l$, the DM is assured that (2.2) will remain a coherent posterior probability because (2.2) is a polynomial of degree at most l. This is not to say that the DM must still use a linear opinion pool, but rather than he/she *need not* determine a more complicated $p^*(q)$ in order to remain coherent. By electing to retain the linear formula, in effect, the DM is simply asserting his/her willingness to set $\lambda_2 = \cdots = \lambda_l = 0$ in Equation (2.13). This, in turn, can be interpreted in terms of his/her beliefs regarding the joint distribution of A and Q.

EXAMPLE 2.3 (continued). Now suppose that the DM calculates $R(0.1) = 0.2$ from (2.10) but that he/she feels that $R(0.1)$ should be equal to 0.3 instead. The effect of this would be to increase the value of $p^*(q)$ for small values of q. To allow this extra amount of flexibility in the modeling of the function $p^*(q)$, the DM may then specify a second moment of the marginal distribution of Q and (in the case $p = 0.5$) solve two linear equations in λ_1 and λ_2 involving the ratio

$$\frac{\Pr(Q = q \mid A)}{\Pr(Q = q \mid \bar{A})} = \frac{1 + 2\lambda_1(q - \mu_1) + 2\lambda_2(q^2 - \mu_2)}{1 - 2\lambda_1(q - \mu_1) - 2\lambda_2(q^2 - \mu_2)}.$$

Here, $\mu_1 = 0.6$ as before and if the DM chooses $\mu_2 = 0.42$, say, then a straightforward calculation shows that $\lambda_1 = 0.2689$ while $\lambda_2 = 0.3287$. The posterior probability $p^*(q)$ is now a quadratic in q given by $0.2006 + 0.2689q + 0.3287q^2$. This is virtually the same as the linear formula $0.1 + 0.6667q$ obtained earlier except for values of q less than 0.2 for which the new $p^*(q)$ takes slightly higher values.

We do not mean to imply that, in general, it is an easy matter for a DM to assess $R(q)$ as in the above example. We do believe, however, that this might prove an easier task than that of assessing the entire conditional distributions of Q given A and $\bar{A}$.

3. Several experts. To treat the case in which a DM wishes to use the opinions of several different experts simultaneously, it is necessary to consider the joint distribution of the experts' opinions and A. To parallel the discussion in the previous section, suppose that $n \geq 2$ experts have agreed to state their probabilities $Q_1, \cdots, Q_n$ for the event A, but the DM does not wish to assess the entire joint distribution of the vector $\mathbf{Q} = (Q_1, \cdots, Q_n)$. Rather the DM will initially specify his/her prior probability p of A and the mean μ_i of Q_i for each $i = 1, \cdots, n$. He/she will then choose a formula $p^*(\mathbf{q})$ for his/her conditional probability of A given $\mathbf{Q} = \mathbf{q}$ among those which satisfy the Consistency Condition of Section 1. As in (2.1), we can write the joint probability measure for A and $\mathbf{Q}$ as

$$\text{(3.1)} \quad \Pr(A, \mathbf{Q} = \mathbf{q}) = \begin{cases} p^*(\mathbf{q})dF(\mathbf{q}) & \text{for } A = 1 \text{ and } \mathbf{q} \text{ in } [0, 1]^n, \\ [1 - p^*(\mathbf{q})]dF(\mathbf{q}) & \text{for } A = 0 \text{ and } \mathbf{q} \text{ in } [0, 1]^n. \end{cases}$$

The posterior probability of A given $\mathbf{Q} = \mathbf{q}$ is then $p^*(\mathbf{q})$, while the conditional probability measure of $\mathbf{Q}$ given $A = 1$ is $p^*(\mathbf{q})dF(\mathbf{q})/p$. We need to ask once again for what functions p^* is $\int p^*(\mathbf{q})dF(\mathbf{q}) = p$ for all distributions dF with mean vector $\boldsymbol{\mu}$. If any such function p^* exists, then (3.1) is a valid joint probability measure for A and $\mathbf{Q}$ and the posterior probability of A given $\mathbf{Q} = \mathbf{q}$ is $p^*(\mathbf{q})$. Lemma 3.1 prescribes the form of p^* in the case where the support of dF is required to be the entire cube $[0, 1]^n$.

LEMMA 3.1. *Let μ be some fixed vector in $(0, 1)^n$, and let Δ_μ denote the collection of all n-dimensional distribution functions F with support $[0, 1]^n$ and mean vector μ. If k is a real-valued Lebesgue measurable function on $[0, 1]^n$ such that*

$$\text{(3.2)} \qquad \int_{[0,1]^n} k(\mathbf{t})\, dF(\mathbf{t}) = p \quad \textit{for all} \quad F \textit{ in } \Delta_\mu,$$

then $k(\mathbf{t}) = \sum_{i=1}^n \lambda_i(t_i - \mu_i) + p$ for some real numbers λ_i, where $\mathbf{t} = (t_1, \cdots, t_n)$.

PROOF. Since we have already proven the result for $n = 1$, we proceed by induction. Suppose the following is true for all $s = 1, \cdots, n - 1$:

INDUCTION HYPOTHESIS: *If $k(\mathbf{t})$ satisfies $\int_{[0,1]^s} k(\mathbf{t})\, dF(\mathbf{t}) = p$ for all distribution functions F with support $[0, 1]^s$ and mean vector $(\mu_1, \cdots, \mu_s)$, then $k(\mathbf{t}) = p + \sum_{i=1}^s \lambda_i(t_i - \mu_i)$ for some constants λ_i, no matter what $0 < \mu_1, \cdots, \mu_s < 1$.*

Let $F_1(t_1)$ be a distribution function with support $[0, 1]$ and mean μ_1, and let $F_{(n-1)}(t_2, \cdots, t_n)$ be a distribution function with support $[0, 1]^{n-1}$ and mean $(\mu_2, \cdots, \mu_n)$. If $L(t_2, \cdots, t_n) = \int_{[0,1]} k(t_1, \cdots, t_n)\, dF_1(t_1)$, then by (3.2) we have that

$$\int L(t_2, \cdots, t_n)\, dF_{(n-1)}(t_2, \cdots, t_n) = p.$$

From the induction hypothesis, it follows that

$$L(t_2, \cdots, t_n) = p + \textstyle\sum_{i=2}^n b_{1i}(t_i - \mu_i).$$

Define $H(t_1 \mid t_2, \cdots, t_n) = k(t_1, \cdots, t_n) - \sum_{i=2}^n b_{1i}(t_i - \mu_i)$. For each $(t_2, \cdots, t_n)$ and each F_1 as above,

$$\int H(t_1 \mid t_2, \cdots, t_n)\, dF_1(t_1) = p.$$

By induction, $H(t_1 \mid t_2, \cdots, t_n) = p + b_1(t_2, \cdots, t_n)(t_1 - \mu_1)$, and hence

$$k(\mathbf{t}) = p + b_1(t_2, \cdots, t_n)(t_1 - \mu_1) + \textstyle\sum_{i=2}^n b_{1i}(t_i - \mu_i).$$

Repeat the above argument separating out t_2 instead of t_1, then t_3 etc. For each $1 \le j \le n$, we obtain

$$k(\mathbf{t}) = p + b_j(t_1, \cdots, t_n \backslash t_j)(t_j - \mu_j) + \textstyle\sum_{i \ne j} b_{ji}(t_i - \mu_i), \tag{3.3}$$

where the variables following the symbol \ are deleted from the list preceding them. Next, set the right-hand sides of (3.3) equal to each other for every distinct pair of values i and j. It can then be seen that

$$k(\mathbf{t}) = p + c_{ij}(t_1, \cdots, t_n \backslash t_i, t_j)(t_j - \mu_j)(t_i - \mu_i) + \textstyle\sum_{m=1}^{n} \lambda_m(t_m - \mu_m),$$

where c_{ij} is defined by the relation

$$\lambda_i + c_{ij}(t_1, \cdots, t_n \backslash t_i, t_j)(t_j - \mu_j) = b_i(t_1, \cdots, t_n \backslash t_i).$$

Continuing in this manner, we obtain

$$k(\mathbf{t}) = p + c \textstyle\prod_{i=1}^{n} (t_i - \mu_i) + \sum_{i=1}^{n} \lambda_i(t_i - \mu_i). \tag{3.4}$$

By choosing a measure dF with appropriate first and nth moments, we can show that (3.2) will be satisfied if and only if $c = 0$ in (3.4). □

It follows from Lemma 3.1 that formulas of the form (1.2) are the only posterior probabilities for A satisfying the Consistency Condition of Section 1. As before, the λ_i's must obey a number of inequalities analogous to (2.3). If, for example, the DM feels that all λ_i's are positive, the most common case, then they must be chosen so that

$$\max\{\textstyle\sum_{i=1}^{n} \lambda_i\mu_i/p, \sum_{i=1}^{n} \lambda_i(1 - \mu_i)/(1 - p)\} \le 1, \tag{3.5}$$

which is implicitly two inequalities. For every possible combination of signs for the λ_i's, there will be two inequalities similar to (3.5) which must be satisfied. With no prior restrictions on the λ_i's, there are thus 2^{n+1} inequalities to be verified. Without listing all of these inequalities, we state the following loosely formulated theorem.

THEOREM 3.2. *Let the initial specifications consist of* $p = \Pr(A)$, $\mu = E(\mathbf{Q})$ *and that the support of* dF *is* $[0, 1]^n$. *Then* $p^*(\mathbf{q})$ *satisfies the Consistency Condition if and only if* $p^*(\mathbf{q}) = \sum_{i=1}^{n} \lambda_i(q_i - \mu_i)$ *for some* $\lambda_1, \cdots, \lambda_n$ *satisfying* 2^{n+1} *inequalities similar to* (3.5).

When $\mu_i = p$ and $\lambda_i \ge 0$ for all i, $p^*(\mathbf{q})$ reduces to the linear opinion pool of Stone (1961) in which the DM is considered as one of the experts. An advantage of our derivation, however, is that it provides an interpretation of the coefficients λ_i in terms of the regression of A on $\mathbf{Q}$. Let Σ_Q denote the covariance matrix of $\mathbf{Q}$, let σ_{AQ} be the vector of covariances between A and $\mathbf{Q}$, and use λ to represent the vector of λ_i's. Then $E[(A - p)(\mathbf{Q} - \mu)] = \Sigma_Q\lambda$, from which it follows that $\lambda = \Sigma_Q^{-1}(\sigma_{AQ})$, the vector of coefficients of the linear regression of A on $\mathbf{Q}$, so long as the experts are not linearly dependent. As in multiple regression, each λ_i can thus be thought of as a measure of the additional information that the ith expert provides over and above the other experts and what the DM already knows.

As we mentioned in the case of one expert, this approach limits the DM's ability to model the joint distribution of A and $\mathbf{Q}$. For example, suppose the DM contemplates the possibility that one or more experts might not respond. In accordance with the Bayesian philosophy, one would deal with this situation by forming the marginal joint distribution of A and those experts who do respond, and calculating the posterior from that distribution; or equivalently, one could average (1.2) with respect to the conditional distribution of the nonrespondents given the respondents. Unfortunately, this procedure will not produce a linear opinion pool unless the conditional means of the nonrespondents given the respondents is linear in the respondents' q_i's. Joint distributions for $\mathbf{Q}$ with this property exist (e.g., when the coordinates of $\mathbf{Q}$ are independent), but certainly not all joint distributions have this property. In Section 4, we will restrict our attention to those joint distributions of A and $\mathbf{Q}$ in which the Q_i are independent given A and $\bar{A}$. In this case, we will obtain combination rules satisfying a Consistency Condition similar to the one in Section 1, and which produce the same formula when averaged over nonrespondents as when only the respondents are considered.

4. Conditionally independent experts. It is common, in modeling dependent observations such as expert opinions, to assume that there is a random variable Y such that, conditional on Y, the observations are independent. The random variable Y then introduces dependence between the observations. Since we assume that all the experts are trying to forecast the random variable A, it might be natural to expect that only dependence between the experts enters through A. In this section, we consider the case in which the DM believes that the experts are independent conditional on A, the indicator of the event of interest. We believe this is the same meaning of independence which Morris (1974, 1977, 1983) adopts in his discussion of independent experts.

We will assume that the DM's initial specifications consist of his/her probability p of A, the marginal mean vector $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$ of $\mathbf{Q}$, as well as the requirement that the experts be conditionally independent given A and $\bar{A}$. In this case, we hope to find a formula $p^*(\mathbf{q})$ for the posterior probability of A which, when averaged over the conditional distribution of nonrespondents given respondents, will reduce to the formula which would have been obtained, had the respondents alone been considered a priori. With these added constraints, we will see that it is impossible for $p^*(\mathbf{q})$ to satisfy the Consistency Condition stated in Section 1. However, Formula (4.3) below will satisfy the following similar condition.

MODIFIED CONSISTENCY CONDITION. No matter what the unspecified univariate marginal distributions for the Q_i's are, there exists a compatible joint distribution for A and $\mathbf{Q}$ which satisfies the initial specifications and is such that $p^*(\mathbf{q}) = \Pr(A \mid \mathbf{Q} = \mathbf{q})$.

The above Modified Consistency Condition is actually equivalent to the original one in the contexts of Sections 2 and 3, where the initial specifications pertained only to the univariate marginal distributions of the Q_i's. In this section, however, the DM has also specified certain aspects of the joint distribution of $\mathbf{Q}$, namely conditional independence and the nonresponse condition.

For ease of notation, we will assume the existence of a dominating measure on [0, 1] and we will express all distributions as densities with respect to either that measure or the product measure which it induces in n dimensions. For instance, the measure could be taken to be the sum of the marginal distributions of the Q_i's. Let $g_i(q)$ be the conditional density of Q_i given A, and let $\bar{g}_i(q)$ be the conditional density of Q_i given $\bar{A}$. The joint density of the vector $\mathbf{Q}$ can then be written as $f(\mathbf{q}) = p \prod_{i=1}^{n} g_i(q_i) + (1-p) \prod_{i=1}^{n} \bar{g}_i(q_i)$. It follows that the conditional probability of A given $\mathbf{Q} = \mathbf{q}$ is

$$p^*(\mathbf{q}) = p[\textstyle\prod_{i=1}^{n} g_i(q_i)]/f(\mathbf{q}). \tag{4.1}$$

Equation (4.1), being a generalization of a formula of Bordley (1982), proves that his formula is a conditional probability for A given $\mathbf{Q} = \mathbf{q}$ which could have been derived within the Bayesian framework, although Bordley himself elected to use an axiomatic approach instead.

Next, let R be a subset of $\{1, \cdots, n\}$ indicating which experts respond, and let $\mathbf{q}_R$ and $\mathbf{Q}_R$ denote the subvectors of $\mathbf{q}$ and $\mathbf{Q}$, respectively, whose components are the elements of R. Its complement represents nonrespondents. The conditional density of the nonrespondents given $\mathbf{Q}_R = \mathbf{q}_R$ is $f(\mathbf{q})/f_R(\mathbf{q}_R)$, where

$$f_R(\mathbf{q}_R) = p \textstyle\prod_{i\in R} g_i(q_i) + (1-p) \prod_{i\in R} \bar{g}_i(q_i)$$

is the marginal density of $\mathbf{Q}_R$. Integrating (4.1) with respect to the conditional density of the nonrespondents given the respondents gives

$$p[\textstyle\prod_{i\in R} g_i(q_i)]/f_R(\mathbf{q}). \tag{4.2}$$

If the functions g_i and $\bar{g}_i$ do not depend on which set of experts is to be the respondents, then (4.2) is precisely the formula that the DM would have used, had R been the original set of experts. In other words, the problem with nonrespondents disappears if and only if the marginal distribution of each Q_i is functionally independent of the other experts. This implies that the choice of the functions g_i and $\bar{g}_i$ used for constructing $p^*(\mathbf{q})$ in (4.1) must be done separately for each i, as in the single expert case.

From Theorem 2.1, it follows that g_i/f_i and $\bar{g}_i/f_i$ must be linear functions of q, viz.

$$g_i(q) = [1 + \lambda_i(q - \mu_i)/p]f_i(q),$$

$$\bar{g}_i(q) = [1 + \lambda_i(\mu_i - q)/(1-p)]f_i(q),$$

where f_i is the marginal density of Q_i and each λ_i satisfies (2.3) with μ replaced by μ_i. Since Theorem 2.1 was used separately for each expert, note that the functions g_i and $\bar{g}_i$ are compatible with every *marginal* density f_i for Q_i, but not necessarily with every *joint* distribution for $\mathbf{Q}$. This is because a set of f_i's, λ_i's and p together determine the entire distribution of A and $\mathbf{Q}$, under the conditional independence assumption. Hence, two different joint densities of $\mathbf{Q}$ with the same marginals cannot both be compatible with a given set of λ_i's and p. As there is no hope of satisfying the Consistency Condition of Section 1 in this case, the Modified Consistency Condition given above was introduced instead. Our result can now be stated as follows:

THEOREM 4.1. *Assume that the DM's initial specifications consist of* $p = \Pr(A)$, $\mu_i = E(Q_i)$ *for each* i *and the assumption that the experts are conditionally independent given* A *and* $\bar{A}$. *Furthermore assume that* $p^*(\mathbf{q})$ *must satisfy the Modified Consistency Condition, and has the property that the mean of* $p^*(\mathbf{Q})$ *with respect to the conditional distribution of a set of nonrespondents given a set of respondents is the same as the formula which could be derived by considering only the respondents from the start. Then*

$$p^*(\mathbf{q}) = \frac{p^{1-n} \prod_{i=1}^{n} n_i}{p^{1-n} \prod_{i=1}^{n} n_i + (1-p)^{1-n} \prod_{i=1}^{n} (1 - n_i)}, \tag{4.3}$$

where, for each $1 \le i \le n$, $\lambda_i n_i = [p + \lambda_i(q_i - \mu_i)]$, *is between the upper and lower bounds of* (2.3) *with* μ *replaced by* μ_i.

When all of the μ_i's equal p, (4.3) becomes a special case of the formula of Bordley (1982). It should be noted that for each i, the value λ_i in (4.3) is the same as the λ in (2.2) that would be used in the single expert case. A polynomial version of Theorem 4.1 analogous to Lemma 2.4 could also be proven.

At this point, we should also comment on the effect of the assumption of conditional independence. Indeed, such an assumption has consequences which may not be apparent to the naked eye. For example, consider the case in which expert i gives $Q_i = \mu_i$, $i = 1, \cdots, n$. Formula (4.3) reduces to $p^*(\mathbf{q}) = p$. This is a comforting result: if every expert says exactly what the DM expected they would, his/her probability of A does not change. In contrast, consider the case in which expert i gives $Q_i = q_i > \mu_i$, $i = 1, \cdots, n$. Assume that each λ_i is positive, and let $z_i = p + \lambda_i(q_i - \mu_i)$. Then z_i would be the posterior probability of A given Q_i alone for each $i = 1, \cdots, n$. Formula (4.1) shows how to combine these values to get the conditional probability of A given $\mathbf{Q}$. Since each q_i is greater than μ_i, each z_i is greater than p. Simple algebra shows that, in this case, $p^*(\mathbf{q})$ will be greater than the largest of the z_i's. For example, if p equals 0.7, and if three experts give $z_1 = z_2 = z_3 = 0.8$, then $p^* = 0.92$. This overcompensation phenomenon is often called a *risky shift* (cf. Bordley, 1983). If this does not correspond to what the DM considers as appropriate behavior, he/she should consider the joint distribution of the expert opinions more carefully.

5. Conclusion. It is the authors' conviction that the Bayesian argument embodied in (1.1) is normatively the only logical method available to a decision maker for exploiting the collected opinions of a number of experts. "This has the immediate advantage," as French (1985, Section 1.5) points out, "that we can use the likelihood function to allow for the possibilities of miscalibration, dishonesty and non-independence of the experts." In practice, however, the DM will rarely have enough data to face the enormous assessment problems caused, in particular, by the dependence of information sources between the experts. The primary objective of this paper, thus, was to show how these difficulties could be avoided without abandoning the Bayesian viewpoint.

This objective was accomplished by showing how a DM could choose a formula $p^*(\mathbf{q})$ for his/her posterior probability of A without specifying an entire joint distribution for A and the expert opinions $\mathbf{Q}$. Assuming that the DM can assess his/her prior probability of A, some aspects of the marginal distribution of $\mathbf{Q}$, as well as some aspects of the joint distribution of A and $\mathbf{Q}$, we showed how a formula could be chosen which is a coherent posterior probability no matter what is the DM's complete marginal distribution of $\mathbf{Q}$. This is feasible because the class of pooling recipes consistent with (1.1) increases as the DM specifies more and more aspects of the marginal distribution dF of $\mathbf{Q}$. At one extreme, it can be seen that a Bayesian DM who leaves dF completely unspecified has no choice but to *ignore* the expert opinions and use $p^*(\mathbf{q}) \equiv p$ if he/she wants to satisfy the Consistency Condition of Section 1. At the other extreme, if dF is determined completely, $p^*(\mathbf{q})$ can be any function such that

$$\int p^*(\mathbf{q})\, dF(\mathbf{q}) = p.$$

An alternative less extreme than either of the above would be, for example, to assess only the first few moments of the marginal distribution of $\mathbf{Q}$. This is the situation which we discussed in Sections 2 and 3. When only the first moment of $\mathbf{Q}$ is specified, $p^*(\mathbf{q})$ must be a linear opinion pool whose coefficients are operationally defined. If the DM is willing to model the experts as conditionally independent given A and $\bar{A}$, a logarithmic opinion pool of the form (4.3) obtains whose coefficients also have a precise interpretation in terms of the joint distribution of A and $\mathbf{Q}$. This last result was derived under a slightly weaker consistency condition, however.

As a final note, we should comment on the restriction that the support of the distributions of the Q_i should be the entire interval [0, 1]. This assumption, which was made on account of realism, could be dropped without altering the conclusions of our theorems and lemmas. Had the results been proven without this requirement, however, a DM might have argued that since his/her distribution for $\mathbf{Q}$ was further restricted by having support [0, 1], he/she had more choices for $p^*(\mathbf{q})$ than implied by our analysis. The assumption was made in anticipation of this criticism, at the expense of slightly more complicated proofs.

Acknowledgement. We are indebted to the referees for suggestions that led to a simpler proof of Lemma 2.2 and improved readability of the paper.

REFERENCES

BORDLEY, R. F. (1982). A multiplicative formula for aggregating probability assessments. *Management Sci.* **28** 1137–1148.

BORDLEY, R. F. (1983). A Bayesian model of group polarization. *Organ. Behav. Human Perform.* **32** 262–274.

FRECHET, M. (1909). Une définition fonctionnelle des polynômes. *Nouvelles Ann. Math.* **9** 145–162.

FRENCH, S. (1980). Updating of belief in the light of someone else's opinion. *J. Roy. Statist. Soc. Ser. A* **143** 43–48.

FRENCH, S. (1981). Consensus of opinion. *European J. Oper. Res.* **7** 332–340.

FRENCH, S. (1985). Group consensus probability distributions: a critical survey. In *Bayesian Statistics* **2** (J. M. Bernardo et al., eds.) North-Holland (Amsterdam).

KARLIN, S. and SHAPLEY, L. S. (1953). Geometry of moment spaces. *Mem. Amer. Math. Soc.* **12**.

LINDLEY, D. V. (1982). The Bayesian approach to statistics. In *Some Recent Advances in Statistics.* (J. T. de Oliviera and B. Epstein, eds.) Academic, New York.

LINDLEY, D. V. (1985). Reconciliation of discrete probability distributions. In *Bayesian Statistics* **2** (J. M. Bernardo et al., eds.) North-Holland (Amsterdam).

LINDLEY, D. V., TVERSKY, A. and BROWN, R. V. (1979). On the reconciliation of probability assessments. *J. Roy. Statist. Soc. Ser. A* **142** 146–180.

MORRIS, P. A. (1974). Decision analysis expert use. *Management Sci.* **20** 1233–1241.

MORRIS, P. A. (1977). Combining expert judgments: a Bayesian approach. *Management Sci.* **23** 679–693.

MORRIS, P. A. (1983). An axiomatic approach to expert resolution. *Management Sci.* **29** 24–32.

POPOVICIU, T. (1934). Sur quelques propriétés des fonctions d'une ou de deux variables réelles. *Mathematica* **8** 1–85.

SCHERVISH, M. J. (1983). *Combining expert judgments.* Technical Report 294, Department of Statistics, Carnegie-Mellon University.

STONE, M. (1961). The linear opinion pool. *Ann. Math. Statist.* **32** 1339–1342.

WINKLER, R. L. (1968). The consensus of subjective probability distributions. *Management Sci.* **15** B61–B75.

WINKLER, R. L. (1981). Combining probability distributions from dependent information sources. *Management Sci.* **27** 479–488.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213