Test (1995)
Vol. 4, *No.* 2, *pp.* 263–313

Coherent Combination of Experts' Opinions

A. P. DAWID*, M. H. DeGROOT** and J. MORTERA***

**Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK*

****Dipartimento di Economia, Università di Roma III, Via Ostiense, 00154 Rome, Italy*

SUMMARY

An *expert* (for You) is here defined as someone who shares Your world-view, but knows more than You do, so that were She to reveal Her current opinion to You, You would adopt it as Your own. When You have access to different experts, with differing information, You require a *combination formula* to aggregate their various opinions. A number of formulae have been suggested, but here we explore the fundamental requirement of *coherence* to relate such a formula to Your joint distribution for the experts' opinions. In particular, in the context of opinions about an uncertain event A, we investigate coherence properties of the linear, harmonic and logarithmic opinion pools. Some general results on coherence of the joint forecast distribution are also developed.

Keywords: EXPERT OPINIONS; COHERENCE; COMPATIBILITY; COMBINING OPINIONS; LINEAR OPINION POOL; LOGARITHMIC OPINION POOL; HARMONIC OPINION POOL.

Received June 95; Revised November 95.

** Morrie DeGroot passed away on 2 November, 1989. This work was initiated by Dawid and DeGroot and was completed by Dawid and Mortera.

1. INTRODUCTION

It is sometimes necessary to construct a single opinion by combining a number of individual opinions. A decision maker might consult a number of experts (financial, meteorological, medical etc.) before reaching a final decision. An ideal Bayesian approach to incorporating the experts' views would be for each of them to report all the data and background knowledge on which his or her opinions are based, and for You, the decision maker, to combine all this information with Your own prior opinions and any additional data You may have, using Bayes's theorem. However, this ideal will almost always be rendered unattainable, by the extent of the data, company confidentiality, or the inability of the experts to identify clearly the empirical basis and background knowledge leading to their intuitive opinions. You are then left with only the experts' stated opinions. We shall suppose that these are expressed as probability distributions over a fixed set of events and quantities of interest. Your task is then to combine these into a suitable distribution to use as Your own.

For an extensive review of methods of combining expert opinions see, among others, Genest and Zidek (1986). The two most widely used pooling recipes are the linear opinion pool (Stone, 1961) and the logarithmic opinion pool (Genest, 1984; Genest *et al.,* 1986; Bordley, 1982). An optimal linear opinion pool was derived in DeGroot and Mortera (1991). An axiomatic approach to opinion pooling is taken by *e.g.* Madansky (1964), McConway (1981) and Morris (1983). For a detailed discussion and criticism of this approach see *e.g.* Winkler (1986), Lindley (1986), Schervish (1986), Clemen (1986) and French (1986). A Bayesian model-based approach to pooling is taken by, among others, Winkler (1981), French (1985), Lindley (1985) and Berger and Mortera (1991).

This paper investigates coherent methods for combining experts' opinions, when these are expressed as probabilities for some fixed event A. Neither axiomatic nor modelling assumptions of the usual kind are made. Instead, we work with a special and somewhat restricted definition of what constitutes "expertise", as seen by You, the decision-maker. We consider an expert to be someone who "shares Your world-view", in the sense that, if you both had identical information, you would both have identical opinions. This is a natural assumption for those who, following Keynes (1921), Jeffreys (1939) and Carnap (1950), hold a "logical" view of personal probability, under which probability is objectively determined by evidence. But even for a subjectivist it is a fairly natural assumption in our context, since it would be rash to take into account (at any rate in too naive a way) the opinion of some one whose world view was at odds with Yours. At any rate, the condition has sufficient *prima facie* appeal to make it worthwhile to investigate its consequences, as we do here. Note that, while we do not require that an expert be "well calibrated", in any frequency sense, we do assume that the probabilities the expert provides are correctly and coherently computed.

With this interpretation, the only reason for consulting an expert is that She may have additional information, thus leading to different opinions, than You do. Although it may not be possible clearly to specify this information, we can utilise the underlying identity of Your distribution and Hers to help take due account of the opinions the additional information generates. This approach allows us to characterise the appropriateness of various formulae for combining different experts' opinions. We apply a "principle of coherence", that all probability statements should be consistent with a single overall probability distribution, common to You and all the experts when you are all in the same initial state of information. This principle does not identify a combination formula, but it does limit the possibilities. In particular, under a variety of assumptions and models, various simple formulae have been suggested in the literature for combining probabilistic opinions. Here we examine when these formulae are coherent, *i.e.* justifiable within the framework we present.

1.1. *Preview*

In Section 2 we set out the definition of expertise with which we work, and note how this implies a coherence condition on formulae for combining expert opinions. For simplicity, we largely confine attention to the case of $k = 2$ experts. The main topic of this paper is the search for compatible pairs, comprising a function $\Phi(\Pi_1, \Pi_2)$ and a joint probability distribution P for Π_1 and Π_2. The definition of compatibility, with the necessary and sufficient conditions for logical consistency of compatible pairs, is given in Section 3. It is also shown how to construct compatible pairs, a problem related to recalibration. Section 4 is devoted to examples. The linear opinion pool, the harmonic opinion pool and the logarithmic opinion pool are shown to arise naturally under certain distributional assumptions. Under our interpretation of the term "expert", a notable result is that in the linear opinion pool it is not possible for all the weights to be strictly positive, contrary to assumptions commonly made. Explicit forms for combination rules Φ compatible with independence and conditional independence assumptions are studied. General theory on coherent combination rules is given in Section 5, which is closely related to results in Strassen (1965) and Gutmann *et al.* (1991). A perhaps surprising result is that for any joint distribution P of Π_1 and Π_2, if there exists a compatible combination formula Φ then there exists one yielding only 0, 1 predictions: that is, Φ states with certainty whether the event A obtains or not. A brief discussion of extensions to more than two experts is given in Section 6, and some concluding discussion in Section 7.

2. USING EXPERTS

2.1. *Who is an Expert?*

DeGroot (1988) gives two definitions of an "expert". Under a completely liberal interpretation, an expert could be anyone who gives you a prediction. Taking a more restrictive view, DeGroot then considers an expert to be

> "someone whose prediction you will simply adopt as your own posterior probability without modification. This will be the case if you believe that the expert has all the information you have that may be relevant to the occurrence or non-occurrence of the event, and possibly additional information."

It is essentially this definition that we shall adopt here, calling such an individual an "expert" (for You).

The requirement in the first part of the above quotation has been termed *probability calibration* (Lindley, 1982), or being *well calibrated* (DeGroot and Eriksson, 1985); the second part is assumed by, *e.g.* Clemen (1985). However, although it is implicitly assumed by DeGroot that the two parts are equivalent, this will in general be so only under further conditions, which we now introduce.

Our *basic assumption* is that, at some past point of time,

(i) You and the expert both started with identical information, and further

(ii) you both had a common subjective distribution $\tilde{P}$, expressing your shared uncertainty about all future events and quantities of interest.

(We do not assume that it is possible fully to articulate $\tilde{P}$.)

Suppose now that the expert alone has since obtained further information, which can be regarded as observation of the value of some (possibly highly multivariate) quantity, X say. Your state of information has not changed, and hence Your current uncertainty is still expressed by the distribution $\tilde{P}$. This is the scenario envisaged in the second part of DeGroot's definition. We shall first show that, under our basic assumption, it implies probability calibration, as expressed in the first part.

Let A be an event of interest, and define $\Pi := \tilde{P}(A|X)$, the expert's revised probability of A. Although Π is now known to the expert, to You Π remains unknown, and is thus a random variable. If You know which additional random quantity X the expert has observed, then Π is a known function of X, and thus has, for You, a well-determined distribution. Note that both You and the expert originally assigned the same prior probability $\pi_0 = \tilde{P}(A)$, which must also be Your current expectation of Π.

Now, using standard notation for conditional expectation given a random variable, and using the symbol "$\equiv$" to denote identity of functions or almost sure identity of random variables, we have (since Π is a function of X) $\tilde{P}(A|X,\Pi) \equiv \tilde{P}(A|X) \equiv \Pi$, which is (trivially) a function of Π alone. Thus, using the notation for and properties of conditional independence in Dawid (1979), under $\tilde{P}$, $A \perp\!\!\!\perp X|\Pi$. It follows that $\tilde{P}(A|X,\Pi) \equiv \tilde{P}(A|\Pi)$, that is

$$\tilde{P}(A|\Pi) \equiv \Pi. \tag{1}$$

The identity (1) is equivalent to probability calibration for the event A, as required by the first part of DeGroot's definition: when the expert reports Her probability Π for A, You will adopt it as Your own. (A more subtle analysis is required when You do not even know which quantity X forms the basis of the expert's report, but the conclusion is unaffected.) More generally, if the expert's report Π is Her full updated distribution over a collection of uncertain quantities and events of interest, then, for *any* such event we shall have probability calibration, fully adopting the expert's reported probability: $\tilde{P}(A|\Pi) \equiv \Pi(A)$.

We now study the converse relationship, assuming probability calibration and deducing, even without making our basic assumption, that we can regard the expert as having started with the same distribution as You, and then having obtained further information. Thus suppose the expert will be required to report Her current distribution Π, defined over some fixed collection $\mathcal{A}$ of events (we might have simply $\mathcal{A} = \{A\}$); and that You will then be willing simply to adopt her probabilities. For You Π is currently uncertain, and we have $\tilde{P}(A|\Pi) \equiv \Pi(A)$ for any $A \in \mathcal{A}$. Define $X \equiv \Pi$, a (somewhat abstract) random quantity for You. Then $\Pi(A) \equiv \tilde{P}(A|X)$. It follows that we can consider the expert's reported probabilities as being generated by Her having started with Your own distribution $\tilde{P}$, and then incorporating the additional quantity X.

We have thus shown the equivalence of the two parts of DeGroot's definition, so long as the second part is expanded to require that You and the expert initially share the same joint subjective distribution $\tilde{P}$ over all the relevant quantities. (We should point out, however, that whereas the second part implies the first no matter what events are being considered, the converse is false: it is possible to have an "expert" who is probability calibrated for some events, but not others.)

In this paper we shall consider only the simplest case of a single event A of interest, and expert forecasts of the form $\Pi \equiv \tilde{P}(A|X)$, for suitable X. We have shown that, under our basic assumption, if the expert knows at least as much as You do, then on learning Her probability Π (and nothing else), You would adopt Π as Your own probability for A.

2.2. *Coherent Combination*

Now suppose You have access to k different experts. If You were to obtain a probability for A from a single one of these, You would adopt it as Your own; but the various experts' probabilities typically differ, since they will be based on differing information.

Before You consult the experts, their various reports $(\Pi_1, \cdots, \Pi_k)$ will be, for You, uncertain random quantities, jointly distributed together with the uncertain event A. Since $(\Pi_1, \cdots, \Pi_k, A)$ will be the only random quantities which we shall here need to consider, it will be enough to consider Your overall joint distribution $\tilde{P}$ as defined over these alone. We shall denote by P the implied distribution for the (Π_i), marginalizing out over A.

How should You assign Your probability for A after learning all of theirs? The laws of coherence imply that You must assign the probability $\tilde{P}(A|\Pi_1, \cdots, \Pi_k)$. That is, we obtain the *combination formula*

$$\Phi(\Pi_1, \cdots, \Pi_k) \equiv \tilde{P}(A|\Pi_1, \cdots, \Pi_k). \tag{2}$$

(Note that, if expert i bases Her probability on observation of X_i, then $\Pi_i \equiv \tilde{P}(A|X_i)$, where here $\tilde{P}$ is extended to encompass the (X_i); but in general the value of X_i will not be fully recoverable from that of Π_i, so that the right hand side of (2) will not usually be the same as $\tilde{P}(A|X_1, \cdots, X_k)$.)

Now suppose that You can specify Your joint distribution P for the Π_i, and contemplate using some combination formula Φ, which You will apply to whatever values of $(\Pi_1, \cdots, \Pi_k)$ the experts report to You. The principal question we address in this paper is: when will the use of Φ be coherently *compatible* with Your joint distribution P for the reports? That is, when will there be an overall joint distribution $\tilde{P}$ under which $\Pi_i \equiv \tilde{P}(A|\Pi_i)$, the implied distribution for $(\Pi_1, \cdots, \Pi_k)$ is P, and (2) holds?

We emphasise that the scope of any combination formula Φ is regarded throughout this work as restricted to a *fixed* event A and underlying probability structure $\tilde{P}$, common to You and all the experts in your common initial state of information. The only variable arguments of Φ are $(\Pi_1, \cdots, \Pi_k)$, the (initially unknown) expert forecasts for A. In particular, and in distinction to much of the literature, we do not require that the same formula should be used for different events; nor that the formula Φ, which might be written in an algebraic form involving particular aspects of $\tilde{P}$ (*e.g.* the prior probability $\pi_0 := \tilde{P}(A)$), should retain the same form when such aspects are changed. In our approach, π_0 is a known constant, not a mathematical variable.

Although we suppose throughout that the experts all know at least as much as You do, we can also apply our results to the case in which You have some information which they do not. In this case, You can regard Yourself as just one more expert, and accordingly incorporate Your own views with the others. It is now necessary to regard the underlying common distribution $\tilde{P}$ as that obtaining before either You or the experts obtained your differing information. In particular, if You learn the opinion of just one "expert", but are Yourself party to information which

She does not have, You would not in general accept Her opinion as Your own, but instead apply the results below for the case of two experts (one being Yourself).

2.3. *Related Problems*

Before addressing the question of compatibility between P and Φ, we remark on some related problems.

2.3.1. *Frequency Calibration.*

Let $A_1, A_2, \cdots$ be a sequence of events: in the archetypical illustration, A_i is the event that it rains on day i. A forecaster provides a probability π_i for A_i. Let $a_1, a_2, \cdots$ be the actual outcomes (0 or 1). The forecaster is *well calibrated* (in the frequency sense) if, when we consider just those events for which π_i takes on some preassigned value π, the limiting relative frequency of the outcome 1 is π; and this for all $\pi \in [0,1]$ (see *e.g.* Dawid, 1982, 1986; DeGroot and Fienberg, 1983).

Whereas probability calibration refers to agreement between a single expert forecast and Your own subjective opinions, frequency calibration requires long-run agreement between a string of expert forecasts and the associated empirical outcomes. Although conceptually distinct, these two properties may be expressed in formally identical terms, an identity exploited by DeGroot (1988). Thus, from the observed pairs $(a_1, \pi_1), (a_2, \pi_2), \cdots$, form the limiting empirical joint distribution: a distribution $\tilde{P}$, say, for a random pair (A, Π). This construction has been used by Dawid (1986), Murphy and Winkler (1992). Then the frequency calibration condition can be expressed as $\tilde{P}(A|\Pi) \equiv \Pi$, identical in form to our definition of expertise, (1).

Now suppose we have k forecasters, each individually well-calibrated, but in general giving a set of different probabilities $(\pi_{1i}, \cdots, \pi_{ki})$ for each A_i. We can extend the above construction to a joint empirical distribution $\tilde{P}$ for $(A, \Pi_1, \cdots, \Pi_k)$, constructed from

$$\{(a_i, \pi_{1i}, \cdots, \pi_{ki}) : i = 1, 2, \cdots\}.$$

Again $\tilde{P}(A|\Pi_i) \equiv \Pi_i$, for each i. Then,

$$\Phi(\Pi_1, \cdots, \Pi_k) \equiv \tilde{P}(A|\Pi_1, \cdots, \Pi_k)$$

is just the limiting relative frequency of 1's in a subsequence of the events for which each Π_i takes on a pre-assigned value. Such joint frequency calibration functions have been studied by Clemen and Murphy (1986).

2.3.2. *Conflicting Reference Sets.*

Suppose You, an insurance broker, wish to assess the probability that Mr. Smith, a 42-year-old travelling salesman, will have a motor accident next year. Your actuarial tables show that 11% of 42-year-olds and that 18% of travelling salesmen have an accident each year; but are not sufficiently finely classified to give an accident rate for 42-year-old travelling salesmen. You regard Mr. Smith as exchangeable with the individuals comprised in the tables. How should You assess the probability for Mr. Smith? The analogy (again a formal identity) with our definition of expertise is clear. Let Π_1 be the actuarial accident-rate based on an individual's age, and Π_2 be that based on his occupation. Let A denote accident, and $\tilde{P}$ be the overall probability distribution to be used for setting insurance premiums. Then $\tilde{P}(A|\Pi_1) \equiv \Pi_1$, $\tilde{P}(A|\Pi_2) \equiv \Pi_2$; but we want $\tilde{P}(A|\Pi_1, \Pi_2)$. If we know the joint distribution of (Π_1, Π_2) in the population, what forms are allowable for this quantity?

An application of the above structure arises in games such as baseball, when we know the average hit rate for a batter off all pitchers, and against a pitcher by all batters, and need to assess the probability of a hit when a given pitcher meets a given batter (Gutmann *et al.,* 1991).

3. COMPATIBILITY

We now investigate in detail the property of compatibility for the case of $k = 2$ experts.

Suppose expert E_1 observes X_1 and tells You $\Pi_1 \equiv \tilde{P}(A|X_1)$. Similarly, E_2 observes X_2 and tells You $\Pi_2 \equiv \tilde{P}(A|X_2)$. Then, as shown in Section 2.1, in the joint distribution $\tilde{P}$ of (Π_1, Π_2, A), $\tilde{P}(A|\Pi_i) \equiv \Pi_i$. Define $\Phi := \tilde{P}(A|\Pi_1, \Pi_2)$, the coherent formula for combining the two expert opinions, and let P denote the distribution of (Π_1, Π_2). One then has

$$0 \leq \Phi \leq 1; \tag{3}$$

$$E_P(\Phi|\Pi_i) \equiv \Pi_i, \quad i = 1, 2. \tag{4}$$

Note that (4) only involves the marginal joint distribution P for (Π_1, Π_2), and (3) and (4) are only required to hold with P-probability 1. By Bayes's theorem, if P and $\tilde{P}$ have densities p and $\tilde{p}$ respectively, then

$$\tilde{p}(\pi_1, \pi_2 | A) \equiv \pi_0^{-1} \Phi(\pi_1, \pi_2) p(\pi_1, \pi_2),$$

where $\pi_0 = \tilde{P}(A)$ is the prior probability common to You and the experts; similarly,

$$\tilde{p}(\pi_1, \pi_2 | \bar{A}) \equiv (1 - \pi_0)^{-1} \{1 - \Phi(\pi_1, \pi_2)\} p(\pi_1, \pi_2).$$

Now suppose that a distribution P for (Π_1, Π_2) and a function $\Phi \equiv \Phi(\Pi_1, \Pi_2)$ satisfying (3) and (4) are given. In particular, (4) requires that

$$E_P(\Pi_1) = E_P(\Pi_2) = E_P(\Phi), =: \pi_0, \text{ say.} \tag{5}$$

Together, P and Φ determine a unique distribution $\tilde{P}$ for (Π_1, Π_2, A), having marginal P over (Π_1, Π_2) and with $\tilde{P}(A | \Pi_1, \Pi_2) \equiv \Phi$. Then (4) ensures that $\tilde{P}(A | \Pi_i) \equiv \Pi_i$, $\tilde{P}(A) = \pi_0$. Hence, taking $X_i \equiv \Pi_i$, so that $\Pi_i \equiv \tilde{P}(A | X_i)$, one has a situation in which Π_1 and Π_2 are experts' probabilities having joint distribution P, and combining coherently according to the formula Φ.

We thus see that conditions (3) and (4) on P and Φ are necessary and sufficient for logical consistency. When they are satisfied, we shall say that P and Φ are *compatible*.

We can now raise several general questions.

1. How can we construct compatible pairs (P, Φ)?
2. Given P, how can we characterize all Φ compatible with P?
3. Given Φ, how can we characterize all P compatible with Φ?
4. Is a given Φ (or P) *coherent, i.e.* does there exist any compatible P (or Φ)?

We do not yet have complete solutions to all these problems: this paper constitutes some first steps towards them.

3.1. *Compatible Pairs*

Construction of a compatible pair is straightforward: starting from any joint distribution $\tilde{P}$ for (X_1, X_2, A), define $\Pi_i := \tilde{P}(A|X_i)$ and let P be the marginal distribution for (Π_1, Π_2) and $\Phi := \tilde{P}(A|\Pi_1, \Pi_2)$. Calculation of Φ is especially easy if Π_i is a one-one function of X_i $(i = 1, 2)$, since then

$$\Phi(\pi_1, \pi_2) \equiv \tilde{P}\left\{A \,\middle|\, X_1 = {\Pi_1}^{-1}(\pi_1), X_2 = {\Pi_2}^{-1}(\pi_2)\right\}.$$

In particular, one might have each of X_1 and X_2 taking values in $[0, 1]$. These could be probabilities produced by a forecaster who is not an expert according to our definition. Let $\Psi : [0, 1]^2 \rightarrow [0, 1]$ be an arbitrary combination formula to be applied to (X_1, X_2), and let Q be the joint distribution of (X_1, X_2). One can then extend Q to a joint distribution $\tilde{P}$ for (X_1, X_2, A), on putting $\tilde{P}(A|X_1, X_2) \equiv \Psi(X_1, X_2)$. Let $\Pi_i := \tilde{P}(A|X_i)$, assumed one-one, and let P be the induced distribution for (Π_1, Π_2). Then Π_i is termed the *recalibration* of X_i, since, whereas $X_i \not\equiv \tilde{P}(A|X_i)$ in general, by construction $\Pi_i \equiv \tilde{P}(A|\Pi_i)$. In the frequency calibration interpretation of Section 2.3.1, $\Pi_i(x_i)$ is the proportion of events occurring in those trials for which forecaster E_i quotes a probability of x_i; E_i may not initially be well-calibrated, but will become so after recalibration. One then has

$$\tilde{P}(A|\Pi_1 = \pi_1, \Pi_2 = \pi_2) \equiv \Psi\{{\Pi_1}^{-1}(\pi_1), {\Pi_2}^{-1}(\pi_2)\} =: \Phi(\pi_1, \pi_2)$$

say. Then the distribution P for (Π_1, Π_2), and the combination formula Φ, form a compatible pair.

3.2. *Characterizations*

In general, the problem of characterizing all Φ (or P) compatible with a given P (or Φ) can be difficult – some relevant theory is discussed in Section 5. Here we merely note that such a set, defined by (3) and (4), is *convex*. It might be empty, or contain just one member, or many (see Section 5). It would obviously be useful to identify the extreme points of the convex set, but this too appears difficult in general. The set is typically not a simplex, so not amenable to analysis by methods found useful in other areas (*e.g.* extreme-point models, see Lauritzen, 1980).

4. SOME SPECIAL CASES

In this Section various well known forms for the combination rule Φ are considered, and a (partial) analysis of their compatibility properties is attempted. All these formulae can be expressed in the generalized linear form

$$g(\Phi) \equiv \alpha_1 g(\Pi_1) + \alpha_2 g(\Pi_2) + c \tag{6}$$

where g is a continuous monotonic function.

The linear, harmonic and logarithmic rules are special cases of (6). We conduct a full analysis of the linear rule. For the harmonic and logarithmic rules compatible distributions are given and some properties noted, but we do not have a full characterisation. We also analyse some special cases compatible with independence and conditional independence properties.

4.1. *The Linear Opinion Pool*

The general linear opinion pool has the form

$$\Phi \equiv \alpha_1 \Pi_1 + \alpha_2 \Pi_2 + c. \tag{7}$$

Suppose that this combination formula arises from an overall distribution $\tilde{P}$, with marginal P for the (Π_i). Taking expectations, and noting (5), we find that $c = \alpha_0 \pi_0$, with $\alpha_1 + \alpha_2 + \alpha_0 = 1$ and $\pi_0 := \tilde{P}(A)$, the prior probability. In particular, if $\alpha_1 + \alpha_2 \neq 1$, $\pi_0 = c/(1 - \alpha_1 - \alpha_2)$ is determined by (7). It follows, perhaps surprisingly, that every distribution $\tilde{P}$ compatible with Φ in (7) must assign the same prior probability to A. (If $c/(1 - \alpha_1 - \alpha_2) \notin [0, 1]$, Φ can not be coherent.)

We can thus express (7) as

$$\Phi \equiv \alpha_1 \Pi_1 + \alpha_2 \Pi_2 + \alpha_0 \pi_0. \tag{8}$$

Note that π_0 may be found from (7) as the unique solution of $\Phi(\pi, \pi) = \pi$. In particular we see that, when both experts agree, You too will adopt their common forecast if and only if it is the exactly same as Your prior probability for A.

It is now straightforward to characterize all distributions P compatible with (8). We shall suppose first that α_1, α_2 and $c = \alpha_0 \pi_0$ are all non-zero, with $\alpha_1 + \alpha_2 + \alpha_0 = 1$ and $0 < \pi_0 < 1$; these restrictions will be removed in Section 4.1.1. The condition (4) implies

$$E_P(\Pi_2|\Pi_1) \equiv \lambda\Pi_1 + (1-\lambda)\pi_0 \tag{9}$$
$$E_P(\Pi_1|\Pi_2) \equiv \mu\Pi_2 + (1-\mu)\pi_0, \tag{10}$$

where $\lambda := (1-\alpha_1)/\alpha_2$, $\mu := (1-\alpha_2)/\alpha_1$, *i.e.* one has a linear regression for each Π_i on the other.

Conversely, any joint distribution P on $[0,1]^2$ satisfying (9) and (10) is compatible with Φ in (7), with

$$\alpha_1 := \frac{1-\lambda}{1-\lambda\mu}, \; \alpha_2 := \frac{1-\mu}{1-\lambda\mu}, \; c =: \frac{-(1-\lambda)(1-\mu)\pi_0}{1-\lambda\mu}, \tag{11}$$

so long as it gives probability 1 to the event

$$0 \leq \alpha_1\Pi_1 + \alpha_2\Pi_2 + \alpha_0\pi_0 \leq 1,$$

thus satisfying condition (3).

Not all choices of the coefficients in (7) are coherent. Since, from (9) and (10), $\lambda\mu$ is the squared correlation ρ^2 between Π_1 and Π_2, one must have

$$0 \leq (1-\alpha_1)(1-\alpha_2)/\alpha_1\alpha_2 < 1. \tag{12}$$

In particular, and in contrast with an assumption commonly made for linear opinion pools, it is not coherent for the weights α_1, α_2 and α_0 all to be strictly positive, since then both regression coefficients in (9) and (10) would exceed 1. The fact that all the weights in a linear opinion pool may not be strictly positive was also noted by Genest and Schervish (1985) and Singpurwalla (1988). For a discussion on the interpretation of possibly negative expert weights see Genest and McConway (1990) (however these authors do not use our definition of an expert).

Conversely, when (12) is satisfied with strict inequality, we can indeed find a distribution P for (Π_1, Π_2) compatible with (7), so that formula is then coherent. We need to show that it is possible to satisfy (3), together with (9) and (10). Now (7) will be between 0 and 1 when (Π_1, Π_2) lies in a band between two parallel lines, this band containing the point (π_0, π_0). We can then achieve the regressions (9) and (10) by, for example, taking (Π_1, Π_2) to be uniformly distributed inside a suitably oriented ellipse centred on (π_0, π_0) and totally contained within this band.

Shrinkage. More generally, suppose we have any distribution P on $[0, 1]^2$ with (9) and (10) holding, but which may not give probability 1 to $0 \leq \Phi \leq 1$, where Φ is given by (8) with relation (11) between (α_1, α_2) and (λ, μ). The set

$$A = \{(\pi_1, \pi_2) : \ 0 \leq \alpha_1\pi_1 + \alpha_2\pi_2 + \alpha_0\pi_0 \leq 1\}$$

contains (π_0, π_0). Construct a square S by shrinking each point in $[0, 1]^2$ towards (π_0, π_0) by a constant scale factor, sufficient to ensure that the whole square lies within A. Under this shrinkage map, P is transformed to a distribution P^* on S. It is easy to see that (9) and (10) continue to hold for P^*, and now $0 \leq \Phi \leq 1$ with P^*-probability 1, so that P^* and Φ are compatible.

It may be checked that, among the coherent linear opinion pools, all and only those cases in which exactly one of α_1, α_2 or α_0 is negative can occur, depending on the signs of $1 - \lambda$ and $1 - \mu$. Thus, if $\lambda > 1$, we have $\alpha_1 < 0$, $\alpha_2 > 0$, $\alpha_0 > 0$. The somewhat paradoxical negative weighting on Π_1 can be partly understood by noticing that in this case $E(\Pi_2 - \pi_0|\Pi_1) \equiv \lambda(\Pi_1 - \pi_0)$ with $\lambda > 1$ so that, in a sense, $\Pi_2 - \pi_0$ already incorporates an exaggerated estimate of $\Pi_1 - \pi_0$, which is then corrected when Π_1 is observed.

Now consider the case that (12) holds with equality. Without loss of generality we can take $\alpha_1 = 1$. We then obtain $\lambda = 0$, implying $\rho = 0$. It follows that, for coherence, we must also have $\alpha_2 = 1$ in (10), and thus $\alpha_0 = -1$. Then we obtain the formula

$$\Phi \equiv \Pi_1 + \Pi_2 - \pi_0. \tag{13}$$

A distribution P will compatible with this Φ if and only if:

(i) it gives probability 1 to the event $\pi_0 \leq \Pi_1 + \Pi_2 \leq 1 + \pi_0$, and
(ii) $E(\Pi_1|\Pi_2) \equiv \pi_0$, $E(\Pi_2|\Pi_1) \equiv \pi_0$.

In particular, (ii) will hold when Π_1 and Π_2 are independent (each with mean π_0). This can be arranged (*e.g.* by shrinkage) to hold with the support of (Π_1, Π_2) a rectangle contained in the set for which (i) holds. Consequently, formula (13) is indeed coherent. An intuitive description of (13) is that Π_1 and Π_2 contribute equally to Your final opinion; since they both incorporate the common prior probability π_0, this must be subtracted to correct for double counting.

The following example, due to Mark Schervish, gives an explicit construction of a distribution P compatible with formula (7), in the case when α_1 and α_2 are both positive rational numbers.

Example 1. Binomial model. Let Θ have the beta distribution $Be[a, b]$, and, given Θ, let X_1, X_2 and A be independent, with $X_i \sim B[n_i; \Theta]$, for $i = 1, 2$, and $\mathrm{pr}(A|\Theta) \equiv \Theta$. Define $n_0 := a + b$. Then $\pi_0 = E(\Theta) = a/n_0$. One has

$$\Pi_1 := \mathrm{pr}(A|X_1) \equiv E(\Theta|X_1) \equiv (a + X_1)/(n_0 + n_1).$$

Similarly,

$$\Pi_2 \equiv (a + X_2)/(n_0 + n_2), \quad \Phi \equiv (a + X_1 + X_2)/(n_0 + n_1 + n_2).$$

Thus Φ has the linear form

$$\alpha_1\Pi_1 + \alpha_2\Pi_2 + \alpha_0\pi_0$$

with $\alpha_1 := (n_0 + n_1)/(n_0 + n_1 + n_2)$, $\alpha_2 := (n_0 + n_2)/(n_0 + n_1 + n_2)$, and $\alpha_0 := -n_0/(n_0 + n_1 + n_2)$.

The joint distribution P of (Π_1, Π_2) is discrete, a transformed bivariate beta-binomial distribution. Although $\Phi(0, 0) < 0$ and $\Phi(1, 1) > 1$, P puts all its mass on points for which $0 < \Phi < 1$.

Example 2. Dirichlet model. Let (X_1, X_2, X_3, X_4) jointly have a Dirichlet distribution $D[a, a, b, c]$, and let $\Pi_1 = X_1 + X_3$, $\Pi_2 = X_2 + X_3$, thus defining a joint distribution P for (Π_1, Π_2) with

$$E(\Pi_1) = E(\Pi_2) = (a + b)/(2a + b + c) = \pi_0.$$

By well known properties of the Dirichlet distribution,

$$E(X_3|\Pi_2 = \pi_2) \equiv E(X_3|X_2 + X_3 = \pi_2) \equiv \{b/(a + b)\}\pi_2.$$

Similarly,

$$E(X_1|\Pi_2 = \pi_2) \equiv E(X_1|X_1 + X_4 = 1 - \pi_2) \equiv \{a/(a + c)\}(1 - \pi_2).$$

Hence

$$\begin{aligned} E(\Pi_1|\Pi_2) &\equiv a/(a+c) + \{b/(a+b) - a/(a+c)\}\Pi_2 \\ &\equiv \lambda\Pi_2 + (1-\lambda)\pi_0 \end{aligned}$$

where $\lambda = (bc - a^2)/(a+b)(a+c)$. Similarly

$$E(\Pi_2|\Pi_1) \equiv \lambda\Pi_1 + (1-\lambda)\pi_0.$$

In this case, the corresponding linear Φ, given by (11), is

$$\Phi \equiv (1+\lambda)^{-1}\{\Pi_1 + \Pi_2 - (1-\lambda)\pi_0\}. \tag{14}$$

However the support of P is the whole of $[0,1]^2$, and (for example) $\Phi < 0$ at $(0,0)$, $\Phi > 1$ at $(1,1)$. We therefore apply shrinkage to P, so obtaining a distribution P^* giving probability 1 to (3), and thus compatible with formula (14). Note that the case $a^2 = bc$, *i.e.* $\lambda = 0$, provides an example of a distribution, compatible with (13), where Π_1 and Π_2 are not independent.

4.1.1. *Zero Weights.*

We now relax the assumption that α_1, α_2 and c in (7) are all non-zero. Suppose first $\alpha_1 = 0$ (the case $\alpha_2 = 0$ is similar). Then under any compatible $\tilde{P}$, $\Phi := \tilde{P}(A|\Pi_1, \Pi_2)$ is a function of Π_2 alone, so that $A \perp\!\!\!\perp \Pi_1|\Pi_2$, and so $\Phi \equiv \tilde{P}(A|\Pi_2) \equiv \Pi_2$ almost surely. Thus we must have $\alpha_2 = 1$, $c = 0$ (unless, with $\Phi \equiv \alpha_2\Pi_2 + (1-\alpha_2)\pi_0$, $\Pi_2 = \pi_0$ almost surely, in which case also $\Pi_1 \equiv E(\Phi|\Pi_1) = \pi_0$ almost surely, the case of two entirely uninformative "experts").

The condition $A \perp\!\!\!\perp \Pi_1|\Pi_2$ is equivalent to $\Pi_1 \perp\!\!\!\perp A|\Pi_2$, implying (Dawid, 1979) that Π_2 is a "sufficient statistic" for inference about A based on both Π_1 and Π_2. That is, Π_1 is giving no additional information about A, once Π_2 is given. Clemen (1985) calls such an expert E_1 "extraneous". A distribution P is compatible with this property if and only if $E(\Pi_2|\Pi_1) \equiv \Pi_1$. Note that a distribution with this property might also be compatible with other combination formulae, which do not ignore the information in Π_1. However, the condition $E(\Pi_2|\Pi_1) \equiv \Pi_1$ does imply that expert E_2 is *more refined* than E_1 (DeGroot, 1988), so that, if You have the choice of consulting either E_1 or E_2, but not both, You should always select E_2.

Now consider the case $c = 0$, so that $\Phi \equiv \alpha_1\Pi_1 + \alpha_2\Pi_2$, with $\alpha_1 + \alpha_2 = 1$ for coherence. Having already treated the contrary cases above, we suppose $\alpha_1, \alpha_2 \neq 0$. For compatible P, the condition $E(\Phi|\Pi_1) \equiv \Pi_1$ implies $E(\Pi_2|\Pi_1) \equiv \Pi_1$, and similarly $E(\Pi_1|\Pi_2) \equiv \Pi_2$. But these properties can only hold together when $\Pi_1 \equiv \Pi_2$ almost surely, so that the two experts have identical information, and either of them is "extraneous" when the other is available (which is not, of course, to say that they can both be dispensed with). Then $\Phi \equiv \Pi_1 \equiv \Pi_2$ almost surely.

Much of the literature on the linear opinion pool assumes $\alpha_0 = 0$ in (8) which, we have seen, is not coherent unless at least one expert is extraneous. However the incorporation of a constant term appears more reasonable when we express (8) as

$$(\Phi - \pi_0) = \alpha_1(\Pi_1 - \pi_0) + \alpha_2(\Pi_2 - \pi_0),$$

relating the deviations of Your posterior probability and those of the experts from your common prior probability π_0. Recall, too, that we are assuming a fixed overall probability structure (common to all parties), so that (8) is not required to continue to apply if π_0 changes.

4.2. *The Harmonic Opinion Pool*

Example 3. Negative Binomial model. Let Θ have a $Be[a, b]$ prior distribution and let X_1, X_2 and A be independent, with $X_i \sim NB[k_i; \Theta]$, for $i = 1, 2$, and $\text{pr}(A|\Theta) \equiv \Theta$. Then

$$\pi_0 := E(\Theta) = a/(a+b),$$

$$\Pi_i := P(A|X_i) \equiv E(\Theta|X_i) \equiv (a+k_i)/(a+b+X_i),\ i = 1, 2,$$

and

$$\Phi \equiv (a + k_1 + k_2)/(a + b + X_1 + X_2).$$

Thus Φ has the form of a harmonic opinion pool

$$\Phi^{-1} \equiv \alpha_1{\Pi_1}^{-1} + \alpha_2{\Pi_2}^{-1} + c \tag{15}$$

where $\alpha_1 = (a+k_1)/(a+k_1+k_2)$, $\alpha_2 = (a+k_2)/(a+k_1+k_2)$, and $c = \alpha_0{\pi_0}^{-1}$ with $\alpha_0 = -a/(a + k_1 + k_2)$.

Note again that $\alpha_1 + \alpha_2 + \alpha_0 = 1$, so that $\Phi(\pi_0, \pi_0) = \pi_0$; and that $\alpha_1, \alpha_2, \alpha_0$ are not all positive.

4.3. *Logarithmic Opinion Pool*

The logarithmic opinion pool has the form

$$\begin{aligned} \Phi &= k{\Pi_1}^{\alpha_1}{\Pi_2}^{\alpha_2}{\pi_0}^{\alpha_0}, \\ 1-\Phi &= k(1-\Pi_1)^{\alpha_1}(1-\Pi_2)^{\alpha_2}(1-\pi_0)^{\alpha_0}. \end{aligned} \tag{16}$$

In what follows we express (16) as

$$\mathrm{logit}\Phi = \alpha_1 \mathrm{logit}\Pi_1 + \alpha_2 \mathrm{logit}\Pi_2 + c \tag{17}$$

where $\mathrm{logit}\pi := \log\{\pi/(1-\pi)\}$ and $c = \alpha_0 \mathrm{logit}\pi_0$.

Example 4. Normal Model. Given A or $\bar{A}$, let (X_1, X_2) be bivariate normal with $\mathrm{var}(X_i|A) = \mathrm{var}(X_i|\bar{A}) = 1$,

$$\mathrm{cov}(X_1, X_2|A) = \mathrm{cov}(X_1, X_2|\bar{A}) = \rho \quad (\rho^2 \neq 1),$$

$E(X_i|A) = \delta_i/2$ and $E(X_i|\bar{A}) = -\delta_i/2$. The prior probability π_0 is arbitrary. The normal model was extensively investigated by French (1980, 1981), but without our criterion of expertise.

Let $\Pi_i := \tilde{P}(A|X_i)$, $\Phi := \tilde{P}(A|X_1, X_2)$. Applying Bayes's theorem we find

$$\mathrm{logit}\Pi_i = \mathrm{logit}\pi_0 + \delta_i X_i$$

and

$$\mathrm{logit}\tilde{P}(A|X_1, X_2) = \mathrm{logit}\pi_0 + (1-\rho^2)^{-1}\{(\delta_1 - \rho\delta_2)X_1 + (\delta_2 - \rho\delta_1)X_2\}.$$

Hence $\Phi = \tilde{P}(A|X_1, X_2)$, and (17) holds, with

$$\begin{aligned} \alpha_1 &= (1-\rho\eta)/(1-\rho^2) \\ \alpha_2 &= (1-\rho\eta^{-1})/(1-\rho^2) \end{aligned}$$

where $\eta = \delta_2/\delta_1$ and $\alpha_0 = 1 - \alpha_1 - \alpha_2$.

Note that the weights in the logarithmic opinion pool depend on δ_1 and δ_2 only through η. Under the compatible joint distribution P in this case, $(\mathrm{logit}\Pi_1, \mathrm{logit}\Pi_2)$ has a mixture of two bivariate normal distributions.

Again we have $\alpha_0 + \alpha_1 + \alpha_2 = 1$, and, if $\alpha_0 \neq 0$, formula (17) determines π_0. Note that, since $\alpha_0 = -(1-\rho^2)\alpha_1\alpha_2$, it is again not possible for $\alpha_0, \alpha_1, \alpha_2$ all to be strictly positive.

4.4. *Conditional Independence*

Let $\Pi_i := \tilde{P}(A|X_i),\ i = 1, 2$ and assume that $\Pi_1 \perp\!\!\!\perp \Pi_2|(A, \bar{A})$, that is the two experts' opinions are conditionally independent given A or $\bar{A}$. Then it can be easily shown that

$$\Phi := \tilde{P}(A|\Pi_1, \Pi_2) \equiv \frac{(1-\pi_0)\Pi_1\Pi_2}{(1-\pi_0)\Pi_1\Pi_2 + \pi_0(1-\Pi_1)(1-\Pi_2)} \tag{18}$$

or equivalently

$$\text{logit}\Phi \equiv \text{logit}\Pi_1 + \text{logit}\Pi_2 + c, \tag{19}$$

where $c = -\text{logit}\pi_0$ (so $\pi_0 = (1+e^c)^{-1}$). We thus obtain a special case of (17) with $\alpha_1 = \alpha_2 = 1$, which can be analysed in more detail. The following theorem characterizes all joint distributions compatible with (19).

Theorem 4.1. *A necessary and sufficient condition for a joint density $f(\pi_1, \pi_2)$ to be compatible with Φ given in (19) is that*

$$f(\pi_1, \pi_2) \equiv \left[\frac{(1-\pi_0)\pi_1\pi_2 + \pi_0(1-\pi_1)(1-\pi_2)}{\pi_0(1-\pi_0)}\right] g(\pi_1, \pi_2) \tag{20}$$

where $\pi_0 = (1+e^c)^{-1}$ and g is a density such that

$$E_g(\Pi_1|\Pi_2) \equiv E_g(\Pi_2|\Pi_1) \equiv \pi_0. \tag{21}$$

In this case π_0 is the common expectation of Π_1 and Π_2 under f and, thus, the prior probability of A.

Proof. See Appendix. $\triangleleft$

It is interesting that the condition on g is mathematically identical to that on f needed for compatibility with the combination formula (13) (although the condition $0 \leq \Phi \leq 1$ is automatically satisfied in the present case). The conditional independence construction of (18) is recovered when $\Pi_1 \perp\!\!\!\perp \Pi_2$ under g.

Note that if (20) holds then

$$f(\pi_1, \pi_2|A) \equiv \pi_1\pi_2 g(\pi_1, \pi_2)/\pi_0^2$$

and

$$f(\pi_1, \pi_2|\bar{A}) \equiv (1-\pi_1)(1-\pi_2)g(\pi_1, \pi_2)/(1-\pi_0)^2.$$

4.4.1. *Independence.*

Finally consider the generalised log-linear formula

$$\log(\Phi + a) \equiv \log(\Pi_1 + a) + \log(\Pi_2 + a) + c \qquad (22)$$

for constants a, c. Equivalently,

$$\Phi \equiv e^c(\Pi_1 + a)(\Pi_2 + a) - a. \qquad (23)$$

Again, the compatibility condition (4) is seen to hold if and only if, under P, $E(\Pi_1|\Pi_2) \equiv E(\Pi_2|\Pi_1) \equiv \pi_0$, where in (22) $c = -\log(\pi_0 + a)$, so that $\Phi(\pi_0, \pi_0) = \pi_0$, and π_0 is the prior probability of A. (Formula (13) is a limit of (22) as $|a| \to \infty$). In particular, the case of independence, $\Pi_1 \perp\!\!\!\perp \Pi_2$, with $E(\Pi_1) = E(\Pi_2) = \pi_0$, satisfies (4).

However for condition (3) to hold, (Π_1, Π_2) must be restricted to a subset of the unit square, contained in the region where $0 \leq \Phi \leq 1$. When $\Pi_1 \perp\!\!\!\perp \Pi_2$ the subset must be a rectangle.

For $a = 0$, $\Phi = \Pi_1\Pi_2/\pi_0$. Thus any distribution for which Π_1 and Π_2 are independent, $E(\Pi_i) = \pi_0$, and $s_1 s_2 \leq \pi_0$, where $s_i = wesssup\Pi_i$, is compatible with $\Phi = \Pi_1\Pi_2/\pi_0$. Note that in this case $\Pi_1 \perp\!\!\!\perp \Pi_2|A$ (but *not* $\Pi_1 \perp\!\!\!\perp \Pi_2|\bar{A}$).

Similarly, for $a = -1$ we get $1 - \Phi = (1 - \Pi_1)(1 - \Pi_2)/(1 - \pi_0)$, equivalent to the case $a = 0$ on interchanging A and $\bar{A}$.

4.5. *Some Conjectures*

In all our variants of (6), we have found that, in fact,

$$g(\Phi) = \alpha_1 g(\Pi_1) + \alpha_2 g(\Pi_2) + \alpha_0 g(\pi_0), \qquad (24)$$

where $\alpha_0 + \alpha_1 + \alpha_2 = 1$ and $\pi_0 = \tilde{P}(A) = E_P(\Phi)$ for any P compatible with Φ; further, not all the α_i's can be positive. This has been fully verified in the linear case (7), and holds for (15) for the construction of Section 4.2. Likewise for (17) under the normal construction, and for (19), by Theorem 4.1, with $\alpha_1 = \alpha_2 = 1, \alpha_0 = -1$. For (22), equivalent to (23), we have shown that, for compatible P, $E_P(\Pi_2|\Pi_1) \equiv \pi_0 = e^{-c} - a$, so that $\pi_0 = E_P(\Pi_2) = E_P(\Phi)$, and again (24) holds with $\alpha_1 = \alpha_2 = 1, \alpha_0 = -1$.

So long as (24) holds with $\alpha_0 + \alpha_1 + \alpha_2 = 1$ and $\alpha_0 \neq 0$, π_0 is the unique solution of $\Phi(\pi, \pi) = \pi$. In particular, although various different distributions P for (Π_1, Π_2) may be compatible with such a Φ, all of them have the same value for $\pi_0 = \tilde{P}(A)$. That is, the form of Φ uniquely determines π_0. It is interesting to speculate how general this result might be.

For a general combination formula Φ, let

$$\mathcal{A} = \{\pi \in (0,1) : \Phi(\pi, \pi) = \pi\}.$$

For any $\pi \in \mathcal{A}$ we can find a distribution P, compatible with Φ, having $\tilde{P}(A) = E_P(\Phi) = \pi$: simply take P to put probability one on the point (π, π) (the conditions (3) and (4) are clearly satisfied). We may conjecture that, when Φ is continuous, then $\pi_0 = E_P(\Phi) \in \mathcal{A}$, for any P compatible with Φ, and such that $P((\Pi_1, \Pi_2) = (0,0) \text{ or } (1,1)) = 0$. If so, then π_0 for such P will be uniquely determined by the form of Φ when $\mathcal{A}$ has a single element. Also, for the generalized linear form (6), we can then deduce the desired property $c = \alpha_0 g(\pi_0)$, with $\alpha_0 + \alpha_1 + \alpha_2 = 1$, and hence determine π_0 whenever $c \neq 0$.

When $c = 0$ in (6), the analysis conducted in Section 4.1.1 for the linear opinion pool shows that at least one of the experts must be extraneous. This appears to hold more generally. Thus, for the construction of Section 4.3, we have $\alpha_0 = -(1 - \rho^2)\alpha_1\alpha_2$. Thus (for $\rho^2 < 1$) if $\alpha_0 = 0$, we have either $\alpha_1 = 0$ or $\alpha_2 = 0$. If $\alpha_1 = 0$ then $\Phi \equiv \Pi_2$, as in the linear case, and similarly if $\alpha_2 = 0$. In these cases $\mathcal{A} = (0,1)$, and any value for π_0 is possible.

The case $\rho^2 = 1$ is a little more complicated. Changing the sign of X_1 and δ_1 if necessary, we may assume $\rho = 1$. If $\delta_1 \neq \delta_2$, formula (17) does not hold: in fact, for $\rho = 1$, we obtain $\Phi = 1$ or 0, according as $X_1 - X_2 = (\delta_1 - \delta_2)/2$ or $(\delta_2 - \delta_1)/2$, the only two possible values, where $X_i = (\text{logit}\Pi_i - \text{logit}\pi_0)/\delta_i$. But if also $\delta_1 = \delta_2$, then $\Phi \equiv \Pi_1 \equiv \Pi_2$ almost surely, and both experts are extraneous (and (17) holds with $\alpha_0 = 0$ and any α_1, α_2 such that $\alpha_1 + \alpha_2 = 1$). Again in this degenerate case any value of π_0 is possible.

In these examples of (6), we can thus only have $c = 0$ when, under any compatible P, at least one expert is extraneous. This is an unexpected and somewhat unsettling property, for which we do not have a convincing explanation. Nevertheless we are led to conjecture that it may be a general property of any formula of the generalised linear form (6). We further conjecture that, as in all our examples, whenever (24) holds it is not possible for all of α_0, α_1 and α_2 to be strictly positive.

5. GENERAL THEORY

In this Section the compatibility of a combination formula Φ and a joint distribution P for (Π_1, Π_2) is studied, by relating this problem to that of existence of probability measures with given marginals.

Theorem 5.1. *Let P be a joint distribution for (Π_1, Π_2) with $E(\Pi_1) = E(\Pi_2) = \pi_0$; and let $\Phi : [0,1]^2 \to [0,1]$ be a combination formula. Define a finite measure Q by $dQ(\pi_1, \pi_2) := \Phi(\pi_1, \pi_2) dP(\pi_1, \pi_2)$ (thus when P is absolutely continuous with density p, Q has density $q(\pi_1, \pi_2) := \Phi(\pi_1, \pi_2)p(\pi_1, \pi_2)$.) Let P_i, Q_i be the marginals for Π_i under P and Q respectively. Then Φ and P are compatible if and only if $dQ_i(\pi_i) \equiv \pi_i dP_i(\pi_i) =: dP_i^*(\pi_i)$ say (or, when marginal densities exist, $q_i(\pi_i) \equiv \pi_i p_i(\pi_i) =: p_i^*(\pi_i)$, $i = 1, 2$).*

Proof. See Appendix. ◁

Given P, Theorem 5.1 thus reduces the problem of finding a compatible Φ to that of characterizing measures Q having specified marginals (P_i^*), and such that $dQ/dP \leq 1$. This has been studied by Kellerer (1961) and Gutmann *et al.* (1991).

5.1. *Coherence of Joint Forecast Distribution*

If Π_1 and Π_2 are both produced by "experts", then one should not expect them to be wildly different. For example, it would seem paradoxical if, with Π_1 say uniform on $[0, 1]$, one always had $\Pi_2 = 1 - \Pi_1$. This suggests that not all joint distributions on $[0,1]^2$ for (Π_1, Π_2) are coherent.

If, for a joint distribution P on $[0,1]^2$, there exists a compatible Φ, then we can construct the associated joint distribution $\tilde{P}$ for (Π_1, Π_2, A), and then $\tilde{P}(A|\Pi_i) \equiv \Pi_i$; so that P is coherent; and conversely. So coherence of P is equivalent to the existence of a compatible Φ, or equivalently, by Theorem 5.1, to the existence of a finite measure Q on $[0,1]^2$ with $dQ/dP \leq 1$ and given margins (P_i^*). A necessary and sufficient condition for this is (see, for example, Strassen, 1965, Theorem 6): for all measurable $D, B \subset [0,1]$,

$$P_1^*(D) - P_2^*(B) \leq P(D \times \bar{B}) \tag{25}$$

or, when densities exist,

$$\int_D \pi_1 p_1(\pi_1) d\pi_1 - \int_B \pi_2 p_2(\pi_2) d\pi_2 \leq \int_{D \times \bar{B}} p(\pi_1, \pi_2) d\pi_1 d\pi_2. \quad (26)$$

Note that on replacing B by $\bar{B}$, and using $\int_0^1 \pi dP_1 = \int_0^1 \pi dP_2 = \pi_0$, (25) becomes

$$\int_D \pi dP_1 + \int_B \pi dP_2 \leq \pi_0 + P(D \times B) \quad (27)$$

The following Theorem shows the incoherence of any distribution for which, for some $0 \leq c_1, c_2 \leq 1$, $\Pi_1 - c_1$ and $\Pi_2 - c_2$ always have opposite signs.

Theorem 5.2. *Suppose* $P(\Pi_1 \leq c_1 \cap \Pi_2 \leq c_2) = P(\Pi_1 \geq c_1 \cap \Pi_2 \geq c_2) = 0$. *Suppose further that* $P(\Pi_j > c_j) > 0$, *where* $c_j = \max(c_1, c_2)$. *Then* P *is not coherent.*

Proof. See Appendix. ◁

Gutmann *et al.* (1991) point out that, whenever P is absolutely continuous, and there exists a measure Q with $dQ/dP \leq 1$ and given marginals, then there exists such a measure with dQ/dP taking values 0 and 1 only. The following result then follows from Theorem 5.1.

Corollary 5.1. *Suppose* P *is an absolutely continuous coherent joint distribution for* (Π_1, Π_2). *Then there exists a compatible combination formula* Φ *taking values* 0 *and* 1 *only.*

This result is somewhat surprising. It implies that for any such coherent distribution of experts' opinions (Π_1, Π_2), it is logically consistent that the combination of these opinions could deliver absolute subjective certainty as to whether the event A holds or not.

Another result of Gutmann *et al.* (1991) shows that, when Π_1 and Π_2 are independent, there exists a compatible Φ which is non-decreasing in each argument (but it might not, under this additional restriction, be possible to ensure that this Φ takes values in $\{0, 1\}$ only).

We further note the following application of Example 2 of Gutmann *et al.* (1991). Suppose Π_1 and Π_2 are distributed, independently, uniformly on [0,1]. Then the unique combination formula compatible with this joint distribution is

$$\Phi = \begin{cases} 1 & \text{if } \Pi_1 + \Pi_2 > 1 \\ 0 & \text{if } \Pi_1 + \Pi_2 \leq 1. \end{cases}$$

6. MORE THAN TWO EXPERTS

We will now consider some extensions of the previous results when there are more than two experts.

For $k > 2$ the linear opinion pool is

$$\Phi \equiv \sum_{i=1}^{k} \alpha_i \Pi_i + \alpha_0 \pi_0$$

with $\sum_{i=0}^{k} \alpha_i = 1$. From the definition of an expert one requires that, under a compatible P, $E(\Phi|\Pi_i) \equiv \Pi_i$, that is

$$\sum_{j \geq 1, j \neq i} \alpha_i E(\Pi_j|\Pi_i) \equiv (1 - \alpha_i)\Pi_i - \alpha_0 \pi_0.$$

In general, one cannot deduce that the regression of Π_i on the remaining Π's is linear. However, one can impose the stronger requirement on P and Φ that the induced coherent combination formula for *any* subset of the Π's be linear, and then deduce that the regression of any Π_i on any subset of the remaining Π's must be linear. For example, if $k = 3$ we have $E(\Phi|\Pi_1, \Pi_2) \equiv \alpha_1\Pi_1 + \alpha_2\Pi_2 + \alpha_3 E(\Pi_3|\Pi_1, \Pi_2) + \alpha_0\pi_0$, and this is linear if $E(\Pi_3|\Pi_1, \Pi_2)$ is; the linear regression of $E(\Pi_1|\Pi_2)$ then follows from the case $k = 2$.

Most of the examples extend quite trivially when $k > 2$. For Example 1, let $X_i \sim B[n_i; \Theta], i = 1, \cdots, k$. One obtains the linear opinion pool

$$\Phi \equiv \sum_{i=1}^{k} \alpha_i \Pi_i + \alpha_0 \pi_0$$

where $\alpha_i = (n_0 + n_i)/(n_0 + \sum_{i=1}^{k} n_i)$, $\alpha_0 = -n_0/(n_0 + \sum_{i=1}^{k} n_i)$.

For Example 2, let $(X_1, \cdots, X_k, X_{k+1}, X_{k+2})$ have a Dirichlet distribution $D[\overbrace{a, \cdots, a}^{k}, b, c]$ and let $\Pi_j := X_j + X_{k+1}, j = 1, \cdots, k$. One can easily show that $E(\Pi_j) = (a+b)/(ka+b+c) = \pi_0$, $E(X_j|\Pi_1) \equiv (1-\Pi_1)a/\{(k-1)a+c\}$, $E(X_{k+1}|\Pi_1) \equiv \Pi_1 b/(a+b)$. Hence

$$E(\Pi_j|\Pi_1) \equiv \{b/(a+b)\}\Pi_1 + [a/\{(k-1)a+c)\}](1-\Pi_1)$$

and, after shrinkage, one obtains a distribution compatible with the linear opinion pool

$$\Phi \equiv [1/\{1+(k-1)r\}] \left\{ \sum_{i=1}^{k} \Pi_i - (k-1)(1-r)\pi_0 \right\}$$

where $r := \{b/(a+b)\} - a/\{(k-1)a+c\}$.

Example 3 is likewise easily extendible to $k > 2$. In all these examples, it again turns out that the α_i's cannot all be strictly positive. This again could be conjectured to be true in general.

Consider now, extending Section 4.4, the case of $k > 2$ expert opinions, conditionally independent given A or $\bar{A}$. One can show that

$$\begin{aligned} \Phi &:= P(A|\Pi_1, \cdots \Pi_k) \\ &\equiv \frac{(1-\pi_0)^{k-1} \prod_{i=1}^{k} \Pi_i}{(1-\pi_0)^{k-1} \prod_{i=1}^{k} \Pi_i + \pi_0{}^{k-1} \prod_{i=1}^{k} (1-\Pi_i)} . \end{aligned} \tag{29}$$

The necessary and sufficient condition for the joint density $f(\pi)$ to be compatible with Φ of (29) is now that

$$\begin{aligned} f(\pi_1, \cdots, \pi_k) \propto & \left[\left(\prod_{i=1}^{k} \pi_i \right) / \pi_0^{k-1} + \right. \\ & \left. \left\{ \prod_{i=1}^{k} (1-\pi_i) \right\} / (1-\pi_0)^{k-1} \right] g(\pi_1, \cdots, \pi_k) \end{aligned}$$

where g is a density such that, for $j = 1, \cdots, k$,

$$\begin{aligned} & E_g \left[\left(\prod_{i \neq j} \Pi_i \right) / \pi_0^{k-1} \middle| \Pi_j \right] \\ & \equiv E_g \left[\left\{ \prod_{i \neq j} (1-\Pi_i) \right\} / (1-\pi_0)^{k-1} \middle| \Pi_j \right]. \end{aligned}$$

In the generalization of (23) to the case of $k > 2$ independent experts, we have the combination formula

$$\Phi \equiv \left\{ \prod_{i=1}^{k} (\Pi_i + a) \right\} / (\pi_0 + a)^{k-1} - a. \tag{30}$$

A distribution P is compatible with (30) if and only if, for $j = 1, \cdots, k$, $E[\prod_{i \neq j}(\Pi_i + a) | \Pi_j] \equiv (\pi_0 + a)^{k-1}$, and P gives probability 1 to $0 \leq \Phi \leq 1$. It is no longer immediately evident, however, that these conditions require $E(\Phi) = \pi_0$.

Theorem 5.1 and Corollary 5.1 extend straightforwardly to the case $k > 2$. However, with more than two experts, there is no simple counterpart of the existence condition (25).

7. DISCUSSION

The purpose of this paper has been to examine coherence properties of various methods for combining opinions expressed as probabilities for some fixed uncertain event A, by relating the combination rules to an overall joint distribution, for all relevant unknown quantities, which we have supposed common to You and the experts. More generally, it would be desirable to extend the analysis to the case where the experts report their probabilities for a number of events, in particular where their opinions take the form of full probability distributions for some uncertain quantity θ of interest. This would require investigation of coherence properties for a joint probability distribution encompassing both θ and the experts' distributions for θ. While various workers have developed models for such joint distributions (*e.g.* Lindley, 1985), the implications of coherence, taken together with our definition of expertise, have not been considered.

Our basic assumption, that You and the experts would agree about your probabilities when you have the same information, could be criticized as too strong in general. One way in which it might be achieved is by interaction between You and the experts at the initial point of time. DeGroot (1974) described an iterative interaction process leading to convergence of individual opinions, when the individuals update their opinions in the light of the others' opinions. While it would be difficult to apply this formally to the problem considered here, since the distributions to be shared would need to apply to the experts' own future opinions, or at least to the variables which they will be based on, it does suggest that our basic assumption should be reasonable in at least some contexts.

Most of the literature on combining opinions uses axiomatic properties or modelling assumptions to derive particular pooling recipes. Compared with these, our assumptions are not so restrictive; furthermore, and completely generally, our analysis offers valuable guidance for assessing pooling formulae that have been suggested from other approaches.

ACKNOWLEDGEMENTS

We are grateful to Robert Clemen, Steffen Lauritzen and Robert Winkler for valuable comments and suggestions. This research was supported by grants from MURST and CNR (Italy) and SERC (UK).

REFERENCES

Berger, J. O. and Mortera, J. (1991). Bayesian limited communication. *J. Statist. Planning and Inference* **28**, 1–24.

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

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University Press.

Clemen, R. T. (1985). Extraneous expert information. *J. Forecasting* **4**, 329–348.

Clemen, R. T. (1986). Calibration and the aggregation of probabilities. *Manag. Sci.* **32**, 312–314.

Clemen, R. T. and Murphy, A. H. (1986). Objective and subjective precipitation probability forecasts: statistical analysis of some interrelationships. *Weather and Forecasting* **1**, 56–65.

Dawid, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. B* **41**, 1–31, (with discussion).

Dawid, A. P. (1982). The well-calibrated Bayesian. *J. Amer. Statist. Assoc.* **77**, 605–610.

Dawid, A. P. (1986). Probability forecasting. *Encyclopedia of the Statistical Sciences* (S. Kotz, N. L. Johnson and C. B. Read, eds.), vol. **7**. New York: Wiley, 210–218.

DeGroot, M. H. (1974). Reaching a consensus. *J. Amer. Statist. Assoc.* **69**, 118–121.

DeGroot, M. H. (1988). A Bayesian view of assessing uncertainty and comparing expert opinion. *J. Statist. Planning and Inference* **20**, 295–306.

DeGroot, M. H. and Eriksson, E. A. (1985). Probability forecasting, stochastic dominance and the Lorenz curve. *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, 99–118, (with discussion).

DeGroot, M. H. and Fienberg, S. E. (1983). The comparison and evaluation of forecasters. *The Statistician* **32**, 12–22.

DeGroot, M. H. and Mortera, J. (1991). Optimal linear opinion pools. *Manag. Sci.* **37**, 546–558.

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

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

French, S. (1985). Group consensus probability distributions: a critical survey. *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, 183–201, (with discussion).

French, S. (1986). Calibration and the expert problem. *Manag. Sci.* **32**, 315–320.

Genest, C. (1984). A characterization theorem for externally bayesian groups. *Ann. Statist.* **12**, 1100–1105.

Genest, C. and McConway K. J. (1990). Allocating the weights in the linear opinion pool. *J. Forecasting* **9**, 53–73.

Genest, C., McConway K. J. and Schervish, M. J. (1986). Characterization of externally Bayesian pooling operators. *Ann. Statist.* **14**, 487–501.

Genest, C. and Schervish, M. J. (1985). Modelling expert judgments for Bayesian updating. *Ann. Statist.* **13**, 1198–1212.

Genest, C. and Zidek, J. V. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.* **1**, 114–148.

Gutmann, S., Kemperman, J. H. B., Reeds, J. A. and Shepp, L. A. (1991). Existence of probability measures with given marginals. *Annals of Prob.* **19**, 1781–1797.

Jeffreys, H. (1939). *Theory of Probability*. Oxford: Clarendon Press.

Kellerer, H. G. (1961). Funktionen und Produkträumen mit vorgegeben Marginal-Funktionen. *Math. Ann.* **144**, 323–344.

Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.

Lauritzen, S. L. (1980). *Statistical Models as Extremal Families*. Aalborg: University Press.

Lindley, D. V. (1982). The improvement of probability judgements. *J. Roy. Statist. Soc. B* **145**, 117–126.

Lindley, D. V. (1985). Reconciliation of discrete probability distributions. *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, 375–390, (with discussion).

Lindley, D. V. (1986). Another look at an axiomatic approach to expert resolution. *Manag. Sci.* **32**, 303–306.

McConway, K. J. (1981). Marginalization and linear opinion pools. *J. Amer. Statist. Assoc.* **76**, 410–414.

Madansky, A. (1964). Externally Bayesian groups. RAND Memo RM-4141-PR. Santa Monica, CA: The Rand Corporation.

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

Murphy, A. H. and Winkler R. L. (1992) Diagnostic verification of probability forecasts. *Int. J. Forecasting* **7**, 435–455.
Schervish, M. J. (1986). Comments on some axioms for combining expert judgments. *Manag. Sci.* **32**, 306–312.
Singpurwalla, N. D. (1988). Discussion of "Gaining weight: A Bayesian Approach" by Bayarri, M. J. and DeGroot, M. H. *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, 39–42.
Stone, M. (1961). The opinion pool. *Ann. Math. Statist.* **32**, 1339–1342.
Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.
Winkler, R. L. (1981). Combining probability distributions from dependent information sources. *Manag. Sci.* **27**, 479–488.
Winkler, R. L. (1986). Expert resolution. *Manag. Sci.* **32**, 298–303.

APPENDIX

Proof of Theorem 4.1.

Note that with definition (18), $0 \leq \Phi(\Pi_1, \Pi_2) \leq 1$ for all (Π_1, Π_2) and (20) is equivalent to

$$f(\pi_1, \pi_2) \equiv \pi_1 \pi_2 g(\pi_1, \pi_2) / \left\{\pi_0 \Phi(\pi_1, \pi_2)\right\}.$$

Suppose (20) and (21) hold. Then

$$\begin{aligned} f_1(\pi_1) &\equiv g_1(\pi_1)\left\{\frac{\pi_1}{\pi_0}\int \pi_2 g(\pi_2|\pi_1)d\pi_2 + \right. \\ &\qquad \left. \frac{1-\pi_1}{1-\pi_0}\int (1-\pi_2)g(\pi_2|\pi_1)d\pi_2\right\} \\ &\equiv g_1(\pi_1)\left[\pi_1{\pi_0}^{-1}E_g(\Pi_2|\pi_1) + \right. \\ &\qquad \left. (1-\pi_1)(1-\pi_0)^{-1}\left\{1 - E_g(\Pi_2|\pi_1)\right\}\right] \\ &\equiv g_1(\pi_1). \end{aligned}$$

Similarly $f_2(\pi_2) \equiv g_2(\pi_2)$. Thus

$$\begin{aligned} E(\Phi|\Pi_2 = \pi_2) &\equiv \int \Phi(\pi_1, \pi_2) f(\pi_1|\pi_2) d\pi_1 \\ &\equiv \int \Phi(\pi_1, \pi_2) f(\pi_1, \pi_2)\{g_2(\pi_2)\}^{-1} d\pi_1 \\ &\equiv \pi_2 {\pi_0}^{-1} \int \pi_1 g(\pi_1|\pi_2) d\pi_1 \\ &\equiv \pi_2 \end{aligned}$$

and similarly $E(\Phi|\Pi_1) \equiv \Pi_1$. Hence $f(\pi_1, \pi_2)$ is compatible with Φ. Conversely suppose Φ and $f(\pi_1, \pi_2)$ are compatible and define $g(\pi_1, \pi_2)$ by (20). Then

$$\int \Phi(\pi_1, \pi_2) f(\pi_1, \pi_2) d\pi_1 \equiv \pi_2 f_2(\pi_2)$$

so $\int \pi_1 g(\pi_1, \pi_2) d\pi_1 \equiv \pi_0 f_2(\pi_2)$. Similarly

$$\int (1 - \pi_1) g(\pi_1, \pi_2) d\pi_1 \equiv (1 - \pi_0) f_2(\pi_2),$$

thus $g_2(\pi_2) \equiv f_2(\pi_2)$. Hence

$$E_g(\Pi_1|\Pi_2) \equiv \pi_0, \quad E_g(\Pi_2|\Pi_1) \equiv \pi_0.$$

Further,

$$E_f(\Pi_1) = E_{f_1}(\Pi_1) = E_{g_1}(\Pi_1) = E_{g_2}\{E_g(\Pi_1|\Pi_2)\} = \pi_0,$$

and similarly $E_f(\Pi_2) = \pi_0$. $\triangleleft$

Proof of Theorem 5.1.

Suppose Φ and P are compatible. Consider the unique joint distribution $\tilde{P}$ for (Π_1, Π_2, A), with marginal P for (Π_1, Π_2) and with $\tilde{P}(A|\Pi_1, \Pi_2) := \Phi(\Pi_1, \Pi_2)$. Then $\tilde{P}(A|\Pi_i) \equiv \Pi_i$ and $\tilde{P}(A) = \int \Phi dP = \pi_0$.

By Bayes's Theorem, under $\tilde{P}$ the distribution of (Π_1, Π_2) given A is just Q/π_0, which is thus a probability distribution; and then Q_i/π_0 is the distribution of $\Pi_i|A$, so that, again by Bayes's Theorem, one has ${\pi_0}^{-1} dQ_i(\pi_i) \equiv \pi_0^{-1} \pi_i dP(\pi_i)$, and hence $Q_i = P_i^*$.

Conversely, suppose $Q \ll P$ is a finite measure having marginals $Q_i = P_i^*$, and with $dQ/dP \equiv \Phi$. Then, for any measurable function k of Π_1,

$$\begin{aligned} E_P\{\Phi k(\Pi_1)\} &= \int k(\pi_1)(dQ/dP)dP = \int k(\pi_1) dQ \\ &= \int k(\pi_1) dQ_1 = \int k(\pi_1) dP_1^* \\ &= \int \pi_1 k(\pi_1) dP_1 = E_P\{\Pi_1 k(\Pi_1)\}. \end{aligned}$$

Hence, $E_P(\Phi|\Pi_1) \equiv \Pi_1$, and similarly $E_P(\Phi|\Pi_2) \equiv \Pi_2$, which establishes compatibility. $\triangleleft$

Proof of Theorem 5.2

Suppose P is coherent. Applying (27) first with $B = [0, c_1], D = [0, c_2]$ and then with $B = [c_1, 1], D = [c_2, 1]$, we have

$$\int_0^{c_1} \pi_1 dP_1 + \int_0^{c_2} \pi_2 dP_2 \leq \pi_0 \tag{31}$$

$$\int_{c_1}^1 \pi_1 dP_1 + \int_{c_2}^1 \pi_2 dP_2 \leq \pi_0 \tag{32}$$

where, since $P(\Pi_i = c_i) = 0, i = 1, 2,$ the integrals may equally include or exclude these end-points. Since the left hand sides of (31) and (32) sum to $2\pi_0$, we must have equality in both. So $\int_{c_1}^1 \pi_1 dP_1 + \int_{c_2}^1 \pi_2 dP_2 = \pi_0$. But $\int_0^{c_1} \pi_1 dP_1 + \int_{c_1}^1 \pi_1 dP_1 = \pi_0$. Hence $\int_0^{c_1} \pi_1 dP_1 = \int_{c_2}^1 \pi_2 dP_2$. Without loss of generality suppose $c_2 \geq c_1$, so that $P(\Pi_2 > c_2) > 0$; and note that $P(\Pi_1 < c_1) = P(\Pi_2 > c_2)$. We then have

$$\int_0^{c_1} \pi_1 dP_1 < c_1 P(\Pi_1 < c_1), \quad \int_{c_2}^1 \pi_2 dP_2 > c_2 P(\Pi_2 > c_2),$$

a contradiction. ◁

DISCUSSION

R. COOKE *(Technishe Universiteit Delft, The Netherlands)*

The fundamental assumption of this article is that everyone has the same prior. As someone dealing with quantitative expert probabilities on a daily basis, I cannot recall a real situation in which this assumption is plausible. In working with practical Bayesian models for expert judgement in which this assumption is not made, I have run up against a problem and I would be grateful if the authors could propose a solution.

Suppose an expert gives invertible cdf's F_1 and F_2 for random variables X_1 and X_2. Let P denote the decision maker's (DM's) probability, and let G_i be DM's (invertible) cdf's for X_i, all this conditional on hearing F_1 and F_2. A Bayesian model should show DM how to update his distribution after observing X_1. The problem is quite general but, for simplicity, let us assume that the events $\{F_1(X_1) \leq r\}$ and $\{F_2(X_2) \leq r\}$ are exchangeable for every r. This entails,

$$P\{F_1(X_1) \leq r\} = P\{F_2(X_2) \leq r\}$$

so that

$$P\{F_1(X_1) \leq r\} = G_1\{F_1^{-1}(r)\} = G_2\{F_2^{-1}(r)\}.$$

It follows that $G_2 = G_1 F_1^{-1} F_2$. Now suppose a second expert has been consulted as well, giving invertible cdf's H_1 and H_2 (conditionalization is now on hearing the advice of both experts). If the above exchangeability assumption held for the second expert as well, then the same reasoning would apply and we should conclude that

$$F_1^{-1} F_2 = H_1^{-1} H_2.$$

If this relation does not hold, and in practice it will not, then exchangeability is not a feasible assumption for both experts. Can someone give an updating scheme which is generally applicable?

S. FRENCH *(University of Leeds, UK)*

Early in their paper, the authors refer to a need for a *shared world view*. To follow the detail of their exposition fully, one needs to share their world view; and alas I do not. So, perhaps, they will forgive me if I concentrate on their opening section and explain where my world view differs from theirs. I should also say that this is not intended as a dismissal of their work. Far from it. Reading their ideas opened a new perspective on the expert problem for me, one that helped me understand more clearly why I am most comfortable with approaches in which the decision maker's beliefs are modelled with probabilities, which he can use in guiding his thinking and his choice of action, whereas experts' statements should be modelled as data, even if they are articulated in the language of probability.

The history of group decision making and the combination of opinion shows that one enters the land of paradoxes immediately one treats all participants in a symmetrical way. Arrow's Theorem shows that to combine weak orders representing beliefs or preferences in an entirely symmetrical manner leads to the contradiction of sensible principles of rationality or fairness. (Arrow, 1951; Cooke, 1993; Dalkey, 1972; Kelly, 1972). Only when one takes a *constructive* approach to subjective expected utility (French, 1994), does it seem possible to avoid difficulties. In a constructive approach the viewpoint taken is internal to a decision maker (You) and the models are constructed to help the decision maker organise, understand and evolve his beliefs and preferences. An expert is external to the decision maker and, while one can help the decision maker model his beliefs about her [1] statements, one avoids modelling her internal beliefs. Thus I am immediately uncomfortable with part (ii) of the basic assumption in this paper: namely, that the decision maker and expert share a common subjective distribution, $\tilde{P}$. I wonder whether the authors have examined the comments of Cooke (1993) which identify some difficulties with exchangeability when there is a single probability distribution owned by the decision maker and the expert(s).

As the authors show, this assumption of a common $\tilde{P}$ implies that for the decision maker the expert is probability calibrated. Of course, *vice versa*, it implies that for the expert the decision maker is probability calibrated. It also implies that the decision maker believes that when they base their judgements on the same information they share the same frequency calibration curve. Thus in matters of belief and information processing, the decision maker and expert are clones of each other. Are we entering a *Brave New World* in which Betas only take advice from Betas and not Alphas or Gammas? But ignoring that rhetorical question, let me focus on why the assumption that the decision maker and expert share a common world view renders this theory inapplicable in some very important circumstances.

More and more evidence is accumulating of the danger of ignoring modelling error (see, e.g., Draper, 1995; Harper *et al.*, 1994). If one lives within a model and forgets the step back to reality in the interpretation of an analysis, one risks underestimating the uncertainty in one's forecasts. Moreover, since no physical model is a perfect reflection of reality, there can be considerable advantages in combining forecasts based upon very different physical models. For instance, a decision maker who is

[1] The decision maker will be referred to in the masculine and the expert in the feminine.

offered two economic forecasts, one based upon Keyesian thinking and the other on monetarist thinking, might be well advised to weight them together in some way (Draper, 1995; Makridikis and Winkler, 1983). Recently a joint CEC and USNRC has shown the advantage of combining atmospheric dispersion-deposition models both in terms of accuracy and in making realistic assessment of modelling error (Harper *et al.*, 1994). In this latter study forecasts based upon different models of atmospheric physics (i.e., distinctly different world views) were combined. Such combinations, the ones that in practice seem to bring the greatest benefit, are outside the scope of the authors' approach.

Despite all my negative comments, I enjoyed and valued this paper.

C. GENEST *(Université Laval, Canada)* and
M. J. SCHERVISH *(Carnegie Mellon University, USA)*

We are grateful for the privilege to comment on this paper authored by three fine statisticians with whom we have had opportunities to reflect on the issue of the combination of expert opinions on various occasions in the past. We share the preoccupation of Dawid, DeGroot and Mortera (DDM) with coherent methods of aggregation and were not surprised to discover many similarities between their approach to the problem and our own attempt to wrestle with it, ten years ago (Genest and Schervish, 1985).

In the simplest situation we considered at the time, You specify

$$\Pr(A) = \pi_0 \text{ and } \mathrm{E}(\Pi) = \mu, \tag{1}$$

where $\Pi = (\Pi_1, \ldots, \Pi_k)$ is a vector whose components represent the subjective probability for event A, as reported by $k \geq 1$ experts. Then You look for combination rules $\Phi(\Pi)$ that can be expressed in the form $\Pr(A|\Pi = \pi)$ for some joint distribution for (Π, A) that satisfies (1). As shown in Theorem 3.2 of our paper (whose statement contains a typographical error), $\Phi(\Pi)$ must be of the form

$$\Phi(\Pi) = \pi_0 + \sum_{i=1}^{k} \lambda_i(\pi_i - \mu_i) \tag{2}$$

in order to be compatible with every distribution for Π satisfying (1). Here, the λ_i's are scalars to be selected by You so that $\Phi(\Pi)$ lies between

0 and 1. Thus they must satisfy 2^{n+1} (possibly redundant) inequalities of the type

$$\max\left(\sum_{i=1}^{k}\lambda_i\frac{\mu_i}{\pi_0},\sum_{i=1}^{k}\lambda_i\frac{1-\mu_i}{1-\pi_0}\right)\leq 1.$$

The components of the vector $\lambda = (\lambda_1, \ldots, \lambda_k)$ may actually be interpreted in terms of the coefficients of a linear regression of A on Π, through the relation $\mathrm{Cov}(A, \Pi) = \mathrm{Var}(\Pi)\lambda$. This implies that some of the λ_i's might be negative, but not that at least one of them is necessarily negative.

What explains this discrepancy between our result and DDM's finding that linear opinion pools are incoherent if their weights are all positive? Their conclusion stems from two properties of the distribution of Π, namely that (i) the expected value of the vector Π is necessarily of the form $\pi_0 1$, where 1 is a k-dimensional vector of ones; and that (ii) the conditional mean of each Π_i given another Π_j must be linear in Π_j. These two properties are consequences of DDM's definition of expert, which is more restrictive than ours. By limiting the class of people who can be called experts, DDM achieve stronger results, but at what cost? In order for someone to be an expert in their sense, two conditions must hold:

1. You and the expert must have had the same opinions about all relevant data at some time in the past, and
2. the expert must have observed all the data that You have observed and possibly more.

Are these conditions plausible in some circumstances? For example, suppose that the event A concerns some rare medical condition about which You have never heard. Perhaps You will consider a specialist to have all information available to You and more. You might even believe that, at some time in the distant past, this specialist knew as little as You do about the condition. At that time, the specialist's opinion about all relevant information might possibly have been the same as Yours is now. We believe, however, that such a case is rare. Surely, You know something about Yourself that a medical specialist does not know or will not treat with the same importance as You do.

These considerations lead rather naturally to the question of what You should do if You regard Yourself as an expert. At the end of Section 2.2, DDM claim that if You are party to information that another expert does not have, then You could treat Yourself as one of the experts and apply their results. This comment is sufficiently intriguing that we wish to pursue it here. You are asked to imagine a time in the past (say T_0) at which You would have accepted the expert's opinion as Your own, had it (and nothing else) become available. However, between T_0 and now, You have obtained some information that You do not believe the expert to have acquired. In the spirit of DDM's discussion at the end of Section 2.1, we impose no conditions on the form of the additional information that You (but not the expert) have obtained. Let Π stand for the current opinion of the expert. Does the problem, as described, place any restrictions on Your joint distribution for Π and A? We think not. To support our position, we offer the following result, whose proof is given at the end of the discussion.

Proposition. Every joint distribution for (Π, A) is consistent with the assumption that, at some earlier time T_0, You would have accepted the expert's opinion as Your own, but that between then and now, You have learned some additional information not available to the expert.

Although it is stated in the one-expert case, this result remains valid for multiple-expert situations. In short, it says that if You are willing to assume that You have some information that is unavailable to the experts, and if You put no restrictions on how that information might have arrived, then the results described by DDM are no longer relevant to You, in Your *current* state of information. This is because at present, the experts need no longer be experts in the sense of DDM, since You have information that they do not have. In fact, Your current state of information is not at all constrained by the assumption that, in the past, the other experts satisfied DDM's definition. In this case, however, You could still apply Theorem 3.2 of Genest and Schervish (1985), since it does not rely on a restrictive concept of expertise. If You believe now that $\mathrm{E}(\Pi_i) = \mu_i$, this theorem implies that (2) is the *only* combination formula that is consistent with every marginal distribution of Π having mean $(\mu_1, \ldots, \mu_k)$.

The situation is somewhat different if You assume that the experts are conditionally independent given A and its complement. In that case, DDM arrive at a logarithmic opinion pool of the form (18) or (29) in their paper. By comparison, an unrestrictive definition of expert leads to Theorem 4.1 of Genest and Schervish (1985) (whose statement also contains a typographical error!). The ensuing logarithmic pool may be written as

$$\Phi(\Pi) = \frac{\pi_0^{1-k} \prod_{i=1}^{k} \overline{\Pi}_i}{\pi_0^{1-k} \prod_{i=1}^{k} \overline{\Pi}_i + (1-\pi_0)^{1-k} \prod_{i=1}^{k} (1-\overline{\Pi}_i)},$$

with $\overline{\Pi}_i = \pi_0 + \lambda_i(\Pi_i - \mu_i)$, $\mu_i = \mathrm{E}(\Pi_i)$, and scalars λ_i's to be chosen by You in such a way that

$$\max\{\pi_0/(\mu_i - 1), (\pi_0 - 1)/\mu_i\} \leq \lambda_i \leq$$

$$\min\{\pi_0/\mu_i, (1-\pi_0)/(1-\mu_i)\}, \quad 1 \leq i \leq n.$$

This formula has the nice property that, if You only learn some of the experts' opinions, the resulting combination rule retains the same form. In particular, if only one Π_i is revealed to You, then

$$\Pr(A|\Pi_i) = \overline{\Pi}_i = \pi_0 + \lambda_i(\Pi_i - \mu_i).$$

If in addition λ_i is set equal to one (which forces $\mu_i = \pi_0$), one gets DDM's condition $\Pr(A|\Pi_i) = \Pi_i$. Conversely, their definition of expert implies $\lambda_i = 1$, and hence $\mu_i = \pi_0$. So, the assumption that the experts are conditionally independent given A and its complement brings DDM's results and ours closer together. That is, their characterization of the logarithmic opinion pool is the special case of ours in which $\lambda_1 = \cdots = \lambda_k = 1$ and no further conditions need to be assumed. This is in contrast to the linear opinion pool in which their result reduces to ours with $\mu_1 = \cdots = \mu_k = \pi_0$ *and* the additional requirement that the conditional mean of each Π_i given another Π_j is linear in Π_j. Of course, conditional independence is itself an assumption (made both by DDM and by us), and quite a strong one at that.

Although the present discussion is critical of some of DDM's findings, we would not want to leave the readers with the impression that we find no merits to their work. Nothing would be further from the truth. In particular, we are grateful to DDM for articulating some of the connections between coherence issues in the combination of probabilistic opinions and compatibility requirements for the existence of distributions with fixed multivariate marginals. In the light of their paper, it now seems obvious that modern developments in the latter area could be profitably applied to expert use, as Clemen and Jouini (1996) have recently begun to explore. DDM are to be congratulated for describing one promising line of attack, and we hope that the publication of their paper will help rekindle the statistical community's interest in these issues.

Proof of the proposition. Let $f(\pi, a)$ stand for the joint density of the pair (Π, A) with respect to some σ-finite measure, and identify A with its indicator function. In order for f to be Your current density, there must have been some data X with conditional density $h(x|\pi, a)$ given $(\Pi, A) = (\pi, a)$, and a joint density $g(\pi, a)$ that You held at time T_0 such that

(i) $g(\pi, 1)/[g(\pi, 1) + g(\pi, 0)] = \pi$ for all π;
(ii) $f(\pi, a) \propto g(\pi, a)h(x|\pi, a)$ as a function of (π, a).

Define $g(\pi, a) = \pi^a(1-\pi)^{1-a}$, so that condition (i) holds. Now, introduce a random pair $X = (X_1, X_2)$ whose components are conditionally independent given (Π, A). Assume that X_1 has a uniform distribution on $[0, 1]$ given $(\Pi, A) = (\pi, a)$ if $f(\pi, a) > 0$, and that it has a uniform distribution on $[2, 3]$ given $(\Pi, A) = (\pi, a)$ if $f(\pi, a) = 0$. Let the conditional distribution of X_2 given $(\Pi, A) = (\pi, a)$ be normal with mean 0 and standard deviation $g(\pi, a)/f(\pi, a)$ if $f(\pi, a) > 0$, and let the conditional distribution be uniform on the interval $[2, 3]$ if $f(\pi, a) = 0$. Then condition (ii) is satisfied whenever $(X_1, X_2) = (1/2, 0)$ is observed. Since we are not required to assume any particular form for the type of information that only You (and not the expert) have, the proof is complete.

D. V. LINDLEY *(Somerset, UK)*

This excellent paper is a real advance in our understanding of what constitutes expert opinion and how it should be used. I was once of the view that experts, in the restricted sense in which the term is used here, were uncommon because of a 'strange' property that you have of them. The present paper convinces me that this view is too extreme and that the restricted sense is valuable. The 'strange' property flows from Bayes's theorem

$$o(A|\Pi) = \frac{p(\Pi|A)}{p(\Pi|\overline{A})}o(A), \qquad (1)$$

where o means your odds on and p your probability. Thus $p(\Pi|A)$ is your probability that the expert will announce Π for her probability of A, were A to be true. By definition, you believe the expert, so the left-hand side is $\Pi/(1-\Pi)$. Hence so is the right-hand side and therefore the likelihood ratio is

$$\frac{p(\Pi|A)}{p(\Pi|\overline{A})} = \frac{\frac{\Pi}{(1-\Pi)}}{o(A)} . \tag{2}$$

This analysis effectively treats Π as you would any other data. (2) says that the likelihood ratio, on the left, depends on the prior odds, on the right. This is the 'strange' feature referred to above. It is 'strange' because standard practice is to fix the likelihoods and consider the prior separately.

To pursue this further, consider what happens when your prior changes. Presumably this is because you have received some additional information. If so, the expert ceases to be an expert, since it is a requirement of expertise that "the expert has all the information you have". So the only way to retain the expertise is for you to share this new information with them. If this is done, it is perfectly reasonable for your likelihood to change, since it now involves $p(\Pi|A, I)$, where I is the new information, and not $p(\Pi|A)$. I can therefore more comfortably accept the definition of an expert, though it does demonstrate that experts need to be treated with care. Do not keep anything from them.

The analysis becomes more complicated when two experts are consulted, the situation that occupies the bulk of the paper. Let them give their opinions in sequence, Π_1 and then Π_2. The question then arises as to whether the second informant is an expert. In a sense, no, because you know Π_1 and therefore have information that the second informant does not. They are experts in isolation but not in sequence. One way to retain the expertise is to tell the second what the first has said, before you receive Π_2. Again, no secrets from experts.

With two experts, Bayes's theorem says

$$o(A|\Pi_1, \Pi_2) = \frac{p(\Pi_1|A)p(\Pi_2|\Pi_1, A)}{p(\Pi_1|\overline{A})p(\Pi_2|\Pi_1, \overline{A})} o(A) \tag{3}$$

$$= \frac{\Pi_1}{(1-\Pi_1)} \frac{p(\Pi_2|\Pi_1, A)}{p(\Pi_2|\Pi_1, \overline{A})} \tag{4}$$

because the first informant is an expert, equation (2). If the second is told Π_1 before pronouncing, then the expertise condition presumably obtains, in the sense that the second informant "has all the information you have". (2) may therefore be applied again, with the result

$$o(A|\Pi_1, \Pi_2) = \frac{\frac{\Pi_1}{(1-\Pi_1)} \frac{\Pi_2}{(1-\Pi_2)}}{o(A|\Pi_1)} \,.$$

But $o(A|\Pi_1) = [\Pi_1/(1 - \Pi_1)]$, so that finally

$$o(A|\Pi_1, \Pi_2) = \frac{\Pi_2}{(1 - \Pi_2)} \,.$$

We have here a solution to the problem of combining expert opinion; namely tell the second expert what the first has said and then accept the second's opinion, ignoring the first. On reflection, this is most sensible; you let the second expert do the combination for you since, after all, he or she is an expert and can do it as well as you. (Notice that it will not matter whether you provide Π_1 or the information that led to it, since just as you consider the informant an expert, the informant thinks you are an expert, as you both share $\tilde{P}$.) Of course, sometimes it will not be possible to establish communication between the two experts and resort must be made to some combination. But expert 2 can do the combining, why can't you? Notice that if you do not communicate Π_1 to the second expert, and if you judge Π_1 and Π_2 independent, given A and also given $\overline{A}$, then from (4), applying (2) again and taking logs

$$\text{logit}\Phi = \text{logit}\Pi_1 + \text{logit}\Pi_2 - \text{logit}\Pi_0$$

as in Section 4.3. This has the curious consequence that even though you accept each separately, you do not accept what they both say, even when they agree, $\Pi_1 = \Pi_2$, unless the common value is π_0.

My own view is that the definition of an expert given here is too restrictive. My preference is to speak of an informant who provides Π, which is treated, like any other data, by Bayes, (1). The informant's expertise for you is expressed through your likelihood ratio. Thus if you thought they were good, then $p(\Pi|A)$ would centre around a large value of Π, certainly in $(1/2, 1)$, whereas $p(\Pi|\overline{A})$ around a low value in $(0, 1/2)$. With two or more informants, the joint likelihood ratio does the same job as in (3). I remain somewhat unhappy about going direct to the combination rule, rather than deducing it from the proper use of the probability calculus. Invention of a rule smacks of adhockery. There is a good precept than I learnt from de Finetti: think about your probabilities; but then don't think any more, merely become a computer and apply the rules of the probability calculus, and nothing else.

K. J. McCONWAY *(The Open University, UK)*

1. *Introduction.* Dawid, DeGroot and Mortera (hereafter DDM) deal with a crucial real-world problem. The world is full of experts telling us what to think and how to act. (Some of these experts are even statisticians.) DDM describe, and to a considerable extent elucidate, a particular model of expert knowledge and of how a decision maker should use it. The theory they present is in itself fascinating. But is it useful? To what extent might the kind of experts they describe be found in the real world?

2. *Do 'experts' exist?.* DDM's *basic assumption* about experts stems from remarks of DeGroot (1988), quoted by DDM in Subsection 2.1, about the use of another person's stated probability and about the availability of information. DDM recognize that DeGroot's statements about probability and about information cannot be equivalent unless further assumptions are made, and they provide such an assumption; that everyone involved has a common subjective probability distribution for everything of interest. The clarification that this brings is very welcome, but it is important to realize just how restrictive it is. It leaves no room, for example, for the decision maker to learn anything about the state of the world from the fact that an expert chose to observe one particular random variable rather than another. It leaves no room for honest subjective disagreement, where two subjectivists have access to identical information but nevertheless state different probabilities. DDM themselves point out (Section 1) that their assumption is more in accord with a logical view of probability than a subjective one; they justify the use of their assumption to a subjectivist by pointing out the rashness of taking into account in a naive way the views of someone 'whose world view was at odds with Yours.' Their assumption is that Your world view and the expert's are exactly in accord; 'at odds with' sounds too strong for views that differ slightly from their assumed equality.

DDM note (Section 7) that their basic assumption might indeed be taken to be too strong in general, but propose achieving it by some sort of consensus-seeking process, carried out before extra information is observed. While it is possible that such a process might indeed validate their assumption, perhaps it limits its real-world validity even more. Under what circumstances would a group of experts and a decision maker be able to co-operate in this possibly time-consuming way at the start of the process, but not interact at all (except through reports to the decision maker) later on? French (1985) pointed out the importance of the distinction between the 'expert problem', where a panel of experts advise a decision maker individually, and the 'group decision problem', where the group are jointly responsible for what goes on. It seems inappropriate in most circumstances to use what is fundamentally a description of group interaction (DeGroot 1974) at one stage in the process, and then to use a model for the 'expert problem' in the rest.

To summarize, my view is that 'experts' in the sense of DDM rarely if ever exist in practical situations. However, this is far from saying that this work is of no practical value, for three reasons. First, as pointed out in Subsection 2.3, the same modelling approach fits other situations; in my view, it may turn out to be more directly useful in the situation of conflicting reference sets than in dealing with the opinions of real experts. Secondly, the work throws important lights on the whole process of using expert opinion, which go beyond the specific assumptions made. Thirdly, every method of attacking this problem yet proposed makes restrictive and probably unrealistic assumptions; there seems to be no other way of getting to grips with it.

3. *Recalibration and the linear opinion pool.* In Subsection 3.1, DDM explain how, under certain circumstances, their method for constructing a compatible pair can be applied, for example, to probabilities produced by forecasters who are not 'experts' on their definition. This is done by recalibrating each of the stated probabilities X_i to $\Pi_i := \tilde{P}(A|X_i)$. Using this construction, one can show, for instance, that if You choose to combine the reports X_1 and X_2 of two non-'expert' forecasters using a linear opinion pool $\Psi(X_1, X_2) := \beta_1 X_1 + \beta_2 X_2$ ($\beta_1 + \beta_2 = 1$) which does not attach weight to Your own view, and if X_1 and X_2 are considered independent, then the Π_i, the recalibrations of the X_i, are linear functions of the X_i. Furthermore, the combination formula for

the recalibrated probabilities is

$$\Phi(\Pi_1, \Pi_2) = \Pi_1 + \Pi_2 - \pi_0.$$

This demonstrates a situation in which one can begin with a linear opinion pool of the kind considered sometimes in the past (no negative weights, zero weight for Your probability), and end up after recalibration with a pool including Your own probability π_0, and indeed giving it negative weight.

A related converse problem is the following. Under what circumstances is it true that the combination formula Ψ for uncalibrated 'probabilities' is a linear opinion pool, and the combination formula Φ, constructed in the manner described in Subsection 3.1 of DDM is also a linear opinion pool as defined in Equation (7) of DDM? My preliminary investigations indicate that, when the opinion pool Ψ gives zero weight to Your own probability, then the only possible form for the opinion pool Φ is $\Phi(\Pi_1, \Pi_2) = \Pi_1 + \Pi_2 - \pi_0$. If this is indeed the case, what forms of pooling operator Ψ and recalibration can lead to other weights in the opinion pool Φ? More generally, under what circumstances is it helpful to think of a pooling operator for uncalibrated probabilities in terms of recalibration and a pooling operator for 'experts' in the DDM sense?

4. *Conclusion* In conclusion, though this paper may not directly provide a practical model for dealing with the opinions of real-world experts, it provides important insights and concepts for that process. It also emphasizes yet again the key role that Professor DeGroot played in this area over such a long period. I am honoured to have been asked to discuss it.

R. L. WINKLER *(Duke University, USA)*

Dawid, DeGroot, and Mortera (hereafter D/DG/M) are to be congratulated on an interesting, thought-provoking paper. I have read it several times with new insights each time.

The issues of what constitutes "expertise" and what might be reasonably assumed to be common knowledge are crucial elements of the D/DG/M structure. In their development, expertise rests on the notion of a "shared world-view." One implication of this notion, that you are extraneous when a single expert is available, is perfectly reasonable. But the extension of the shared world-view to imply that at some point, you and the expert can be viewed as having a common subjective distribution about all future events and quantities of interest, is an awfully strong assumption. It is hard to imagine this scenario without going back to a point before the expert acquired any of her expertise, in which case the common starting point is based on a diffuse state of information with respect to the substantive questions of interest. After all, expertise is not just having access to relevant data, but being able to understand and interpret that data. Without going back to a state where everyone is diffuse, can you realistically share the same grand probability distribution without being an expert yourself? Also, keeping in mind that assumptions of common probabilities restrict the generality of many results in game theory and economics, we should be cautious in interpreting results based on such assumptions.

A key question, to my mind, is the role of the prior. Strictly speaking, in a Bayesian formulation the prior is certainly relevant, and the prior seems to play a major role in many of the developments in D/DG/M. In a practical sense, however, should the prior be so important in this context? If any prior information you have will be overwhelmed by the judgments of an expert, then your state of prior information is effectively diffuse. Once the expert has given you her probability, you will immediately forget about whatever (weak) judgments you may have held before obtaining the expert's probability. With more than one expert, your prior information is likely to be relevant only to the extent that it will help you sort out any relationships among the sets of information (hence among the probabilities) of the experts. (See Clemen (1987) for an example in which one expert is extraneous when compared with either of two other experts individually but not when all three experts are considered together.) But once again, if these are truly experts, your meagre information about A, as reflected by π_0, shouldn't be of much help. I therefore get nervous when I find that π_0 seems to play more than a trivial role in the combination of experts' opinions.

For the above reasons, discussions of the combination of experts' opinions in the literature have typically used (explicitly or implicitly) diffuse priors and focused on combinations of the experts' probabilities, not of the experts' probabilities and your prior probability. If your prior information is non-diffuse, then it is easier to treat yourself as another expert (e.g., see Winkler (1981), p. 481, and Section 2.2 of D/DG/M).

The typical dependence among experts is a contributing factor here, since it is easier to model the overall dependence structure among the new pieces of information (including your own probability as one of the new pieces of information) than to think about the dependence structure only for the k experts and then to have to consider separately the dependence between your prior and the new information. This explains why, for instance, the linear opinion pool generally does not include a term involving π_0. When a constant term is included, it is usually to correct for systematic bias, not to bring in π_0.

As a concrete example, suppose that I am interested in the probability of rain tomorrow and that the only prior information I have is climatology, the historical long-run proportion of days with rain at the location of interest for this time of year. Thus, my π_0 is simply equal to climatology. Then I see a probability assessed by a weather forecaster with considerable experience at this location and in this season. Surely the weather forecaster knows the value of climatology and takes that into account, along with all sorts of current weather information, in coming up with her probability. My π_0 is now totally irrelevant. If I were using a linear opinion pool, I would give myself a weight of zero and the weather forecaster a weight of one (possibly after recalibrating the weather forecaster if I judge her to be miscalibrated). With probabilities from two forecasters, I would give myself a weight of zero and would give weights to the two forecasters based on my judgments about their relative accuracy and the degree of dependence between them.

Why does my result with two experts differ from D/DG/M in that π_0 is not included in the combination rule? Because of my diffuse state of information. In terms of the binomial model in Section 4.1 of D/DG/M, I am saying that n_0 is zero for all practical purposes. If having an improper prior for θ is unappealing, let n_0 get arbitrarily close to zero with $\pi_0 = a/n_0$ held constant. If n_0 is not zero, then I would expect the term with π_0 to have a negative coefficient, as in D/DG/M, to allow for the fact that this information is already being included twice, once in each expert's probability. For the same reason (multiple-counting of my prior information), shouldn't the numerator of α_0 be $-(k-1)n_0$ instead of just $-n_0$ in the extension of the binomial example to k experts in Section 6? When n_0 is zero, of course, the term involving π_0 drops out because double- (or multiple-) counting "no information" is not a problem. As an aside, negative weights can also be traced to high dependence, but that is not the issue here.

It is worth noting that things seem to work differently in a model such as the above binomial model, which is hierarchical in the sense that the experts and I have probability distributions for the probability of interest, than in a more basic non-hierarchical model in which we assess our probabilities for A and do not consider "probabilities of probabilities." These different ways of modelling were discussed in Winkler (1986) and the accompanying series of papers in *Management Science*.

One quibble I have with D/DG/M is that I would have liked to see more examples, interpretation, and intuition to supplement the mathematical results. This is particularly true when the results seem somewhat counterintuitive. In some such cases, it is not too difficult to sort out what is happening. For example, consider the result in Section 4.1.1 that if $c = 0$ and both experts get non-zero weights, then the two experts must agree on their probabilities almost surely. One explanation here is that under D/DG/M's assumptions, a negative c is needed to avoid double-counting unless the two experts have identical information and there is no double-counting. This is a direct consequence of D/DG/M's definition of a shared world-view. (See Clemen and Winkler (1990) for further thoughts on situations in which experts give identical probabilities.)

In another case, the result that if there exists a compatible combination rule for any joint distribution P, then there exists one which predicts with certainty whether A will occur or not, I wonder if D/DG/M could provide a bit more insight? Under what conditions will this particular rule hold, and are those conditions likely to occur in practice? More generally, do seemingly pathological results such as this indicate that coherence as defined by D/DG/M is a very weak condition that is consistent with non-plausible combination rules, that there is something inherently faulty with the basic assumptions (e.g., common probabilities), or that my intuition that labels the results as pathological is itself faulty?

In conclusion, despite my concerns about the basic assumptions in D/DG/M, it is always valuable to look at a familiar problem (in this case, the combination of experts' opinions) from a new perspective (D/DG/M's coherence). The main results of the paper, showing probability models and combination rules that are compatible with each other given

D/DG/M's assumptions, are certainly of interest. It would be nice to see them supplemented with some speculation on which models and rules seem reasonable and on implications for practice. Beyond this, the paper can add to our understanding of the combination of experts' opinions by stimulating further thought about notions of expertise and coherence and their implications.

REPLY TO THE DISCUSSION

(by A. P. Dawid and J. Mortera)

A common theme of much of the discussion is a feeling of discomfort over the relevance and usefulness of our definition of "expert". We can strongly sympathize with this, and in no way would we wish to suggest that our analysis makes other approaches redundant. We claim merely that a study of the implications of our definition casts valuable light on the problem of combining expert opinions. In particular, a point McConway takes up, given any combination formula for "non-expert" probabilities (X_1, X_2), together with a joint distribution for those probabilities, we can always convert them to "expert" probabilities (Π_1, Π_2) by recalibration, as in our §3.1, and thus obtain a new, coherent, combination formula. Such constructions deserve further analysis. However, McConway's specific conjecture, that a linear recalibration of a linear "non-expert" combination formula will always have the form $\Phi \equiv \Pi_1 + \Pi_2 - \pi_0$, is easily seen to be false. Suppose we start with $\Psi \equiv \beta_1 X_1 + \beta_2 X_2$, and the joint distribution of (X_1, X_2) has linear regressions of each X_i on the other. Then $\Pi_i \equiv E(\Psi | X_i)$ is linear in X_i, and substituting for X_i in terms of Π_i into the original combination formula Ψ leads to a new, coherent, "expert" linear combination formula $\Phi(\Pi_1, \Pi_2)$, now generally with a constant term, in which the coefficients of Π_1 and Π_2 are unconstrained.

French correctly points out the desirability of combining diverse judgments that may not reflect a shared world-view. Application of recalibration, to obtain new assessments which now behave (technically at least) like expert judgments in our sense, demonstrates that he is incorrect in asserting that such problems are outside the scope of our approach.

Also closely related to the recalibration construction is the formula for the general logarithmic pool for conditionally independent experts, given by Genest and Schervish. In this case, $\bar{\Pi}_i$ is just the recalibration of the "non-expert" Π_i, and their formula then follows immediately from our equation (29).

Genest and Schervish claim that we obtain stronger results than theirs by imposing stronger conditions. This is not clearly so. They are looking for a combination formula which will simultaneously be compatible (according to their interpretation) with every one of a large class of joint distributions, constrained only to have given means. This requirement appears extremely strong, and we are not clear when or why one might wish to impose it.

Genest and Schervish also argue that, when You have data that the experts don't, coherence puts no constraints on Your combination formula. This is true, but only in the same sense that, in ordinary Bayesian analysis, coherence puts no constraints on Your posterior distribution given the data to hand. All coherence does is relate what You would do or believe under various different circumstances, all but one of which must be counter-factual. It is thus still worthwhile to try to go back in time to before You got Your additional data, when Your own posterior probability was still random for You, and jointly distributed with those of the other experts. (Assessing this joint distribution now is logically similar to assessing a prior, after seeing the data.) The theory of our paper then applies to delimit Your choice of (prior) combination formula. You might choose one such formula by considering its behaviour for various hypothetical data that You and/or the experts might obtain. Then, when You are happy with it, You can plug in Your own new probability based on the actual data to hand, thus obtaining the "coherent" new combination formula for incorporating the other experts' probabilities. The procedure runs exactly parallel to the way a thoughtful Bayesian might use coherence to help her decide what to believe after seeing the data.

Lindley contrasts our analysis with the more common paradigm in which the informant's opinion is treated as data, and subjected to a fairly standard Bayesian analysis. However, we find such an approach, popular though it is, too facile. Is it really sensible to consider Your "sampling model" for Π given A or $\bar{A}$ as an "objective" ingredient, which can be specified independently of Your prior probability $P(A)$? Surely, if You think A very likely, that in itself should lead You to expect that so too does the informant, and it would be reasonable to bias Your conditional distributions for Π more towards higher values. (Similar considerations cast doubt on the realism of the model of conditional independence of the informants). What Lindley terms a "strange property" of our approach is in fact entirely reasonable. It is only "strange" when referred to an inappropriate paradigm whose wide use in this area appears to have no better justification than an uncritical analogy with something familiar in a totally distinct context. An additional implementational difficulty of this "Bayesian" approach is the need to understand and assess the conditional distributions of Π given A and $\bar{A}$. How realistic are the strong model assumptions that are commonly made in order to assess these conditional distributions? How is one to assess the interdependence among the experts? It is all too easy to write down and play with arbitrary mathematical formulae for these conditional distributions, but we consider the true psychological effort involved in any real-world assessment to be totally unfamiliar and mind-boggling. Frankly, it seems to us far more natural to introspect directly about $P(A|\Pi)$. This can then be recalibrated, if necessary, and the approach of our paper applied to the problem of finding combination rules.

Lindley remarks on the "curious" property of our formula (19) (which is shared with the various other formulae we consider): that even when both experts give You the same probability, each of which by itself You would adopt, Your coherent probability based on hearing both experts will differ from their common value, unless that is the same as the prior probability π_0. This does not seem to us at all curious, as the following argument shows. Except in degenerate cases, the experts must have seen different data. If expert 1 has seen data that justify an increase in the probability of A from, say, 0.5 to 0.75, and expert 2 has seen different data with the same effect, then You might feel justified in concluding that the overall effect of all the data (an effect You are trying to reproduce in so far as it is possible on the basis of the experts' reports) should be to raise the probability of A still further. See, however, the discussion on shrinkage on page 276 for a case in which a different conclusion is appropriate.

Lindley also suggests that a simple solution for combining two experts' opinions is to tell the second expert the first expert's opinion and then adopt the second expert's updated opinion, letting her do the combination for You. However, this does not resolve the problem, it merely displaces it. The second expert still needs a coherent method for combining the first expert's opinion with her own. The coherent methodology we have proposed can be used by whoever is required to perform the combination.

Cooke describes a problem of his own making, in a different set-up from ours, and asks for our comments. We find it hard to appreciate why he should be so concerned to discover that it is possible to make a collection of assessments that are inconsistent with each other. Is he saying that he does not wish his freedom to be constrained by the mathematical theory of probability? French refers to Cooke (1993), which we have examined, but we cannot reconcile it with French's description of it, so that we do not know what "difficulties" we are being asked to address.

Winkler claims that the only satisfactory use for our approach is when the initial state of information is "diffuse". However, this concept plays no rôle whatsoever in our theory, and we do not even know how to interpret it clearly within our general framework. But if we posit that we all start out at birth as equally blank Bayesian babies, and learn coherently from whatever life throws at us, we might have a scenario which links his ideas with ours.

Winkler is mistaken in claiming that "the prior seems to play a major rôle in many of the developments". The prior probability cannot play any rôle at all, because, in our framework, it is just a fixed number; something that cannot vary cannot have any influence on anything. True, that number, π_0, does come up in various formulae, but this should not be overinterpreted. Winkler's intuition, to combine two different expert forecasters using a linear opinion pool with no constant term is, as we have shown, misguided. A constant term is required, and the fact that it turns out to be related to the prior probability should be neither surprising nor worrying.

Winkler regards the possibility of a formula which combines two uncertain judgments to obtain certainty as "pathological". It is not, although it is perhaps, a little artificial. Suppose that expert i observes X_i ($i = 1, 2$). Let A be an event of the form "$(X_1, X_2) \in S$", determined by (X_1, X_2) jointly, but not by either singly. Then, assuming $\Pi_i \equiv P(A|X_i)$ is an invertible function of X_i, we have exactly the "pathological" case described. We also remind Winkler that, except for special cases such as that at the end of our §5, the 0–1 formula associated with a given distribution for (Π_1, Π_2) will only be one of a wide variety of compatible combination formulae.

In conclusion, we should like heartily to thank José Bernardo for organising this discussion on our paper, and all the discussants for contributing to it. In doing so, they honour the memory of a wonderful friend and colleague, Morrie DeGroot. Had he been with us today, this rejoinder would certainly have been far more interesting and insightful.

ADDITIONAL REFERENCES IN THE DISCUSSION

Arrow, K. J. (1951). *Social Choice and Individual Values.* New York: Wiley.

Clemen, R. T. (1987). Combining overlapping information. *Manag. Sci.* **33**, 373–380.

Clemen, R. T. and Winkler, R. L. (1990). Unanimity and compromise among probability forecasters. *Manag. Sci.* **36**, 767–779.

Clemen, R. T. and Jouini, M. N. (1996). Copula models for aggregating expert opinions. *Operations Research* **44**, (to appear).

Cooke, R. M. (1993). The ill advised Bayesian. *ISBA Newsletter* **2**, 2–3.

Dalkey, N. C. (1972). An impossibility theorem for group probability functions. *Tech. Rep.* **P-4862**. Santa Monica, CA: RAND Corporation.

Draper, D. (1995). Assessment and propagation of model uncertainty. *J. Roy. Statist. Soc. B* **57**, 45–97, (with discussion).

French, S. (1994). Utility: probability's younger twin. *Aspects of Uncertainty: a Tribute to D. V. Lindley* (P. R. Freeman, and A. F. M. Smith, eds.). Chichester: Wiley, 171–180.

Harper, F. T.; Goossens, L. H. J.; Cooke,R. M.; Helton, J. C.; Hora, S. C.; Jones, J. A.; Kraan, B. C. P.; Lui, C.; McKay, M. D.; Miller, L. A.; Päsler Sauer, J. and Young, M. L. (1994). *Joint USNRC/CEC Consequence Uncertainty Study: Summary of Objectives, Approach, Application and Results for the Dispersion and Deposition Uncertainty Assessments.* NUREG/CR-6244, EUR 15855, SAND94-1453. Sandia National Laboratories and Delft University of Technology.

Kelly, F. S. (1978). *Arrow Impossibility Theorems.* New York: Academic Press.

Makridakis, S. and Winkler, R. L. (1983). Average of forecasts: some empirical results. *Manag. Sci.* **29**, 987–996.