Statistical Science
1986, Vol. 1, No. 1, 114–148

Combining Probability Distributions: A Critique and an Annotated Bibliography

Christian Genest and James V. Zidek

Abstract. This paper addresses the problem of aggregating a number of expert opinions which have been expressed in some numerical form in order to reflect individual uncertainty vis-à-vis a quantity of interest. The primary focus is consensus belief formation and expert use, although some relevant aspects of group decision making are also reviewed. A taxonomy of solutions is presented which serves as the framework for a survey of recent theoretical developments in the area. A number of current research directions are mentioned and an extensive, current annotated bibliography is included.

Key words and phrases: Bayesian inference, consensus, expert opinions, group decisions, pooling, reconciliation.

1. INTRODUCTION

The objective of this paper is to provide an overview and an extensive, annotated bibliography on the normative aspects of the formation of an aggregate opinion in the context of group interaction, consensus belief emergence, or managerial expert use. In any of these contexts, the word "opinion" will refer to an arbitrary collection of numerical statements expressing an individual's degrees of belief about the world. Encompassed by this definition are such things as prior, vague prior, posterior, structural, or fiducial distributions, as well as odds ratios, finitely additive probability measures, and belief functions. Preference-based opinions are not admitted, however, as they would be in the treatment of the group decision problem.

Philosophical concerns partly explain the interest in the problem of combining opinions. The frequency theory of statistics and its attendant concept of repeated sampling reflect and make operational the notion of objectivity which is enshrined in the scientific method. Strictly speaking, an experiment is never repeatable; there are only experiments which are deemed to be governed by the same probabilities. To rescue the ideal of objectivity, therefore, it is necessary to call upon the law of large numbers to show how interexperiment variations will "average out" in the long run. Nevertheless, there are circumstances in which the interpretation of frequencies is stretched to the limits of plausibility. For example, the probability of failure of a structure like a nuclear or hydroelectrical power-generating facility presents greater difficulty of interpretation than the probability of a head in the toss of a newly minted coin. Moreover, the frequency theory fails to deal adequately with increasingly common practical problems in which data is sparse, unavailable, or subject to nonsampling errors. Retrospective studies, for example, often present serious difficulties in this respect.

These difficulties have led to the vigorous development of the Bayesian theory. Ignoring practical problems of implementation which are the object of current research, the Bayesian program would seem to be entirely satisfactory as a normative theory for the individual. However, groups of individuals are left stranded; no concept equivalent to the classical notion of objectivity is available to them.

Weerahandi and Zidek (1981, 1983) propose such a concept. Their idea is related to what Dawid (1982a) defines and calls "intersubjectivity." According to this definition, the opinion or conclusion reached by an individual from the results of an experiment would be called "objective" or perhaps "intersubjective," if the same conclusion were reached by a succession of individuals faced with the same results. But just as the classical notion of objectivity is challenged by inevitable variations in the results of repeated experiments, so intersubjectivity needs to contend with variations in the conclusions derived by the succession of individuals viewing the evidence. This calls for an ana-

Christian Genest is Assistant Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. James V. Zidek is Professor and Head, Department of Statistics, University of British Columbia, 2021 West Mall, Vancouver, B. C., Canada V6T 1W5.

logue of the law of averages, that is, a method of "averaging" the possibly diverging opinions of a group of analysts and a limit theory for the long run.

Apart from the philosophical considerations outlined above, there are obvious practical reasons for wanting to merge opinions. When faced with a complex decision problem, a decision maker who has sought the advice of a group of experts might want to incorporate their opinions into his/her own. The "jury problem" of Savage (1954) exemplifies a problem of an entirely different sort in which a committee or parliament needs to summarize the final, possibly conflicting opinions of its members "at the end of the day." More informally, the group might simply consist of a number of individuals separated in time or space and whose opinions are already on record.

Of critical importance is whether the group must agree to the resulting aggregate opinion as an expression of consensus. If so, this becomes a particular type of joint decision problem, one which has not yet been treated in the statistical literature. At the outset, the large variety of situations encompassed by the aggregation problem seems to rule out the existence of a uniquely "rational" way of reaching a consensus, although notable claims to the contrary have been made by Lehrer (1975, 1976, 1983). In any case, it seems fair to say that the consensus question is at the origin of a large proportion of the research described in this article. We will not even attempt to survey the literature on group decision making per se despite its even greater practical importance than (and obvious connection with) the opinion pooling problem.

Before opinions can be combined, they must be elicited and expressed in some quantitative form. As there are fundamental difficulties with the representation and elicitation of the opinions of a single individual, it should be anticipated that a theory describing the behavior of a group of individuals of necessity will inherit these same difficulties. In fact, the form in which the group's opinions are expressed influences, to some extent, the selection of a pooling method, since it would be natural to express the consensus judgment in the same form as the originals. For that reason, our survey would be incomplete without at least some mention of the issues at stake when modeling uncertainty. This is the object of Section 2, which concentrates on the case of one individual and points out the controversies which seem most relevant to the general case. This enables us at the same time to introduce the necessary terminology and notation. Then, in Sections 3, 4, and 5, we describe the plethora of methods which have been proposed for combining opinions, using the way in which these opinions have been expressed as a somewhat arbitrary, but convenient thread for exposition. As everywhere else in the literature, it will be assumed that all the subjects have expressed their judgments in a common form.

In Section 3, we present a number of formulas which may be used to summarize a set of opinions which have been encoded as subjective probability measures, densities, or mass functions. Each formula is derived from rather weak and qualitative assumptions about what properties a pooling operator should have. In each case, different sets of assumptions are explored and these often yield different formulas, but the subjective distributions themselves are always taken as the primitive objects of study, the aim being to determine an acceptable summary of these objects.

The domain and objectives of the methods which are presented in Section 4 are different from those of Section 3. Here, opinions are still represented by subjective probability distributions, but the focus of study is the background information which led to the formation of those opinions. The goal is the accumulation of the knowledge which is jointly held by the individuals in the group. So whereas in Section 3 (and in the assumptions stated therein) there is no explicit concern for the nature of the information which led each of the group members to their individual opinion, this becomes a matter of fundamental importance in Section 4. Moreover, the relative quality of the respective background experiences of the individuals is now an issue, as is the degree of interdependence of the information these backgrounds provide. As a result, difficulties in judging and interpreting the elicited opinions can then be addressed more adequately.

In Section 5, we survey the few solution concepts which have been proposed in cases where the opinions provided by a group are not expressed directly in terms of probability. As this article is being written, this topic is in the early stages of its development, but future research efforts are likely to go in that direction. For completeness, a brief survey of the literature on the modelization of behavioral group consensus formation through feedback and interaction is included in Section 6.

This survey complements those of Winkler (1968), Pill (1971), Beach (1975), Hogarth (1975), Weerahandi and Zidek (1981), and French (1985). It is based in part on lectures given by Zidek in May of 1983 at the University of London, on the invitation of its Board of Studies for Statistics.

2. QUANTIFYING DEGREES OF BELIEF

Most of the solutions to the aggregation problem require that each individual's opinion be encoded as a *subjective probability distribution*. Because of its fundamental importance, we will begin this section with a review of the multiple aspects of this concept and an indication of the controversies which surround its existence. If Θ denotes a collection of mutually exclusive statements about the world, exactly one of which is true, a probability measure P assigns a number $0 \le P(E) \le 1$ to each possible subset E of Θ, according to the degree to which this subset is believed to contain the fixed, but unobserved realization $\theta \in \Theta$. Generally, P is constructed in such a way that

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i) \tag{2.1}$$

when the E_i's are mutually disjoint subsets of Θ. When Θ is infinite, condition (2.1) is usually referred to as *countable additivity* and may only hold on a restricted class of subsets E_i of Θ. This class of subsets, henceforth denoted by Σ, is called the σ-field of events of interest on Θ.

If there exists a convenient reference measure μ on Θ (such as Lebesgue measure "dx" on the real line), then P can be expressed as a probability density function p with respect to μ. In that case, $P(E) = \int_E p(\theta)\, d\mu(\theta)$ and $\int_\Theta p(\theta)\, d\mu(\theta) = 1$. Moreover, if Θ is discrete, i.e., at most countable, no generality is lost in assuming that μ is the counting measure which assigns value 1 to each singleton $\{\theta\}$ in Θ. The literature then refers to p as a probability mass function.

Whether they are expressed as probability measures, densities, or mass functions, subjective probability distributions are widely employed in both theory and practice. Just how such a distribution can be chosen to capture the structure of an individual's beliefs is another matter however. This question has been investigated by Savage (1954), de Finetti (1970), Fine (1973), French (1982), and Lindley (1982), to name but a few. To them, the existence of probabilities is the ineluctable expression of coherent beliefs, which in turn are reducible to rational preference. Moreover, they insist that their normative theories of preference are weak, and that they accommodate all but self-defeating desires. Yet, even in the relatively simple case where Θ contains 5 points, Kraft, Pratt, and Seidenberg (1959) have been able to show that comparability, transitivity, and additivity of a preference relation on the subsets of Θ may not be sufficient to guarantee the existence of a strictly agreeing probability measure. Moreover, many philosophers have argued, especially in view of the probability amalgamation question (e.g., Baird, 1985), that probability distributions do not adequately summarize a person's or a group's information.

Leaving existence questions aside, it is also possible to challenge the meaning of a probability statement, even something as simple as "$P(E) = 1/3$." It is not clear, for instance, whether $P(E)$ can be regarded as a measurement, or even whether one subject's probabilities have any meaning for another person (de Finetti, 1970). Admitting that they don't, how could we then hope to compare, let alone combine the opinions of several people? On the other hand, assuming that $P(E)$ can be interpreted as a measurement, it will not necessarily be calibrated (Dawid, 1982b; Morris, 1977, 1983; Schervish, 1983) and how best to elicit it remains unclear, despite recent attempts by Kadane et al. (1980) and Chaloner and Duncan (1983), among others. Finally, common observation and experimental studies (Winkler, 1967; Tversky and Kahneman, 1974; Slovic, Fischoff, and Lichtenstein, 1977) tend to confirm that even if an individual has a good knowledge of the relative likelihood of the various possible states of nature, it cannot generally be expected that he/she will also master the calculus of probabilities and encode his/her opinion accordingly. This problem has recently been studied by Lindley, Tversky, and Brown (1979) and some of their techniques have bearing on the Bayesian approach to pooling which will be discussed in Section 4. Savage (1971) reviews some other major difficulties encountered in the experimental elicitation of preferences and opinion. Those include "motivational bias" in the formulation of an individual distribution (Benson and Nichols, 1982), as well as difficulties in assessing the likelihood of rare events.

Some of the aforementioned controversies have been thought to derive from the definition of a subjective probability distribution. Thus, finite additivity, rather than countable additivity as in (2.1), has been advocated (de Finetti, 1970), although there are indications that this may be too strong as well (Fishburn, 1983). Insofar as combining opinions is concerned, however, the distinction between finitely and countably additive probability measures will not be important in our paper, because all criteria considered to this day for aggregating the latter will work just as well with the former. On the other hand, an individual's total probability is usually taken to be 1 on the assumption that this may be done without loss of generality. But, says Fine (1973, p. 66),

> "... nothing necessitates this choice; it is a mere convention. The problem with conventions is that we may loose sight of their arbitrary origin. The unit normalization and nonnegativity conventions begin to appear as substantive properties ... "

Thus there is no particular reason why a number of individuals should use the same scale for their probabilities, and assuming that they do is not necessarily unconsequential. For example, different members of a group may well wish to have different scales when the parameter set Θ is not exhaustive or logically complete, or when their levels of expertise vary widely. It should be remembered that all probabilities are conditional and that it may not be possible for an individual to articulate all the elements in his/her conditioning set even if he/she is permitted to enter into dialogue with the rest of the group. This comparability issue, which does not arise in conventional one-person decision analysis, would seem to be analogous to the well known and unresolved problem of comparing utilities in social-welfare economics.

Other alternatives to probability distributions might include betting odds or "ratio judgments" (Smith, 1961; Aczél and Saaty, 1983), belief functions (Shafer, 1976), or perhaps vague priors and posteriors $p(\theta)$ for which $\int_\Theta p(\theta)\, d\mu(\theta) = \infty$. A number of papers which exploit these forms of opinions are reviewed in Section 5. However, it should be kept in mind that use of these concepts might lead to controversy, as illustrated by Stone (1976) in the case of uniform priors. In our annotated bibliography, we have indicated whether an author has assumed that degrees of belief were expressed as discrete probabilities (i.e., when Θ is discrete), probability measures (Θ infinite), probability density functions (with respect to some reference measure μ), cumulative distribution functions, or simply as odds or log odds.

3. COMBINING SUBJECTIVE PROBABILITY DISTRIBUTIONS

Suppose given n subjective probability measures $P_1, \cdots, P_n$, and suppose that a summary of the P_i's is required. There turns out to be a significant difference in the mathematical treatment of cases where the underlying set of events Σ on Θ contains more than four or contains exactly four sets, i.e., $\Sigma = \{\varnothing, E, E^c, \Theta\}$. In the former case, Θ is said to be *tertiary* (Wagner, 1982), and since this situation is more general, we will restrict ourselves to this case in the present section.

If Σ is a σ-field and if $P_1, \cdots, P_n$ are expressed as probability measures on Σ, the *linear opinion pool*

$$T(P_1, \cdots, P_n) = \sum_{i=1}^{n} w_i P_i \tag{3.1}$$

with non-negative weights w_i such that $\sum_{i=1}^n w_i = 1$ was proposed by Stone (1961) and is attributed to Laplace by Bacharach (1979). This formula would seem natural because $T(P_1, \cdots, P_n)$ is then a probability measure on Σ. Moreover, McConway (1981) and Wagner (1982) have shown independently that a probability measure $T(P_1, \cdots, P_n)$ is necessarily of the form (3.1) if

$$T(P_1, \cdots, P_n)(A) = F[P_1(A), \cdots, P_n(A)] \tag{3.2}$$

for some arbitrary function $F: [0, 1]^n \to [0, 1]$ and every event A of interest in Σ.

Condition (3.2) above is called the *strong setwise function property* (SSFP) by McConway (1981), "strong label neutrality" by Wagner (1982), and "context-free assumption" by Bordley and Wolff (1981). It is similar to an independence of irrelevant alternatives hypothesis commonly postulated by decision theorists (e.g., Bacharach, 1975), but it has consequences whose implications may be considered too strong. For instance, it follows immediately from (3.2) that

$$P_1(A) = \cdots = P_n(A) = 0 \Rightarrow T(P_1, \cdots, P_n)(A) = 0, \tag{3.3}$$

because $F(0, \cdots, 0) = T(P_1, \cdots, P_n)(\varnothing) = 0$ by definition of a probability measure. This assumption, referred to as the *zero preservation property* (ZPP), is one of a general class of axioms which would require the consensus distribution to embrace any aspect of the subjects' personal opinions that are already the object of an (implicit) agreement between them. Genest and Wagner (1984) have pointed out the dubious nature of such requirements in cases where the group is reporting to an external decision maker, but preservation axioms sometimes entail serious difficulties even without reference to a third party.

Laddaga (1977), for instance, favors the *independence preservation property* (IPP), according to which one should have

$$T(P_1, \cdots, P_n)(A \cap B) = T(P_1, \cdots, P_n)(A)T(P_1, \cdots, P_n)(B) \tag{3.4}$$

whenever $P_i(A \cap B) = P_i(A)P_i(B)$ for some A and B in Σ, $i = 1, \cdots, n$. As he points out, however, linear opinion pools cannot satisfy (3.4) unless they are *dictatorial*, i.e., unless $w_i = 1$ for some i and 0 for all others. This result, which was proved formally by Lehrer and Wagner (1983), is called an "impossibility theorem" due to the unsatisfactory nature of dictatorial aggregation methods. This result is in fact closely related to an earlier finding of Dalkey (1972, 1975), who observed that dictatorships alone could accommodate condition (3.2) when it is also required to hold for conditional probabilities, viz.

$$T(P_1, \cdots, P_n)(A \mid B) = F[P_1(A \mid B), \cdots, P_n(A \mid B)].$$

The severity of the above conditionality inconsistency is uncertain. McConway (1981) and Genest (1984c), who both discuss its impact, argue that the same expert weights $w_1, \cdots, w_n$ need not be used before and after the occurrence of B is revealed. This leads McConway to resurrect a Bayesian revision of the weights which had been developed in somewhat different contexts by Madansky (1964) and Roberts (1965). In the presence of additional data for which there is a commonly agreed likelihood, however, Raiffa

(1968, Section 8.11) demonstrates in a special case how this procedure might entice the subjects to delay disclosing their opinions until after the data is reported, so as to increase the relative weight of their respective distribution in the pool. This problem has led Madansky (1964, 1978) to the notion of externally Bayesian groups which we will discuss shortly.

Basing their analysis on Bayesian considerations, Bordley and Wolff (1981) have attempted to explain the ill effects of independence preservation by suggesting that condition (3.2) be enlarged to allow the function F to depend on the identity of the set A, viz.

$$T(P_1, \cdots, P_n)(A) = F_A[P_1(A), \cdots, P_n(A)], \quad A \in \Sigma. \tag{3.5}$$

McConway (1981), who calls (3.5) the *weak setwise function property* (WSFP), shows that it is equivalent to a *marginalization property* (MP) according to which the pooling procedure should commute with the process of reducing the size of the σ-field Σ to a sub-σ-field, say $\Sigma^* \subset \Sigma$. His mathematical formulation of MP, which involves a potentially infinite family of pooling procedures, can be stated somewhat more simply in terms of only one combination method by requiring that

$$T(P_1, \cdots, P_n)(A) = T([P_1 \mid \Sigma^*], \cdots, [P_n \mid \Sigma^*])(A) \tag{3.6}$$

for all events $A \in \Sigma^*$ whenever $[P_i \mid \Sigma^*]$ is a Carathéodory extension of $P_i \mid \Sigma^*$ to Σ.

Surprisingly, perhaps, pooling formulas which obey (3.6) can be characterized completely, and their form does not depart greatly from (3.1). In fact, Aczél, Ng, and Wagner (1984), and Genest (1984c) in greater generality, have proven that if T has the MP, there must exist a probability measure P_0 and weights $w_0, \cdots, w_n \in [-1, 1]$ such that

$$T(P_1, \cdots, P_n) = \sum_{i=0}^{n} w_i P_i, \tag{3.7}$$

where the weights add up to one and must satisfy other consistency conditions to insure that (3.7) is a probability measure. An operator of this form may be called a *generalized linear opinion pool*, since putting $w_0 = 0$ in (3.7) reduces it to (3.1). In fact, the only operators of the form (3.7) which satisfy the ZPP are also regular linear opinion pools, as remarked by McConway (1981). The possibility of negative weights is interesting, however, since generalized linear pools could vary inversely with the P_i's or could even ignore all the P_i's! Unfortunately, Genest (1984c) has established that only dictatorships and a very restricted class of imposed consensus functions will verify both the MP and the IPP. In cases where Σ and Θ are countable and Θ contains at least five points, Genest and Wagner (1984) have recently derived an even stronger result to the effect that no pooling operator of the form

$$T(P_1, \cdots, P_n)(\theta) = \frac{F_\theta[P_1(\theta), \cdots, P_n(\theta)]}{\sum_{\eta \in \Theta} F_\eta[P_1(\eta), \cdots, P_n(\eta)]} \propto F_\theta[P_1(\theta), \cdots, P_n(\theta)] \tag{3.8}$$

except dictatorships can have the IPP. In an attempt to find independence-preserving formulas, Wagner (1984) has weakened the IPP by requiring that the probabilities $P_i(A)$ and $P_i(B)$ in line (3.4) lie strictly between 0 and 1, but that too has been unsuccessful. Thus the problem of finding reasonable pooling methods which have that property remains open. Furthermore, no definite indications can be given concerning the choice or interpretation of the weights or the arbitrary measure P_0 in (3.7), despite meritorious consideration of the question of weight assessment by Roberts (1965), Winkler (1968), and Raiffa (1968).

Returning to condition (3.5) for a moment, the dependence of F upon the identity of the event A in Σ might conduce the mistaken belief that the marginalization property admits the possibility of pooling operators with *variable* "expert weights." Bordley and Wolff (1981), for instance, argue that a pooling operator of the form

$$T(P_1, \cdots, P_n)(A) = \sum_{i=1}^{n} w_i(A) P_i(A), \; A \in \Sigma \tag{3.9}$$

could be constructed with variable weights $w_i(A)$ depending on the event A. Since (3.9) has the WSFP, however, it must be of the form (3.7) and hence $w_i(A) \equiv w_i$, $i = 1, \cdots, n$. French (1985) views this lack of flexibility of the generalized linear opinion pool as a flaw, arguing intuitively that subjects trained in different fields of expertise may be "more expert" at forecasting events belonging to certain sub-σ-fields of Σ than others. In fact, McConway (1981) states that the MP might appear counter-intuitive in situations where Θ is a product parameter space and the measures $P_1, \cdots, P_n$ are product measures on this space. For this reason, operators of the form (3.8) might be considered a reasonable alternative to (3.5) or (3.6). It might be, for instance, that this class of aggregation methods admits formulas with θ-variable weights. Alas, we are not aware of any axiomatic justification leading to (3.8), and it is not clear how one could even define this class in cases where Θ is uncountable and the opinions must be expressed as probability measures on Σ.

When a mutually agreeable reference measure μ exists on Θ, people's opinions are sometimes expressed as probability density functions with respect to μ

rather than by probability measures. The linear opinion pool (3.7) has an obvious counterpart in terms of densities, viz.

$$(3.10) \qquad T(p_1, \cdots, p_n) = \sum_{i=0}^{n} w_i p_i,$$

where p_i can be described technically as the Radon-Nikodym derivative of P_i with respect to μ, $1 \le i \le n$, and p_0 is an arbitrary density. The required axiomatic support for the use of (3.10) is given in Genest (1983) in the case where $w_0 = 0$, but the more general case remains to be treated. Another obvious difficulty associated with the density version of the linear opinion pool is that it is typically multimodal, so that no clear-cut choice for a jointly preferred action emerges in decision making situations. In addition, formulas (3.7) and (3.10) depend critically on the assumption that every individual has expressed his/her opinion using the same 0–1 probability scale. So, for example, vague priors cannot be accommodated. Finally, a number of experimental studies (e.g., Staël von Holstein, 1970; Winkler, 1971) have pointed to the fact that linear pools are relatively insensitive to the choice of expert weights, although that could be considered either advantageous or disadvantageous.

Many of the above difficulties associated with the use of (3.7) or (3.10) are overcome by the *logarithmic opinion pool* which Bacharach (1972) attributes to Peter Hammond. This method of pooling applies to densities, say $p_1, \cdots, p_n$ and gives

$$(3.11) \quad T(p_1, \cdots, p_n) = \prod_{i=1}^{n} p_i^{w_i} \Big/ \int \prod_{i=1}^{n} p_i^{w_i}\, d\mu,$$

where $w_1, \cdots, w_n$ are weights such that the integral in the denominator of (3.11) is finite. In this section, we will confine our discussion to the case where $\sum_{i=1}^{n} w_i = 1$. As we will see in the following section, however, formulas analogous to (3.11) but with unrestricted weights can be derived within the Bayesian framework. As pointed out by Winkler (1968), the logarithmic opinion pool also has a natural-conjugate interpretation. Unlike the linear opinion pool, it is typically unimodal and less dispersed. Thus it is more likely to indicate consensual values when decisions must be made, and in that respect, it may be considered more than a simple representation of the diverse opinions of the members of the group. Moreover, (3.11) is invariant under rescaling of individual degrees of belief. Thus, if $T(p_1, \cdots, p_n)$ were input into a formal decision analysis, it would preserve an important credo of uniBayesian decision theory according to which the optimal decision should not depend upon the choice of scale for the utility function or prior probability distribution (Weerahandi and Zidek, 1981; 1983). In addition, it should be noted that when the parameter space Θ is finite and a 0–1 utility function is adopted in the formal analysis to which we just referred, equation (3.11) is equivalent to the Nash (1950) product.

In our opinion, the most compelling reason for using a logarithmic opinion pool is that it is *externally Bayesian*, prior to posterior coherent, or that it has the data independence property, in the terminologies of Madansky (1964, 1978), Weerahandi and Zidek (1978), and McConway (1978), respectively. This means that finding the consensus distribution commutes with the process of revising distributions using a commonly agreed likelihood. Thus, to be externally Bayesian $T(p_1, \cdots, p_n)$ must satisfy

$$(3.12) \quad T\left[\frac{lp_1}{\int lp_1\, d\mu}, \cdots, \frac{lp_n}{\int lp_n\, d\mu}\right] = \frac{lT(p_1, \cdots, p_n)}{\int lT(p_1, \cdots, p_n)\, d\mu},$$

where $l\colon\Theta \to (0, \infty)$ is an arbitrary likelihood function and the equality in (3.12) may be required to hold only μ almost everywhere on Θ. The plausibility of this axiom derives from the latent objective behind pooling priors: if $T(p_1, \cdots, p_n)$ does in fact represent the beliefs of all the members of the group, the order in which the pooling and the updating are done should be immaterial. To this extent, the behavior of the group could be perceived as that of a single Bayesian. On the other hand, external Bayesianity would make much less sense if the responsibility of the pooling is that of an external decision maker, since the arrival of new data might change his/her evaluation of the relative expertise of the subjects who were consulted. Other elements will be added to this debate in Section 4.

Genest (1984b) has shown that (3.11) is the only externally Bayesian formula which satisfies

$$(3.13) \quad T(p_1, \cdots, p_n)(\theta) \propto F[p_1(\theta), \cdots, p_n(\theta)],$$

where, as in (3.8), the symbol $\propto$ is interpreted to mean "proportional up to a constant independent of θ." In particular, it follows that (3.10) is not externally Bayesian unless it reduces to a dictatorship (Genest, 1984a). Genest (1984b) interprets condition (3.13) as a version of the "likelihood principle" for pooling operators, arguing, much in the same spirit as Bernardo (1979), that outcomes not observed should play no role in the determination of the consensus at a particular point θ. However, a better case for looking at formulas of this form derives from the complexity of the most general solution, which was recently obtained by Genest, McConway, and Schervish (1986). Here, the class of all externally Bayesian aggregation methods is characterized without any regularity conditions on the space Θ, but the general solution is seen to require the specification of a consensual density for an infinite number of opinion vectors, each of which must be selected by the axiom of choice! The work of Genest, McConway, and Schervish (1986) brings about an interesting extension of (3.11), however, when condition (3.13) is relaxed to allow the function F to change with θ. The result is a *generalized logarithmic opinion pool,*

$$(3.14)\quad T(p_1, \cdots, p_n) = g \prod_{i=1}^{n} f_i^{w_i} \Big/ \int g \prod_{i=1}^{n} f_i^{w_i}\, d\mu,$$

where g is some essentially bounded function on Θ. This formula may be viewed as the logarithmic analogue of (3.7) in which g plays the role of P_0, although Genest, McConway, and Schervish (1986) caution their readers against this interpretation and suggest regarding g as a likelihood instead. In (3.7), (3.11), and (3.14), it should be pointed out that the weights are demonstrably non-negative unless the underlying σ-field on Θ is essentially finite.

Despite its advantages, the logarithmic opinion pool suffers from the same problem as its linear counterpart in that it clearly lacks a normative basis for choosing the pooling weights, except perhaps when the resulting formula could also be derived using the Bayesian techniques which are discussed in the following section. Whatever scheme is elected for assigning the weights in (3.11) or (3.14), however, zeros will constitute vetos and unduly great emphasis will tend to be placed on the opinions of single individuals. For an analogy, think of the familiar situation faced in a uniBayesian analysis where the supports of the likelihood and prior functions are disjoint.

4. THE SUPRA BAYESIAN APPROACH

If an individual's probability distribution is the expression of that person's subjective beliefs, it is nevertheless based on whatever "objective" prior experience the individual has had with the problem at hand. Indeed, experts who have been trained in the same specialty will typically share a fair amount of information. Their current opinions may differ because they do not share all the same evidence or perhaps because they do not interpret common data in the same way. However, none of the pooling formulas which we have presented above takes explicit account of this important feature of the subjects' judgments. "Expert" weights do allow for some discrimination on that basis, but only in vague, somewhat ill defined ways.

The absence of a group leader or perhaps an external decision maker may be one practical reason for avoiding the issue of opinion dependence. One might argue, justly we think, that a group of individuals who have conflicting beliefs are unlikely to reach an agreement on the amount of overlap existing between their opinions. In some cases, however, there does exist a decision maker to whom the panel is reporting. In other circumstances, such as a "jury problem" for instance, there may also be enough affinity in the group to justify the assumption of a fictitious decision maker representing the "synthetic personality" of the group (Hogarth, 1975, p. 282). Keeney and Raiffa (1976) call this fictitious being the "supra Bayesian." It is thus the seemingly impossible task of this supra Bayesian decision maker to evaluate the individuals, their prior information sets, the interdependence of these information sets, the experts' "calibration" or honesty, etc.

It was recognized very early (Winkler, 1968; Morris, 1974, 1977) that if such an altruistic supra Bayesian exists, the pooling process itself is not a problem. Indeed, once this decision maker has determined his/her prior p and the appropriate likelihood $l(p_1, \cdots, p_n \mid \theta)$ for the experts' opinions, he/she can then treat the *stated* opinions of the group as data and update his/her prior via Bayes' theorem:

$$T(p_1, \cdots, p_n)(\theta) = p(\theta \mid p_1, \cdots, p_n) \propto p(\theta) l(p_1, \cdots, p_n \mid \theta).$$

Here, the pooling operator is simply Bayes' rule and the decision maker's posterior distribution is the "consensus." The problem, of course, is that if the supra Bayesian is only virtual, the delicate choice of an appropriate likelihood would fall on the group. Worse, the choice of the decision maker's prior distribution itself would have to be the object of a consensus! Note also that the obvious solution which consists of using an "uninformative prior" is not entirely satisfactory nor free from controversies (Seidenfeld, 1979).

Lindley (1985) admits the need for a pooling method in situations of the type described in the last paragraph. He argues that there is no "normative theory" available for group decision making as there is for single decision makers. So he concludes that the introduction of the supra Bayesian

> ". . . is not merely an artifice but essential, at least at the present stage of development. Any other approach has an element of adhockery about it because it does not derive from the inherent logic of the Bayesian paradigm."

This judgment seems unduly severe. Whatever the merits of the "Bayesian paradigm," groups would not appoint a supra Bayesian merely to achieve a "normative method," in Lindley's terminology. Rather, a truly normative theory would prescribe how a group ought to behave if its members agree upon certain basic, qualitative conditions which we may call "axioms." The methods proposed in Section 3 are based on such axioms, and just because they do not flow from the "inherent logic of the Bayesian paradigm" does not make them non-normative. They too are logical consequences albeit of different axioms than those which yield Bayesianity in the case of a single decision maker. Acceptance of the axioms compels the use of the derived formula, and we suspect that the axioms discussed in Section 3 might be less controversial than the appointment by the group of a supra Bayesian.

The supra Bayesian approach is nevertheless valuable because it provides a way of understanding the axioms on which aggregating formulas like the linear opinion pool rest. However qualitative these axioms may seem, they embrace ideas which are difficult to grasp. The marginalization property and external Bayesianity, for instance, both have appealing glosses, but they do lead to two markedly different pooling recipes. On the other hand, the seemingly innocuous condition that a pooling formula preserve independence (condition (3.4)) implies dictatorships. Thus, it appears that these conditions have a sharp and unexpected bite, even if choosing among them seems difficult. This is why, in our opinion, a supra Bayesian analysis of these properties is worthwhile and insightful.

The technical foundations for a Bayesian analysis of the problem of aggregating opinions are laid down by Morris (1974, 1977) and built upon by French (1980, 1981), Winkler (1981), Lindley (1983, 1985), Agnew (1985), and Clemen and Winkler (1985). The practical problem addressed by these authors is that of formulating a consensus model which is mathematically tractable while accounting for expert dependence and maintaining a reasonably broad domain of applicability.

The case of a finite $\Theta = \{\theta_1, \cdots, \theta_m\}$ is analyzed by Lindley (1985), who assumes a joint multivariate normal distribution for the quantities $q_{ij} = \log[p_i(\theta_j)]$, $1 \le i \le n$; $1 \le j \le m$. He achieves additional simplicity with assumptions about expectations and conditional expectations. His results subsume those of French (1980, 1981) for the case $m = 2$, in which the difference between the decision maker's prior and posterior log odds is seen to be a linear function of $Q = (q_{11} - q_{12}, \cdots, q_{n1} - q_{n2})$, the vector of the experts' log odds. Explicitly, if $E[Q \mid \theta_i] = \mu_i$ and if Ξ denotes the covariance matrix of Q given θ_i, $i = 1, 2$, then

$$\begin{aligned} &\log[T(p_1, \cdots, p_n)(\theta_1)/T(p_1, \cdots, p_n)(\theta_2)] \\ (4.1)\qquad &= \log\,[p(\theta_1)/p(\theta_2)] \\ &\quad + [\mu_1 - \mu_2]\Xi^{-1}[Q - (\mu_1 + \mu_2)/2]. \end{aligned}$$

Taking antilogarithms on both sides of (4.1) gives (essentially) the logarithmic opinion pool (3.11) equipped with weights proportional to the independent "information content" of each assessment (Freeling, 1981). French's formula, which was inspired by earlier work of Lindley, Tversky, and Brown (1979), is thus a special case of (5.6). It can be seen readily that (4.1) does not have the MP and is not externally Bayesian except in cases which Lindley (1985) identifies. This leads him to reject both external Bayesianity and the marginalization property as "essentially adhockeries." On this point, French (1985) remarks that this model obeys a variation of the externally Bayesian criterion in which the known value of the log likelihood of the data is "filtered out of the experts' posterior beliefs."

Winkler (1981) and Lindley (1983) tackle the more complicated case where Θ is infinite, a continuum possibly. Both authors suppose that the experts' opinions can be represented by a vector of numbers, such as the mean and standard deviation of the individuals' prior distributions, and their results overlap. To represent the decision maker's likelihood, Lindley (1983) uses a multivariate normal distribution which, conditional upon the true value of θ and knowledge of the experts' stated standard deviations σ_i, has mean $\alpha_i + \beta_i\theta$ and variance $\gamma_i^2\sigma_i^2$, $i = 1, \cdots, n$. The parameters α_i, β_i and γ_i introduced here allow for bias or miscalibration of the group's opinions, and Lindley discusses how uncertainty about their value could be incorporated by specifying a hierarchical prior. When data is available on the experts' past performance, however, estimation of these parameters could proceed along lines similar to those proposed by Morris (1977) for the assessment of his calibration functions, although Clemen (1986) argues that Morris' "joint calibration" adjustment is equivalent to assessing a complicated likelihood for the experts' opinions in all cases but those in which the decision maker has no prior information.

Assuming the above model and the use of an improper diffuse prior, Lindley (1983) shows that the decision maker's subjective posterior is a normal distribution whose mean μ^* is a weighted sum of the experts' means, viz.

$$(4.2)\qquad \mu^* = \sum_{i=1}^{n} w_i\mu_i,$$

where $w_i = \sum_{j=1}^n \xi^{ij}/\sum_{k=1}^n \sum_{j=1}^n \xi^{kj}$, $\Xi^{-1} = (\xi^{ij})$ and, as before, the off-diagonal entries of the covariance matrix Ξ must be specified by the decision maker. Winkler (1981), who also derives this result from slightly different premises, observes that (4.2) is not necessarily a convex combination of the experts' means, since the weights can sometimes be negative. In this context, Clemen and Winkler (1985) analyze the impact of dependence on the precision and value of the information supplied by the experts. They show that dependent sources of information are generally equivalent to a much smaller number of independent sources, and that the decision maker's posterior distribution can be quite sensitive (as measured by its variance, say) to the degree of dependence among the experts. Of late, Agnew (1985) has also considered an extension of Winkler's work to situations in which the experts provide assessments about more than one random variable.

It should be emphasized that the Winkler-Lindley model requires as inputs the covariances between subjects. If these parameters are not known, they suggest the use of an inverted Wishart distribution, except perhaps in cases where the experts are deemed to be independent given θ. This assumption leads to a t distribution for the consensus (or a product of t distributions in the independence case). Lindley (1983) contrasts the use of normal versus t distributions and finds the latter "more in accord with intuition," although he sees problems with the use of inverted Wishart densities. The subjective assessment of correlation coefficients is studied by Gokhale and Press (1982) and their role in probability reconciliation is further discussed by Freeling (1981).

Genest and Schervish (1985) attempt to derive Bayesian pooling operators, which do not require the elicitation of a complete likelihood function for the experts' judgments. Their analysis, which is carried out in the case of a single uncertain event E, starts from the assumption that the decision maker has a prior probability p for E, and that he/she is willing or capable of specifying some moments of his/her *marginal* distribution F for $(p_1, \cdots, p_n)$, the vector of experts' probabilities for the occurrence of E. Genest and Schervish then look for those pooling formulas $T(p_1, \cdots, p_n)$ which yield a posterior probability that is *guaranteed* to be consistent with the unelicited F, in the sense that a joint distribution for E and $(p_1, \cdots, p_n)$ which is compatible with F can always be found such that $T(p_1, \cdots, p_n)$ is the posterior probability of E given p and the experts' opinions.

When only the first moment $(\mu_1, \cdots, \mu_n)$ of F is specified by the supra Bayesian, the only procedures which survive this test are of the form

$$T(p_1, \cdots, p_n) = p + \sum_{i=1}^{n} w_i(p_i - \mu_i) \tag{4.3}$$

with possibly negative weights w_i representing the coefficients of (multiple) linear regression between 1_E, the indicator random variable of the event E, and the p_i's. When $\mu_1 = \cdots = \mu_n = p$, it is worthwhile to note that equations (4.3) and (3.7) agree, although the choice of weights in (4.3) obeys different restrictions. The linear opinion pool, therefore, does correspond to an application of Bayes' theorem in certain circumstances delineated by Genest and Schervish (1985).

In the same paper, Genest and Schervish apply their approach to the situation in which a decision maker is also willing to assume that the subjects are conditionally independent given E (and given E^c). Interestingly, the result is a kind of logarithmic pool with possibly negative weights, viz.

$$T(p_1, \cdots, p_n) = A/(A + B), \tag{4.4}$$

where

$$A = p^{1-n} \prod_{i=1}^{n} [p + w_i(p_i - \mu_i)]$$

and

$$B = (1 - p)^{1-n} \prod_{i=1}^{n} [1 - p - w_i(p_i - \mu_i)].$$

When all the μ_i's equal p, (4.4) becomes a special case of (4.1), the formula derived by French (1980, 1981) using normality and justified axiomatically by Bordley (1982) via the theory of conjoint measurement. Clearly, formula (4.4) is not externally Bayesian. In fact, when $p_i > \mu_i$ for each $1 \leq i \leq n$ and the weights are positive, it exhibits the same "group polarization" phenomenon described and analyzed by Bordley (1982, 1983a). Bordley's contribution will be discussed in greater detail in the following section.

Because the approach of Genest and Schervish (1985) provides a Bayesian answer while requiring a minimum of *a priori* assessments, it would seem well suited for helping groups to arrive at a consensus. It is certainly easier to conceive of circumstances when a panel of experts would agree on the first few moments rather than on an entire predictive distribution modeling the dependence of their information sets. The group might then combine its opinions using (4.3) or (4.4), say, and the resulting consensus could be perceived by all of them as being coherent because of its Bayesian origin. Nevertheless, the problem of selecting the supra Bayesian's prior would remain, and the approach is still limited to those cases where the only source of uncertainty is a single event. Extensions of this model would be of considerable interest.

5. OTHER POOLING RECIPES

So far, our discussion has focused primarily on pooling operators acting on a set of probability measures or probability density functions. In this section, we will examine situations where the subjects in the group have expressed their beliefs in the form of odds, log odds, cumulative distribution functions, or perhaps improper probability density functions. Since probability measures can be encoded as a sequence of odds or log odds when the space Θ is countable, there is a certain amount of arbitrariness in the choice of papers which are reviewed here.

Ratcliff (1979) and Thomas and Ross (1980) consider procedures for combining a number of continuous distribution functions $F_1, \cdots, F_n$ on IR. Their basic method, called *Vincentization*, consists of averaging the α per cent quantiles of the experts' distributions in order to construct the α per cent quantile of the consensus, $0 < \alpha < 1$. Thus if q_i is the α per cent quantile of F_i, i.e., $F_i(q_i) = \alpha$, the consensual distribution F would be defined by setting $F^{-1}(\alpha) = \sum_{i=1}^{n} w_i q_i$. As demonstrated by Thomas and Ross (1980), the main advantage of this approach is that Vincentization is closed under location scale families, i.e., when the F_i's are of the form $F(t) = \Phi[(t - \lambda)/\gamma]$, where λ is a centering parameter and γ represents the scale. Thus, if the subjects' distributions were normal, Cauchy, exponential, or logistic, the same would be true of the Vincent consensus. In fact, the parameters of the group distribution would then be arithmetic averages of its members' λ_i's and γ_i's, much as in Winkler's (1981) formula (4.2). More generally, Thomas and Ross give conditions under which the two parameters of the experts' distributions could be combined using other kinds of averages, e.g., geometric or harmonic means. A clear disadvantage of this entire approach, however, is that its use is currently restricted to continuous distribution functions and real valued parameter spaces Θ. Moreover, it is not obvious how the method could be adapted to handle more general spaces.

In cases where the number of alternatives in Θ is finite, Smith (1961) has suggested that an obvious measure of belief could be obtained by testing at what odds a subject is prepared to bet between two competing alternatives. This has led Aczél and Saaty (1983) to use odds ratios $p_1, \cdots, p_n > 0$ as the starting point of their analysis. They assume, quite generally, that the influence of each individual opinion on the consensual odds ratio $T(p_1, \cdots, p_n)$ can be separated in such a way that

$$(5.1) \qquad T(p_1, \cdots, p_n) = f(p_1) \diamond \cdots \diamond f(p_n),$$

where f is a real valued function and $\diamond$ is some unspecified operation on IR_+ which is both continuous and associative. In particular, $\diamond$ could be ordinary addition or multiplication. Aczél and Saaty then look for those pooling formulas of the form (5.1) which *preserve unanimity*, i.e., $T(p, \cdots, p) = p$ for all $p > 0$, and obey the reciprocal property

$$(5.2) \qquad T(p_1^r, \cdots, p_n^r) = T(p_1, \cdots, p_n)^r,$$

where $r = -1$. The result is a class of quasilogarithmic pools

$$(5.3) \qquad T(p_1, \cdots, p_n) = \exp\left[\psi^{-1}\left\{\sum_{i=1}^{n} \frac{\psi(\log[p_i])}{n}\right\}\right]$$

where ψ is an arbitrary continuous, strictly monotonic odd function. By imposing a certain *homogeneity requirement*, viz.

$$(5.4) \qquad T(kp_1, \cdots, kp_n) = kT(p_1, \cdots, p_n), k > 0,$$

Aczél and Saaty (1983) then succeed in reducing (5.3) to a non-normalized geometric average of the p_i's. In these formulas, each expert is assigned an equal weight of $1/n$, because (5.1) does not permit treating the subjects' opinions asymmetrically. Aczél (1984) has attended to this issue by allowing the function f in (5.1) to be different for each group member, and the expected generalization of (5.3) ensues with unequal weights adding up to one. In the same vein, Aczél and Alsina (1984) have obtained extended forms of (5.3), notably by assuming that (5.2) holds for two arbitrary values r_1 and r_2 such that $\log(|r_1|)/\log(|r_2|)$ is irrational. As far as we can see, the latter requirement is nothing more than a convenient regularity condition. Thus, until a justification is offered, the work of Aczél and Alsina (1984) should be regarded as a purely mathematical exercise.

The analysis of Bordley (1982), founded on the theory of additive conjoint measurement, shares some common features with those papers described just above. Here too the discussion focuses on the amalgamation of personal odds, and it is also assumed that the subjects' odds estimates for an event E versus its complement can be "separated," in the sense that if a number of individuals give the same probability assessments in two different situations, their opinions can be ignored insofar as determining which one of the two situations the group (or the decision maker) prefers. Under certain regularity conditions, this leads to the formula

$$(5.5) \qquad T(p_1, \cdots, p_n) = \psi\left[\sum_{i=1}^{n} g_i(p_i)\right],$$

in which ψ is once again continuous and monotonic while the g_i's are undetermined continuous functions. Requiring that (5.5) obey the law of odds (and thus (5.2) with $r = -1$) and an axiom called the *weak likelihood ratio axiom* which amounts to (5.4), Bordley then demonstrates that the consensual odds must be proportional to a weighted geometric average of the experts' odds p_i, viz.

$$(5.6) \qquad T(p_1, \cdots, p_n) \propto \prod_{i=0}^{n} p_i^{w_i}$$

where, once again, the weights are subject only to the constraint $\sum_{i=0}^{n} w_i = 1$. Here, p_0 might also be interpreted as the decision maker's prior odds, because this is exactly the formula derived by French (1980, 1981) when (4.1) is expressed in terms of probabilities. As pointed out by French (1985), however, the p_0 in (5.6) is not elicited from the decision maker before the subjects are consulted; rather, it is a by-product of Bordley's procedure which is determined only *after* collecting the advice of the group. Thus, Bordley's formula is not quite Bayesian in spirit, but it might be better adapted to situations discussed before where the decision maker is only virtual. Worthy of further consideration here would be the elucidation of the hidden relationships which must exist between the axiomatic approaches leading to (5.3) and (5.5), as well as the connection there is between (5.4) and external Bayesianity.

In 1959, Eisenberg and Gale presented an ingenious scheme for combining discrete probability distributions which also involves betting, but in a different way. In their model, each individual is endowed with a fixed budget, which he/she is then required to bet on the various elements of Θ in such a way as to maximize his/her subjective expectation. The consensus probabilities on Θ are then determined by the same principles which produce track odds or "track probabilities" for pari-mutuel in horse race betting. Norvig (1967) has even developed a dynamic mechanism for reaching this consensus. As noted by Eisenberg and Gale (1959), however, their pari-mutuel method may sometimes enable a member of the group to dictate the consensus odds. For this reason, this approach has never enjoyed much popularity.

Following an idea of Shafer (1976), Walley (1981, 1982) elaborates a personalistic theory of lower and upper probabilities by which degrees of belief may be bracketed (see also Walley and Fine, 1982). In his treatment, which includes a solution to the problem of aggregating opinions, the *lower probability*, $P_l(E)$ of E is defined to be the supremum of prices an individual would be willing to pay to receive a lottery ticket bearing the indicator function of E. The corresponding *upper probability* is defined by $P_u(E) = 1 - P_l(E^c)$. These "probabilities" need not be additive and $P_l(E) \le P_u(E)$ need not hold either. The focus of Walley's theory is the "opinion" $M(P_l)$, the set of all finitely additive probability measures π such that $P_l(E) \le \pi(E)$ for all events E. His aggregated opinions, M, depend on which of 15 criteria are deemed to apply and include, for example, $M = \cap_{i=1}^{n} M_i$ where M_i is the opinion of subject i, $1 \le i \le n$. In the special case where the M_i's and M are all singletons, the linear opinion pool obtains under a version of the unanimity preservation criterion.

Walley argues that the conventional decision paradigm requires an unduly precise specification of individual opinion, inasmuch as a single additive probability function must be specified. In contrast, theories like his admit a whole range of such probability functions, e.g., all measures P such that $P_l(E) \le P(E) \le P_u(E)$ for events E. This benefit is conferred on the group, the diversity of opinions being represented in M. However, the jury remains out on the theory of Walley (1981, 1982). Similar theories like that of Smith (1961) have not enjoyed much success, so the value of Walley's method of combining opinions is in some doubt. In particular, it is unclear how the class M could be used "at the end of the day." A standard error is given without a point estimate, as it were.

West (1984) studies the aggregation problem in the context of a simple decision making situation involving an event, E, assessed individual probabilities $p_1, \cdots, p_n$ for E, and utilities $U_1, \cdots, U_n$. The ith individual faces a gamble X_i which pays him/her A_i, $i = 1, \ldots, n$, or 0 according as E occurs or not. Subject i's certainty equivalent is thus S_i: $U_i(S_i) = p_i U_i(A_i)$ if $U_i(0)$ is taken to be 0 without loss of generality. Hence, the group is indifferent between $\mathbf{X}$ and $\mathbf{S}$, the vector of X_i's and S_i's, respectively.

Assuming the existence of a jointly accepted group utility function U for which $0 < U(\mathbf{S}) = U(\mathbf{X}) < U(\mathbf{A})$, West (1984) shows that there exists a number $0 < p < 1$ such that $pU(\mathbf{A}) = U(\mathbf{S})$. If it is also assumed that U can be expressed as a function of the U_i's, it follows that $p = \prod_{i=1}^{n} p_i^{w_i}$ and that U must be proportional to $\prod_{i=1}^{n} U_i^{w_i}$. West calls p the group belief and emphasizes that $q = \prod_{i=1}^{n} (1 - p_i)^{w_i}$ is the group belief in E^c, so that $p + q < 1$.

West does not take his analysis to support the logarithmic opinion pool. He views it as an argument that group belief cannot be probability. However, let us emphasize that West's analysis is based on the restrictive assumption that the group utility U is expressed as a function of the subjects' U_i's. Furthermore, West's model calls implicitly for each individual's reward to be determined according to their *own* utility function, rather than according to the group's agreed utility. We suspect, therefore, that these assumptions are the source of West's conclusions, and we would argue that his result should be recast as an impossibility theorem. Despite the potential limitation of his model, however, West's analysis illustrates very clearly a point which is becoming increasingly better understood in group decision theory, namely that a group of Bayesians cannot always be fully Bayesian even when its members would want it to be.

In some respects, the work of Genest, Weerahandi, and Zidek (1984) goes even further than West's. Here, the space Θ is allowed to be an arbitrary carrier set such as the space of available decision rules, subject possibly to some feasibility requirement. Alternatively, Θ might index a set of sampling densities. As a means of quantifying the relative degree of support for various elements of Θ, these authors then define what they call *P functions* as everywhere positive functions $b:\Theta \rightarrow (0, \infty)$ on Θ. For example, $b(\theta)$ could be the observed significance level when θ is the null hypothesis that the true state of nature is θ. If Θ is a set of decision rules, then $b(\delta)$ could represent $u(\delta) - u_0$, the expected gain in utility when the decision rule δ is used and δ is required to be Nash-feasible (Nash, 1950) so as to make $b(\delta)$ strictly positive. In other applications, b could be a likelihood function, the density of a diffuse prior, the density of a proper (discrete or continuous) prior, or posterior distribution.

Given n P functions, say $b_1, \cdots, b_n$, Genest, Weerahandi, and Zidek (1984) define the relative propensity profile of a point θ versus another point η in Θ to be

$$RP(b_1, \cdots, b_n; \theta, \eta) = [b_1(\theta)/b_1(\eta), \cdots, b_n(\theta)/b_n(\eta)].$$

They then look for P function pooling operators which are *relative propensity consistent* (RPC), i.e., such that

$$T(b_1, \cdots, b_n)(\theta)/T(b_1, \cdots, b_n)(\eta) \geq T(c_1, \cdots, c_n)(\theta^*)/T(c_1, \cdots, c_n)(\eta^*)$$

whenever

$$RP(b_1, \cdots, b_n; \theta, \eta) \geq RP(c_1, \cdots, c_n; \theta^*, \eta^*).$$

This axiom introduces a kind of scale invariance which is akin to, but stronger than the homogeneity property (5.4) postulated by Aczél and Saaty (1983). If the number of points in the carrier set Θ is at least three, so that Θ is a tertiary space, it then follows that T must be of the form

$$(5.7) \quad T(b_1, \cdots, b_n)(\theta) = C(b_1, \cdots, b_n)g(\theta) \prod_{i=1}^{n} b_i^{w_i},$$

where the w_i's are non-negative constants and g is an unspecified P function. Like (3.14), this is a logarithmic opinion pool with non-negative weights, but the very large domain on which it is defined and the intuitive appeal of the RPC requirement make it possible to apply this result to a variety of unusual types of opinions. Special cases of (5.7) yield, for example, Fisher's celebrated formula for combining significance levels, the product of Nash (1950) for determining an equilibrium in multiperson bargaining problems, as well as the usual rule for combining independent likelihoods. Here again, however, the theory does not prescribe what may meaningfully be pooled. It does not rule out, for instance, the possibility of amalgamating "completely dependent" likelihoods (e.g., produced by French's 1980 "ultimate yes-man"), even though other considerations suggest that this is unwise. In this sense, therefore, the whole theory remains "user-dependent."

6. GROUP INTERACTION MODELS

In the previous three sections, we reviewed a number of normative methods for amalgamating opinions which could be used either when the decision maker is a third person who is consulting a group of experts (Bayesian methods, Section 4) or when the third person is simply the "synthetic personality of the group" (Sections 3 and 5). In both cases, it was implicitly assumed that the group was in a state of "dialectical equilibrium," i.e., a state in which no subsequent discussion among the group members would produce any further change in their opinions. Thus, a pooling formula was used to combine the individual judgments, so that the resulting opinion could be entered into a conventional uniBayesian analysis by the decision maker or what Hogarth (1975) calls the "synthetic personality of the group." It should be emphasized, however, that this synthetic opinion need not be the object of a consensus among the members of the group. In this section, we will briefly survey a number of models in which individuals are allowed to interact with or without feedback, and through dialogue, attempt to reach a consensus.

Dialogue with unrestricted feedback or "group reassessment" (Winkler, 1968) is natural and easy if the individuals are able to communicate with each other. (Recall that this will not necessarily be the case, however; the opinions might have been recorded at different times and places.) Its chief merit, the free exchange of information, may result in a reduction in the range of views. Furthermore, the group may well be synergetic and generate, through its dialogue, relevant new insights that could change their opinions in a dramatic way. But on the negative side, this same interaction may induce conformity, i.e., a degree of agreement beyond that which would be commensurate with the amount of information that is exchanged. Moreover, unrestricted dialogue permits strategic manipulation, bluffing, intimidating tactics, and threats to be employed. Thus, to be effective, the interacting group needs a strong director. A Bayesian model describing "group polarization" effects in expert dialogues has recently been discussed by Bordley (1983a).

The Delphi technique and its variants (Pill, 1971, and references therein) exemplify methods of dialogue with feedback restrictions. In this setup, open discussion is not permitted. The feedback, which may or may not be anonymous, consists in summary statistics such as group means or quantiles when, for example, the task is that of point estimation. Each individual then reassesses his/her distribution and the process is repeated until, hopefully, the different opinions converge toward a common, consensual distribution. This approach can be inexpensive since the individuals need not communicate directly and social pressure is reduced. As pointed out by Winkler (1968), however, it is difficult to limit the feedback to a few summary statistics unless the opinions are known to belong to a specific family of distributions. For this reason, it would seem advisable to use the individual distributions themselves as the feedback.

Attempts have been made to formalize group sharing of distributions using iterative schemes. In one particular model, which was proposed independently by DeGroot (1974) and Lehrer (1976), it is assumed that after k iterations, the ith individual's distribution $P_{(i,k)}$ equals $\sum_{j=1}^{n} w_{(i,j)} P_{(j,k-1)}$, where the $w_{(i,j)} \geq 0$ are fixed weights such that $\sum_{j=1}^{n} w_{(i,j)} = 1$. In matrix notation, we have $P_k = WP_{k-1} = W^k P_0$ for all $k \geq 1$, a model in which feedback is *not* anonymous. If the ith row of W^k converges to $(\pi_1, \cdots, \pi_n)$ as k tends to infinity, "consensus" is said to have been reached, and $\sum_{j=1}^{n} \pi_j P_{(j,0)}$ is the group's distribution. Berger (1981) establishes conditions for the required convergence to take place, and Forrest (1985) discusses the assessment of zero weights in this framework.

In a more general model, W^k is replaced by $W_1 \cdots W_k$ and the W_i's themselves are to be arrived at by a dialogue (Chatterjee, 1975; Chatterjee and Seneta, 1977; Wagner, 1978, 1980). Recently, Cohen, Hajnal, and Newman (1983) have investigated the case where the choice of the matrices W_i is dictated by a probability distribution. Lehrer and Wagner (1981), who have devoted a book to this model, make the controversial claim that under certain circumstances, it provides the *unique* rational way of combining the opinions of experts into a consensus. They argue that even though the experts disagree, they may be *rationally* committed to this type of consensus. This position has been attacked rather fiercely by Baird (1985), Levi (1985), Schmitt (1985), as well as Loewer and Laddaga (1985). Replies to these criticisms about the philosophical implications of the DeGroot-Lehrer model on the "theory of consensus" are offered by Lehrer (1983, 1985) and Wagner (1985). Nurmi (1985) also criticized the method of weighted averaging as a social choice mechanism.

Other iterative-interaction models which have been proposed and analyzed include those of Aumann (1976), Bacharach (1979), and Geanakoplos and Polemarchakis (1982). Press (1978, 1980) offers a more statistical solution to this problem.

7. CONCLUDING REMARKS

In this paper, we have reviewed a host of normative methods for combining the quantitative judgments of several individuals. Our main purpose has been to identify the issues to be addressed and to examine under a critical light the framework within which different solution concepts have been proposed to date. Along the way, we also tried to raise a number of technical points which demand future attention. In this final section, we would like to offer brief comments of a more general nature on the problem of opinion synthesis.

Perhaps the first point to note is that expressing opinions as probability measures seems to lead invariably to arithmetic averaging, while geometric means almost always emerge when odds ratios or densities are used. It may be, therefore, that the form in which the subjects' opinions are expressed plays an unsuspected role in group consensus formation. In view of the difficulties exposed in Section 2 regarding the elicitation and codification of human judgments, this issue deserves our most serious consideration. Is it possible to formulate axioms which would lead to the same consensual opinion irrespective of the form in which the subjects' judgments were encoded? If not, which sorts of opinions should we priviledge? To be useful, an answer to the latter question should be substantiated by the conclusions of experimental research in cognitive psychology (Hogarth, 1975).

Another (obvious?) lesson which we learn from this entire exercise is that the appropriateness of the solution depends in good part on the problem which it addresses. There is a world of difference between the *epistemological* consensus sought by philosophers like Lehrer and the *operational* consensus of group decision theorists. In the first case, the "group" could be all of humanity or the scientific community, and the demonstrable existence of a rational consensus might have profound consequences on the philosophy of science or the whole of human perception. In the second case, consensus of opinion is merely a means of reaching a decision which maximizes the actors' utilities. Agreement now is a matter of convenience only and implies no commitment whatsoever for the future.

Ideally, well directed group interaction with unrestricted feedback would be the best approach to opinion aggregation in the context of decision making. Bargaining theory (Weerahandi and Zidek, 1983) may offer a viable alternative, but the problem mentioned in the introduction of finding an intersubjective equivalent to the classical notion of objectivity remains. The averaging methods which have been described in Sections 3 to 6 of this paper attempt to address this issue, although their success is at best mitigated. A serious obstacle to their implementation is the arbitrariness of the pooling weights, the existence of which is implied by the axiomatic approach. These constants are analogous to the parameters commonly found in physical models and, like physical constants, they would have to be fitted to the particular context in which a formula is being applied. However, the theory lacks a basis on which to carry out such a fit, save for the case described in Section 4 where the pooling formulas have a Bayesian interpretation and their parameters derive from a supra Bayesian's assessment of the subjects' opinions. Only when this problem is resolved can an appropriate analogue of the law of averages be formulated.

When the group is reporting to an external decision maker, the Bayesian paradigm does seem to be the only acceptable method for accumulating the information that lies behind the consultants' conflicting advice. The main concerns, here, are similar to those of the classical Bayesian theory, and the literature has only begun to address them. How can the decision maker elicit his/her own opinion? Can the experts' judgments be taken at face value, or must they be "calibrated?" How is one to assess the dependence between information sources? These and other research directions in group decision making are given by Winkler (1982).

ACKNOWLEDGMENTS

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the National Science Foundation under Grant MCS 83-01807.

REFERENCES

To avoid unnecessary repetitions, references which appear in the annotated bibliography are not listed here.

BEACH, B. H. (1975). Expert judgment about uncertainty: Bayesian decision making in realistic settings. *Org. Behav. Hum. Perform.* **14** 10–59.

BENSON, P. G. and NICHOLS, M. L. (1982). An investigation of motivational bias in subjective predictive probability distributions. *Decision Sci.* **13** 225–239.

BERNARDO, J. M. (1979). Expected information as expected utility. *Ann. Statist.* **7** 686–690.

CHALONER, K. M. and DUNCAN, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *Statistician* **32** 174–180.

DAWID, A. P. (1982a). Intersubjective statistical models. In *Exchangeability in Probability and Statistics* (KOCH, G. and SPIZZICHINO, F., eds). North Holland, Amsterdam.

DAWID, A. P. (1982b). The well-calibrated Bayesian (with discussion). *J. Amer. Statist. Assoc.* **77** 605–613.

DE FINETTI, B. (1970). *Teoria delle probabilità.* Giulio Einaudi editore, Turino. English language edition: *Theory of Probability*, vol. 1 (1974), vol. 2 (1975). John Wiley, New York.

FINE, T. L. (1973). *Theories of Probability.* Academic Press, New York.

FISHBURN, P. C. (1983). Ellsberg revisited: a new look at comparative probability. *Ann. Statist.* **11** 1047–1059.

FRENCH, S. (1982). On the axiomatisation of subjective probabilities. *Theory and Decision* **14** 19–33.

GOKHALE, D. V. and PRESS, S. J. (1982). Assessment of a prior distribution for the correlation coefficient in a bivariate normal distribution. *J. R. Statist. Soc. Ser. A* **145** 237–249.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934). *Inequalities.* Cambridge University Press, Cambridge.

KADANE, J. B., DICKEY, J. M., WINKLER, R. L., SMITH, W. S. and PETERS, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *J. Amer. Statist. Assoc.* **75** 845–854.

KEENEY, R. L. and RAIFFA, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley, New York.

KRAFT, C. H., PRATT, J. W. and SEIDENBERG, A. (1959). Intuitive probability on finite sets. *Ann. Math. Statist.* **30** 408–419.

LEHRER, K. (1975). Social consensus and rational agnoiology. *Synthese* **31** 141–160.

LINDLEY, D. V. (1982). Scoring rules and the inevitability of probability (with discussion). *Int. Statist. Rev.* **50** 1–26.

NASH, J. F., Jr. (1950). The bargaining problem. *Econometrica* **18** 155–162.

PILL, J. (1971). The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Econ. Plan. Sci.* **5** 57–71.

PRESS, S. J. (1978). Qualitative controlled feedback for forming group judgment and making decisions. *J. Amer. Statist. Assoc.* **73** 526–535.

PRESS, S. J. (1980). Bayesian inference in group judgment formulation and decision making using qualitative controlled feedback. *Proceedings of the First Valencia International Meeting on Bayesian Statistics*, 397–414.

ROBERTS, H. V. (1965). Probabilistic prediction. *J. Amer. Statist. Assoc.* **60** 50–62.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* John Wiley, New York.

SEIDENFELD, T. (1979). Why I am not an objective Bayesian: some reflections prompted by Rosenkrantz. *Theory and Decision* **11** 413–440.

SHAFER, G. R. (1976). *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ.

SHAFER, G. R. (1983). A subjective interpretation of conditional probability. *J. Phil. Logic* **12** 453–466.

SLOVIC, P., FISCHOFF, B. and LICHTENSTEIN, S. (1977). Behavioral decision theory. *Annu. Rev. Psychol.* **28** 1–39.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision. *J. R. Statist. Soc. Ser. B* **23** 1–37.

STONE, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71** 114–125.

TVERSKY, A. and KAHNEMAN, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* **185** 1124–1131.

VINCENT, S. B. (1912). The function of the viborissae in the behavior of the white rat. *Behav. Monogr.* **1** No. 5.

WALLEY, P. (1981). *Coherent Lower (and Upper) Probabilities.* Statistics Research Report, University of Warwick.

WALLEY, P. and FINE, T. L. (1982). Towards a frequentist theory of upper and lower probability. *Ann. Statist.* **10** 741–761.

WEERAHANDI, S. and ZIDEK, J. V. (1983). Elements of multi-Bayesian decision theory. *Ann. Statist.* **11** 1032–1046.

WINKLER, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *J. Amer. Statist. Assoc.* **62** 776–800.

WINKLER, R. L. (1971). Probabilistic prediction: some experimental results. *J. Amer. Statist. Assoc.* **66** 675–685.

WINKLER, R. L. (1982). Research directions in decision making under uncertainty (with discussion). *Decision Sci.* **13** 517–553.

ANNOTATED BIBLIOGRAPHY

The bibliography contains 92 entries and was up to date at the time of submission. It is intended only as a *guide* to the vast theoretical literature on the problem of pooling opinions. In particular, we cannot claim that it is comprehensive or entirely impartial, although every effort has been made to include all relevant contributions. For references to the literature on the psychological aspects of group probability assessments, the Delphi method, and the usefulness of expert assessments in applied research, the reader should consult Hogarth (1975, 1977), Pill (1971), and Beach (1975), respectively.

Wherever possible, each article in the following list has been classified according to the approach which it adopts, as well as the sort of opinions that it considers. Papers marked with an * include a review of the literature and/or an extensive bibliography.

Approach Adopted

AT: Axiomatic treatment (especially with regard to group consensus belief formation)
BU: Bayesian updating of opinion (in the presence of a decision maker)
CI: Consensus reached iteratively (group interaction)
DM: Decision making aspects are stressed (e.g., considerations of utilities are included)
RC: Concerned with the reconciliation of probability assessments (involves only one individual)
UB: Use of bargaining theory to reach a joint decision

Expression of Opinions

C: Cumulative distribution functions
D: Probability density functions
L: Odds ratios or log odds
M: Probability measures
P: Discrete probabilities

The abbreviation **AL** is used when a paper is concerned with the somewhat different problem of "allocation of resources." For a description, see Aczél and Wagner (1981).

ACZÉL J. (1984). On weighted synthesis of judgements. *Aequat. Math.* **27** 288–307.

AT; L Although the author provides some new evidence in favor of the logarithmic opinion pool in cases where the expert judgments are expressed in the form of odds, this paper should be viewed primarily as a contribution to the theory of functional equations. Aczél's joint results with Saaty (1983) and Alsina (1984) are generalized to the case where experts are not treated symmetrically. An extra hypothesis is added which guarantees that the aggregating function is locally sensitive to each of the individual opinions. The potential readers must expect mathematical hurdles such as cancellative semigroups.

ACZÉL, J. and ALSINA, C. (1984). Characterizations of some classes of quasilinear functions with applications to triangular norms and to synthesizing judgements. *Methods Oper. Res.* **48** 3–22.

AT; L The functional equations solved here are motivated by and generalize those which were studied in Aczél and Saaty (1983). However, this article says very little about the pooling problem per se.

ACZÉL, J., KANNAPPAN, P. L., NG, C. T. and WAGNER, C. G. (1981). Functional equations and inequalities in rational group decision making. *General Inequalities 3: Proceedings of the Third International Conference on General Inequalities*, Oberwolfach.

AT; AL, P Reformulation of the results exposed in two previous papers by Aczél and Wagner (1980, 1981), but with resolution of the case in which only two resources are to be allocated. This corresponds to the situation faced when the underlying probability space is not tertiary.

ACZÉL, J., NG, C. T. and WAGNER, C. G. (1984). Aggregation theorems for allocation problems. *SIAM J. Alg. Disc. Meth.* **5** 1–8.

AT; AL, P This fourth collaborative paper between Aczél and Wagner drops the "consensus of rejection" hypothesis (the ZPP, condition (3.3)). This leads to a generalized linear opinion pool of the form (3.7) for allocation problems. See also Genest (1984c).

ACZÉL, J. and SAATY, T. L. (1983). Procedures for synthesizing ratio judgements. *J. Math. Psych.* **27** 93–102.

AT; L By assuming that the influence of each individual opinion can be "separated" (condition (5.1)) and that the experts can be treated symmetrically, the authors characterize a large class of aggregating functions for odds ratios (line (5.3)). All the formulas they present obey the reciprocal property $f(x_1, \cdots, x_n)^{-1} = f(1/x_1, \cdots, 1/x_n)$ and preserve unanimity. When an additional homogeneity requirement is imposed, the symmetric geometric mean emerges as the unique solution.

ACZÉL, J. and WAGNER, C. G. (1980). A characterization of weighted arithmetic means. *SIAM J. Alg. Disc. Meth.* **1** 259–260.

AT; AL, P Axiomatic derivation of weighted arithmetic means within the context of group allocation of resources. When the "resource" is a probability mass, the authors' so-called k-allocation property amounts to saying that pooling probability distributions must yield a probability distribution. The other axioms used are equivalent to the SSFP and the ZPP, i.e., conditions (3.2) and (3.3).

ACZÉL, J. and WAGNER, C. G. (1981). Rational group decision making generalized: the case of several unknown functions. *C. R. Math. Rep. Acad. Sci. Can.* **3** 139–142.

AT; AL, P Sequel to the above paper in which the same result is derived under a weaker assumption that reduces to the WSFP, condition (3.5), when probability distributions are pooled.

AGNEW, C. E. (1985). Multiple probability assessments by dependent experts. *J. Amer. Statist. Assoc.* **80** 343–347.

BU; D The Bayesian updating model of Winkler (1968) is extended to account for situations in which the experts are providing assessments about several random variables. When assessment errors have a multivariate normal density, the posterior distribution depends only on a weighted average of the experts' stated means. Special cases of interest are those in which the individual's distributions are dependent but the random variables are not, and vice versa.

AUMANN, R. J. (1976). Agreeing to disagree. *Ann. Statist.* **4** 1236–1239.

BU; P Formalizes the idea of common knowledge and uses this definition to prove a theorem asserting that if two people have the same prior, and if their posteriors for an event are common knowledge, then these posteriors are equal. "Might be considered a theoretical foundation for the reconciliation of subjective probabilities" (p. 1238). See also Geanakoplos and Polemarchakis (1982), as well as Shafer (1983).

BACHARACH, M. (1972)*. *Scientific Disagreement.* Unpublished manuscript, Christ Church, Oxford.

AT, CI, DM; D This penetrating essay articulates some of the philosophical and psychological assumptions underlying the theory of consensus, whether it be concerned with probabilities or with utilities. After criticizing past proposals for the resolution of difference of opinion, the author sets up a general model describing the structure of individual beliefs and their interplay within a group. An iterative method of "Bayesian dialogue" (*cf.* also Bacharach, 1979) ensues and questions of convergence are examined. It is shown, among other things, that the consensus is a logarithmic opinion pool with equal weights when disagreement between the experts is uniquely attributable to different observations.

BACHARACH, M. (1975). Group decisions in the face of differences of opinion. *Manag. Sci.* **22** 182–191.

AT, DM; D Obtains two characterizations of the linear opinion pool by imposing conditions on the group preference relation. Among these conditions, the most important are Pareto optimality and an independence of irrelevant alternatives hypothesis. An impossibility theorem akin to that of Zeckhauser (1968) is also derived.

BACHARACH, M. (1979). Normal Bayesian dialogues. *J. Amer. Statist. Assoc.* **74** 837–846.

BU, CI; P A study of the interaction of opinions which are differing because, and *only* because, their defenders have made different observations. Formulates the concept of a "Bayesian dialogue" at each stage of which one or more features of the individual distributions are shared and incorporated into the opinion of each group member. Under normality, it is shown that the parts of the subjects' distributions that are iteratively revealed converge to a consensus that embodies all private information, and the rate of convergence is obtained.

BAIRD, D. (1985). Lehrer-Wagner consensual probabilities do not adequately summarize the available information. *Synthese* **62** 47–62.

CI; P A critique of Lehrer and Wagner (1981). As the title indicates, the author argues that a single probability distribution cannot adequately summarize the information available to the members of a group. He suggests the use of nonprobabilistic reports of consensus, which could include a measure of dispersion in addition to the usual average. On this point, see also Loewer and Laddaga (1985).

BERGER, R. L. (1981). A necessary and sufficient condition for reaching a consensus using DeGroot's method. *J. Amer. Statist. Assoc.* **76** 415–418.

CI; C Points out that the convergence conditions given in the paper of DeGroot (1974) can be weakened when the prior opinions of the group are already given and it is asked whether the DeGroot-Lehrer iterative process will yield a consensus for this particular set of priors. A convenient computational shortcut is also provided.

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

AT; L Derives a version of the logarithmic opinion pool using axioms from the theory of additive conjoint measurement. It is assumed that the decision maker has an intuitive weak ordering on the set of expert odds ratios, and that a "noninteraction" property similar to (5.1) holds. The problem of determining the weights is given some attention.

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

AT; L The author's 1982 version of the logarithmic opinion pool is offered once again as a model for how individual attitudes tend to become more polarized as a result of group discussion. This is what is sometimes referred to as the risky (or cautious) shift phenomenon. Similar to the discussion in Section 4 of the earlier paper, but with a different focus.

BORDLEY, R. F. (1983b)*. *Bayesian Group Decision Theory.* Conference on Information Pooling and Group Decision Making, Irvine, CA, March 1983.

This review paper focuses on two different aspects of group decision making: aggregation of individual opinions and aggregation of individual preferences. As demonstrated by the author, the two must be kept separate in order to preserve the group maximum expected utility criterion. A parallel is drawn in the event case between the pooling recipe of Bordley (1982) and the product formula which results from the Bayesian approach developed by Morris (1974, 1977). Probability amalgamation is also briefly considered as a problem of parameter estimation. (The section devoted to the separate but related issue of determining a group utility function is very informative; it should definitely be read for the valuable insights it provides.)

BORDLEY, R. F. and WOLFF, R. W. (1981). On the aggregation of individual probability estimates. *Manag. Sci.* **27** 959–964.

AT; P Offers a critical review of Norman Dalkey's work on aggregating opinions. The authors argue, among other things, against what they term the "context-free assumption" (e.g., the SSFP property) according to which the group estimate of an event's probability is a function only of the estimates of that event provided by the individuals.

CHATTERJEE, S. (1975). Reaching a consensus; some limit theorems. *Proc. Int. Statist. Inst.* 159–164.

CI; C Considers a generalization of the DeGroot-Lehrer model in which the individuals can change their weights at each iteration. Opinions here can be the value of an unknown parameter (such as in forecasting) as well as probability distributions. See also the paper just below and Wagner (1978).

CHATTERJEE, S. and SENETA E. (1977). Towards consensus: some convergence theorems on repeated averaging. *J. Appl. Prob.* **14** 89–97.

CI; C Extends and makes rigorous the results of Chatterjee (1975). The problem of the tendency to consensus is connected with the ergodicity problem for backward product of stochastic matrices.

CLEMEN, R. T. (1984). *Modeling Dependent Information: a Bayesian Approach.* Ph.D. dissertation, Indiana University.

CLEMEN, R. T. (1986). Calibration and the aggregation of probabilities. *Manag. Sci.* **32** in press.

BU; D A criticism of the multiplicative formula developed by Morris (1977, 1983). The author demonstrates that the notion of "joint calibration" on which the formula rests is equivalent to independence between the decision maker's opinion and those of the experts. Thus it is argued that what Morris calls joint calibration is equivalent to, and no easier to carry out than the direct assessment of a likelihood for the experts' probabilities, conditionally upon the decision maker's own prior information. For further comments on Morris' definition of calibration, see Schervish (1983).

CLEMEN, R. T. and WINKLER, R. L. (1985). Limits for the precision and value of information from dependent sources. *Oper. Res.* **33** 427–442.

BU, DM; D Investigates the way in which the presence of dependence in the model of Winkler (1981) impacts the decision maker's posterior distribution. It is seen that dependence among experts severely reduces the amount of information the decision maker receives, as measured by the variance of his/her posterior. The authors develop the notion that a number of dependent experts can be considered equivalent, in an information content sense, to a generally smaller number of independent experts. The expected value of information from dependent sources is also studied.

COHEN, J. E., HAJNAL, J. and NEWMAN C. M. (1983). Approaching consensus can be delicate when positions harden. Unpublished manuscript.

CI; C This high pitched mathematical paper gives examples in which the ergodic behavior of a nonstationary product of random non-negative matrices depends discontinuously on a continuous parameter. The examples are interpreted in terms of the DeGroot-Lehrer iterative model of consensus. At each stage, random weights are assigned by all the experts to themselves and their peers. When these weights fall between certain lower and upper bounds, the authors show how a very small change in the weighting process can mean a difference between moving toward a consensus with probability 1, or remaining in dissension with probability 1.

DALKEY, N. C. (1972). An impossibility theorem for group probability functions. P-4862, The Rand Corporation, Santa Monica, CA.

AT; P This paper's main result is stated with a sketch of proof in Dalkey (1975).

DALKEY, N. C. (1975). Toward a theory of group estimation. In *The Delphi Method: Techniques and Applications.* (LINSTONE, H. A. and TUROFF, M., eds). Addison-Wesley, Reading, MA.

AT; P In the first part of this essay, Professor Dalkey attempts to define a number of desirable features of expert opinions: honesty, accuracy, definiteness, realism, certainty and freedom from bias. Assuming that expert judgments are "the best information available," the author then proceeds to describe three approaches to aggregation. Two versions of the logarithmic opinion pool are thus obtained, one (each weight equals $1/n$) by analogy to the "theory of errors," and the other (each weight equals 1) by an application of Bayes' theorem to an expression measuring the degree of dependence among the experts. In the final section, entitled "The Axiomatic Approach," the author's 1972 impossibility theorem is sketched and commented upon.

DALKEY, N. C. (1977). Group decision theory. Technical Report No. 7749, School of Engineering and Applied Science, University of California, Los Angeles, CA.

DE FINETTI, B. (1954). Media di decisioni e media di opinioni. *Bull. Inst. Int. Statist.* **34** 144–157.

DM; P Using a simple hypothesis problem, a theorem of Wald on Bayes decision rules (*Ann. Math. Statist.* **10** 299–326, 1939) is interpreted in the context of group decision making. It is seen that if all the members of a group share the same utilities, there must exist a decision based on an "average opinion" which all the individuals will prefer to an "average decision" based on those which each individual would found upon their own judgement. Compare with Stone (1961).

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

CI; C Exploits the theory of Markov chains to determine under which circumstances a consensus will be reached iteratively when each member of a group updates his/her opinion using a linear opinion pool with fixed, non-negative weights. See Lehrer (1976) for an independent consideration of this model, and Berger (1981) for weaker convergence conditions which can be used when only one set of prior opinions is considered.

DICKEY, J. M. and FREEMAN, P. (1975). Population-distributed personal probabilities. *J. Amer. Statist. Assoc.* **70** 362–364.

CI; P Uses the Dirichlet distribution as a model for studying the variability of personal probabilities under coherent transformations induced by a common likelihood. A tendency to consensus is seen to result from the availability of increasing amounts of data.

EISENBERG, E. and GALE, D. (1959). Consensus of subjective probabilities: the pari-mutuel method. *Ann. Math. Statist.* **30** 165–168.

AT; P Suggests the pari-mutuel betting method as a principle for combining opinions. Assuming that each "bettor" has a fixed budget and a fixed private opinion, the authors show the existence and uniqueness of equilibrium totalisator odds. An example shows the existence of individual opinions which allow their holders to dictate the equilibrium odds.

FISCHER, G. W. (1981). When oracles fail—A comparison of four procedures for aggregating subjective probability forecasts. *Org. Behav. Hum. Perform.* **28** 96–110.

Experimental comparison of an arithmetic average of probability forecasts with the Delphi procedure, face-to-face discussion, and another procedure called "estimate-talk-estimate."

FORREST, P. (1985). The Lehrer/Wagner theory of consensus and the zero weight problem. *Synthese* **62** 75–78.

A short critique of Lehrer's suggestion (*cf.* Lehrer and Wagner, 1981, p. 20) that a zero weight should be assigned to a person's opinion whenever a merely random selection would be preferred to that person's guidance.

FREELING, A. N. S. (1981). Reconciliation of multiple probability assessments. *Org. Behav. Hum. Perform.* **28** 395–414.

RC; L Summarizes and discusses the method of reconciliation proposed by Lindley, Tversky, and Brown (1979) in view of its practicability. In certain circumstances, this method is seen to be equivalent to a logarithmic opinion pool, with possibly negative weights. The author goes on to suggest, without being specific, that the approach could be adapted to yield the same formula in the context of group probability assessments.

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

BU; P Some of the difficulties associated with updating one's beliefs using a correlated opinion are described. The model of Lindley, Tversky, and Brown (1979) is modified accordingly. The resulting prescription is formula (4.1), a logarithmic opinion pool completely equipped with interpretable coefficients.

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

This review paper vindicates the use of a strictly Bayesian approach to expert resolution. Criticisms of the linear and logarithmic opinion pools are offered in the light of the author's 1980 normal model for modifying individual beliefs using Bayes' rule. Conditions are derived under which successive iterations of this model will provide each member of the group with a final opinion, but the procedure demonstrably falls short of creating a consensus.

FRENCH, S. (1985)*. Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2* (BERNARDO, J. M., ET AL., eds). North Holland, Amsterdam, 183–201.

BU, DM; P, MA This paper distinguishes between three versions of the expert problem and reviews a number of related issues and principles, including marginalization, external Bayesianity and calibration. In the first part, the author argues in favor of the Bayesian paradigm for situations in which a decision maker collects the opinions of a number of individuals. After giving some consideration to the axiomatic approaches of Bordley (1982) and Morris (1983), French shows how his own Bayesian model (1980, 1981) obeys a variation of the externally Bayesian criterion in which the known value of the likelihood for the data is filtered out of the experts' posterior beliefs. In the second part, French puts under close scrutiny the vast literature on group decision-making per se. He stresses the role played by the concepts of honesty (*cf.* Dalkey, 1975) and calibration (*cf.* Schervish, 1983) in iterative methods of consensus. Finally, a brief mention is made of the expert problem which French dubs the "text-book" problem.

GEANAKOPLOS, J. D. and POLEMARCHAKIS, H. M. (1982). We can't disagree forever. *J. Econ. Theory* **28** 192–200.

CI; P Extends the result of Aumann (1976) to show that if two agents simply communicate their posterior beliefs back and forth, then they will be led to make revisions that converge, in finitely many steps, to a common, equilibrium posterior. Shafer (1983) provides a natural setting for these theorems.

GENEST, C. (1983). Towards a consensus of opinion. Ph.D. dissertation, University of British Columbia.

GENEST, C. (1984a). A conflict between two axioms for combining subjective distributions. *J. R. Statist. Soc. Ser. B* **46** 403–405.

AT; D This note points out that an externally Bayesian group must concentrate all its power in the hands of a single person if consensus is to be determined using a function which depends only locally on the individual probability densities. A suggestion aimed at resolving this conflict was subsequently investigated by Genest (1984b).

GENEST, C. (1984b). A characterization theorem for externally Bayesian groups. *Ann. Statist.* **12** 1100–1105.

AT; D Determines conditions under which an externally Bayesian group must use a logarithmic opinion pool. The main condition states that the consensual density at each point should be proportional to a function of the individual densities at that point, i.e., condition (3.13). A discussion of the implications of this result is included. A suggestion regarding a possible extension of the theorem was picked up in Genest, McConway, and Schervish (1986).

GENEST, C. (1984c). Pooling operators with the marginalization property. *Can. J. Statist.* **12** 153–163.

AT; M The marginalization property of McConway (1981) (without zero preservation) is shown to imply a consensus which is a weighted average of the subjects' opinions plus an arbitrary, imposed component. The weights can be negative and must satisfy certain consistency inequalities. The result is then used to extend a theorem of Wagner (1984) that the only pooling procedures which preserve independence are dictatorial in nature. It is argued that independence preservation is not a reasonable requirement to impose on consensus-finding procedures.

GENEST, C., MCCONWAY, K. J. and SCHERVISH, M. J. (1986). Characterization of externally Bayesian pooling operators. *Ann. Statist.* **14** in press.

AT; P, D Basically, this is an extension of the result contained in Genest (1984b). Externally Bayesian pooling operators are characterized without resorting to any regularity condition whatsoever. A condition is given under which the generalized linear opinion pool (3.14) emerges.

GENEST, C. and SCHERVISH, M. J. (1985). Modeling expert judgments for Bayesian updating. *Ann. Statist.* **13** 1198–1212.

BU, AT; P The authors examine how a decision maker could go about using the Bayesian paradigm to incorporate expert opinions into his/her own, even though he/she might be incapable or reluctant to assess completely his/her beliefs about the value of their judgments. In certain cases, the result is a linear opinion pool with possibly negative weights. In others, such as when the experts are deemed to be conditionally independent, the log-type formula (4.4) emerges. Discussion.

GENEST, C. and WAGNER, C. G. (1984). Further evidence against independence preservation in expert judgement synthesis. Technical Report No. 84-10, Department of Statistics and Actuarial Science, University of Waterloo.

AT; P Extends previous impossibility theorems on independence preservation (*cf.* Wagner, 1984) to large classes of normalized pooling operators which include the logarithmic opinion pool. An independence preserving formula is also derived when the underlying space contains exactly four points. A short *plaidoyer* against preservation axioms in general concludes the paper. See also Genest (1984c).

GENEST, C., WEERAHANDI, S. and ZIDEK, J. V. (1984). Aggregating opinions through logarithmic pooling. *Theory and Decision* **17** 61–70.

AT Contains results on the synthesis of opinions quantified as arbitrary positive functions on a carrier set. The authors prove that positing a single, often intuitively reasonable axiom of aggregation (which preserves the ordering of certain ratios) restricts the acceptable methods of pooling to a class of more or less geometric means given by (5.7). A logarithmic opinion pool with variable weights is also derived in the finite case using a variation of the concept of external Bayesianity.

HOGARTH, R. M. (1975)*. Cognitive processes and the assessment of subjective probability distributions (with discussion). *J. Amer. Statist. Assoc.* **70** 271–294.

Excellent survey of some recent findings in the psychology of judgment and their implications for the assessment of subjective distributions. Section 3.5 reviews the theoretical developments in group probability assessments and supplies a number of references to relevant experimental work. The main empirical results are summarized and commented upon.

HOGARTH, R. M. (1977)*. Methods for aggregating opinions. In *Decision Making and Change in Human Affairs* (JUNGERMANN, H. and DEZEEUW, G., eds). D. Reidel Publishing Co., Dordrecht, Holland.

Reviews models and methods for aggregating opinions in the form of point estimates as well as probability distributions. Well worth reading for its consideration of relevant literature in social psychology, even though Section 3 on group probability assessments is largely repetitive of the author's 1975 paper.

LADDAGA, R. (1977). Lehrer and the consensus proposal. *Synthese* **36** 473–477.

AT; P Criticism of the linear opinion pool founded on the observation that it does not generally preserve the independence of events, condition (3.4), even though such independence may have been agreed to by all the individuals. On this point, refer to Lehrer and Wagner (1983), Genest (1984c), or Genest and Wagner (1984). The suggestion is made that probability statements be decomposed into their most elementary parts before attempting to reconcile divergence of opinions.

LEHRER, K. (1976). When rational disagreement is impossible. *Noûs* **10** 327–332.

CI; P A philosopher's independent consideration of the model described by DeGroot (1974), with some discussion of its underlying assumptions.

LEHRER, K. (1983). Rationality as weighted averaging. *Synthese* **57** 283–295.

Philosophical essay defending weighted averaging as the only rational representation of the "totality of information" possessed by an individual or a group of individuals. The mathematical justification behind this assertion is basically Theorem 7 of Wagner (1982), i.e., the derivation of the linear opinion pool based on condition (3.2).

LEHRER, K. (1985). Consensus and the ideal observer. *Synthese* **62** 109–120.

Lehrer's rejoinder, the last item in an issue of *Synthese* which is completely devoted to his consensus method and the book he wrote on this topic in collaboration with Carl Wagner (1981).

LEHRER, K. and WAGNER, C. G. (1981)*. *Rational Consensus in Science and Society.* D. Reidel Publishing Co., Dordrecht, Holland.

The collected works of these two authors on the consensus problem. The first part, written by Lehrer, discusses the philosophical implications and the applications of an elementary mathematical model of consensus (an iterative procedure based on the linear opinion pool which was also formulated by DeGroot, 1974). In the second, shorter part, Wagner develops the foundations of the so-called extended model, in which different weights are allowed at each stage. New here is the noniterative argument on behalf of employing as consensual weights the elements of a fixed-point weight vector. Chapter 6 of the book contains a characterization of the linear opinion pool which can also be found in Wagner (1982).

LEHRER, K. and WAGNER, C. G. (1983). Probability amalgamation and the independence issue: a reply to Laddaga. *Synthese* **55** 339–346.

AT; P Offers a rigorous proof that linear opinion pools do not preserve independence. Argues that independence preserving pooling operators are undesirable on the grounds that independence of events is frequently fortuitous and not associated with "prior theoretical commitment" on the part of the assessor. Compare with Genest (1984c) and Genest and Wagner (1984).

LEVI, I. (1985). Consensus as shared agreement and outcome of inquiry. *Synthese* **62** 3–11.

In this paper, Levi attempts to define two types of consensus, whence the title. He argues that whether preservation axioms are sensible or not depends largely on which one of his two types of consensus is being sought. As he notes, however, "according to strict Bayesian dogma . . . there can be no analogue in contexts of probability judgment of the two senses of consensus I identify. . . " This article will probably be difficult to read if your training lies largely in the mathematical sciences.

LINDLEY, D. V. (1983). Reconciliation of probability distributions. *Oper. Res.* **31** 866–880.

BU; D This paper overlaps with that of Winkler (1981). It aims to show how a decision maker can update his/her opinion of an uncertain quantity after several subjects revealed their mean and standard deviation of their distributions for this quantity. To do this, the decision maker's beliefs about the experts are assumed to be normally distributed, but no assumption is made about the experts' opinions themselves. When the parameters are estimated, the resulting opinion is a t distribution. The advantages of this procedure are illustrated and the role of the information provided by the standard deviations is clarified.

LINDLEY, D. V. (1985). Reconciliation of discrete probability distributions. In *Bayesian Statistics 2* (BERNARDO, J. M., ET AL., eds). North Holland, Amsterdam, 375–390.

BU; P Contends that the only available normative approach to consensus requires reduction to a single decision maker to produce the probabilities. In particular, the marginalization property, condition (3.6), and external Bayesianity, condition (3.12), are looked upon as mere "adhockeries." Instead, a normal model is offered as a consequence of which the expert opinions should be averaged linearly, but on the logarithmic scale. Implementation of this method is discussed when there is symmetry within the set of events of interest.

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

RC, BU; P This paper is mainly concerned with the problem of how to reconcile probability assessments that are incoherent or mutually inconsistent. As noted by several discussants and the authors themselves, however, the methods introduced here could easily be applied to opinion pooling. This is discussed in greater detail by French (1980, 1981) and Freeling (1981).

LOEWER, B. and LADDAGA, R. (1985). Destroying the consensus. *Synthese* **62** 79–95.

CI, BU; P Another attack on the "rationality" of reaching a consensus iteratively using weighted averages. Bayes' theorem is promoted as an alternative, more sensible rule for updating one's opinion and it is argued that there is no reason to iterate that procedure. On this point, see Geanakoplos and Polemarchakis (1982). Since the DeGroot-Lehrer method produces a probability distribution which may conflict with whatever consensus already exists (e.g., failure to preserve independence as noted by Laddaga, 1977), it is suggested that "consensus" should be represented by a set of probability assignments rather than by a single one. (This suggestion is also made by Baird, 1985.)

MADANSKY, A. (1964). Externally Bayesian groups. RM-4141-PR, The Rand Corporation, Santa Monica, CA.

Revised and expanded in 1978; *cf.* below.

MADANSKY, A. (1978). Externally Bayesian groups. Unpublished manuscript, University of Chicago.

AT, DM; D It is argued that if the members of a group share a common likelihood function, compounding their posteriors should yield the same result obtained by first aggregating the priors and then applying Bayes' rule. This is condition (3.12) in our text. A number of group decision rules are then examined and pooling formulas which have this property are called "externally Bayesian." Madansky also determines the manner in which the weights in the linear pool must be updated in order to achieve external Bayesianity. Compare with Roberts (1965).

MCCONWAY, K. J. (1978). The combination of experts' opinions in probability assessment: some theoretical considerations. Ph.D. dissertation, University College London.

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

AT; M The zero preservation property, condition (3.3), and the assumption that finding the consensus commutes with marginalization of the individual distributions, condition (3.6), together imply that the consensus must be found using a linear opinion pool. The result is discussed in the light of the celebrated impossibility theorem of Dalkey (1972).

MORRIS, P. A. (1971). Bayesian expert resolution. Ph.D. dissertation, Stanford University.

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

BU; D, P Formulates a theory of expert use which is entirely consistent with the Bayesian philosophy. Here, each expert probability distribution is treated as a random variable whose value is to be revealed to the decision maker. To obtain the consensus distribution, this decision maker must then proceed to introspect a likelihood function representing his/her assessment of the different experts' knowledge and combine their opinions with his/her own using Bayes' rule. Some examples illustrate the mechanics of the theory.

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

BU; D Sequel to the 1974 paper in which a pooling formula of a multiplicative nature is derived in the case of the location-scale family. The idea of calibration is introduced and a method for subjectively calibrating an expert is also presented. Some fictitious examples illustrate the results. Morris' definition of "joint calibration" has been criticized by Schervish (1983) and Clemen (1986).

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

AT, BU; P, D An axiomatic approach to combining expert judgments which supports the use of a multiplicative formula derived by the author in his 1977 paper. The central requirement is similar in spirit to the concept of external Bayesianity due to Madansky (1964, 1978). According to Schervish (1983), however, Morris' system of axioms would appear to be inconsistent.

Moskowitz, H. and Bajgier, S. M. (1978). Validity of the DeGroot model for achieving consensus in panel and Delphi groups. Unpublished manuscript, Purdue University.

CI An empirical evaluation of the predictive, descriptive, and normative values of the DeGroot-Lehrer model of consensus in panel discussions and Delphi groups. The results presented suggest that individuals revise their judgments using linear opinion pools, although the weights used changed significantly over the first few iterations. The distributions produced by the model tended to be more highly dispersed than the actual reassessed distributions. "Given the human tendency to be overconfident and a need to avoid surprises, a DeGroot model seems like a reasonable rule" (p. 18).

Norvig, T. (1967). Consensus of subjective probabilities: a convergence theorem. *Ann. Math. Statist.* **38** 221–225.

CI; P This sequel to the Eisenberg and Gale (1959) paper proposes a dynamic model for reaching a pari-mutuel type of consensus.

Nurmi, H. (1985). Some properties of the Lehrer-Wagner method for reaching rational consensus. *Synthese* **62** 13–24.

CI, DM; P Examines some properties of the DeGroot-Lehrer method of consensus from a decision making point of view. When the weight matrix is fixed, the method is shown to be monotonic in preferences, and although it leads to Pareto optimal decisions, the author shows that it does not satisfy the so-called Condorcet winner and loser criteria. If von Neumann-Morgenstern utilities are used, the model is also seen to violate a weak axiom of revealed preference and is path dependent as a social choice selection mechanism.

Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Addison-Wesley, Reading, MA.

DM; P Part II of Chapter 8 is devoted to group decision making. Section 11 provides a nice summary of the work of Madansky (1964), illustrating what can happen when a probability aggregation method fails to commute with the process of updating probabilities upon arrival of new, jointly perceived information. In Sections 12 and 13, paradoxes are described where a number of experts who might share a common utility function agree on the choice of an action which is not optimal with respect to their reconciled feelings. A result due to Zeckhauser (1968) is stated (*cf.* also Bacharach, 1975) and the issue of independence preservation is briefly raised.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psych. Bull.* **86** 446–461.

AT; C This article resurrects a pooling procedure named "Vincent averaging" after the biologist S. B. Vincent who introduced it at the beginning of the century (Vincent, 1912). Translated into the language of opinion aggregation, Vincent's original idea was to generate a group opinion about some real valued quantity by taking an arithmetic average of the quantiles across distributions. Ratcliff points out that the Vincentized average of n exponential, logistic, or Weibull distributions belongs to the same family. This is not true of the gamma distribution, however. Vincentization was studied in greater detail by Thomas and Ross (1980).

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

Section 10 of this paper provides some useful "armchair" comments on the problem of consulting experts.

Schervish, M. J. (1983). Combining expert judgments. Technical Report No. 294, Department of Statistics, Carnegie-Mellon University.

BU; D The approach to expert use developed by Morris (1977, 1983) is examined under a critical light. In particular, the system of axioms proposed by Morris (1983) in support of the linear opinion pool is seen to be flawed. Important discussion of the concept of calibration in connection with the treatment of expert probability assessments. See also Clemen (1986).

Schmitt, F. F. (1985). Consensus, respect, and weighted averaging. *Synthese* **62** 25–46.

CI, AT; P Focuses on the work of Lehrer and Wagner (1981). Their iterative method for reaching a consensus using linear opinion pools is criticized on philosophical grounds. It is argued, in a nonmathematical way, that although the average of a number of expert judgments may form an adequate summary of their opinions, the group members who want to update their beliefs in the light of the others' are "rationally committed" *not* to average.

Staël Von Holstein, C-A. S. (1970). Assessment and evaluation of subjective probability. Economics Research Institute at the Stockholm School of Economics.

AT; D Chapter 6 of this book reviews the various approaches to the aggregation problem which were available as of 1970. Problems associated with the choice of weights in the linear pool are addressed and new weighting rules based on the assessors' past performance are offered. Some results obtained using these weighting schemes are discussed in Sections 10.7 and 11.8, where the number of experts to be included in a prediction group is also examined.

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

DM; D Repeatedly cited for the use of the term "opinion pool," this paper supplies conditions under which a decision based on the linear opinion pool will yield an improvement over at least one of the individual optimal decisions. A common utility function is assumed and calculations are performed with respect to the true, unknown distribution of the process. See also de Finetti (1954).

Thomas, E. A. C. and Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *J. Math. Psych.* **21** 136–152.

AT; C This follow-up article on the work of Ratcliff (1979) focuses on the Vincentization pooling procedure. The authors show that location scale families alone are closed under this aggregation method. They also suggest a generalized Vincentization procedure based on the notion of quasiarithmetic mean discussed in Hardy, Littlewood, and Pólya (1934).

TODA, M. (1956). Information-receiving behavior of man. *Psych. Rev.* **63** 204–212.

AT; P In a section entitled "Synthesis of Information," the linear opinion pool is put forward on the basis of expected information-loss minimization.

WAGNER, C. G. (1978). Consensus through respect: a model of rational group decision-making. *Phil. Stud.* **34** 335–349.

CI; AL Describes the problem of "resource allocation," which includes aggregation of probabilities as a special case. The major criticisms of the DeGroot-Lehrer method of reaching a consensus are reviewed and the use of "proxy matrices" is suggested by which individuals can change their weights at each iterative stage. Sufficient conditions for consensus are described which parallel those of Chatterjee (1975) and Chatterjee and Seneta (1977).

WAGNER, C. G. (1980). The formal foundations of Lehrer's theory of consensus. In *Profile* (BOGDAN, R. J., ed). D. Reidel Publishing Co., Dordrecht, Holland.

CI; AL Summarizes and assesses the DeGroot-Lehrer method of reaching a consensus using linear opinion pools. Successive revisions of opinion are regarded as incorporating increasingly complex amounts of information about individual judgmental skills. It is shown how consensus may be obtained when the matrix of weights is ergodic but nonregular. The question of providing some axiomatic support for weighted averaging is also raised. The author's own answer to this question was given in the paper just below.

WAGNER, C. G. (1982). Allocation, Lehrer models, and the consensus of probabilities. *Theory and Decision* **14** 207–220.

AT; AL, M Axiomatic support for the linear opinion pool as a solution to the problem of allocation of resources. When interpreted in a measure-theoretic context, the so-called "strong label neutrality" axiom is equivalent to the SSFP of McConway (1981), condition (3.2).

WAGNER, C. G. (1984). Aggregating subjective probabilities: some limitative theorems. *Notre Dame J. Formal Log.* **25** 233–240.

AT; P Assuming that the consensual probability assigned to each event depends only on the probability assigned by individuals to that event (WSFP), and assuming that consensual distributions always preserve instances of independence common to all individual distributions (IPP), the author shows that the only aggregation methods available yield a consensus which is either dictatorial or imposed. Slightly weaker forms of the independence preservation requirement are explored. See also Genest and Wagner (1984).

WAGNER, C. G. (1985). On the formal properties of weighted averaging as a method of aggregation. *Synthese* **62** 97–108.

AT, CI; P Wagner's reply to his critics, namely Baird, Levi, Loewer and Laddaga, Nurmi, and Schmitt (all papers in *Synthese* **62** 1985). This article is a must if you read the latter papers, as Wagner makes a commendable effort to delineate the range of applicability of the so-called DeGroot-Lehrer model. Sections 2 and 4 are particularly well written and interesting.

WALLEY, P. (1982)*. The elicitation and aggregation of beliefs. Statistics Research Report, University of Warwick.

AT Develops a solution to the aggregation problem when opinions are expressed not as additive probability distributions, but rather as a range of such distributions. Different solutions emerge depending on which of 15 axioms is adopted. The basic underlying theory is presented by the author in an earlier unpublished work (1981). A fragmentary, critical review of other solutions to the pooling problem is given along with a discussion of how opinions represented by classes of distributions might possibly be elicited.

WEERAHANDI, S. and ZIDEK, J. V. (1978). Pooling prior distributions. Report No. 78-34, Institute of Applied Mathematics and Statistics, University of British Columbia.

AT, DM; D Contains an incorrect, but correctable derivation of the logarithmic opinion pool based on external Bayesianity, i.e., condition (3.12). Shows the importance of considering randomized decision rules in the context of point estimation and suggests a method for extending the domain of pooling operators accordingly.

WEERAHANDI, S. and ZIDEK, J. V. (1981)*. Multi-Bayesian statistical decision theory. *J. R. Statist. Soc. Ser. A* **144** 85–93.

Surveys the problem of pooling opinions within the framework of multiagent statistical decision theory. Shows how solution concepts proper to bargaining theory may be adapted profitably in this context. Presents valuable statistical applications including hypothesis testing.

WEST, M. (1984). Bayesian aggregation. *J. R. Statist. Soc. Ser. A* **147** 600–607.

AT, DM; P Suppose that a group of Bayesians must jointly choose an action and that the consequence of each possible action depends on the occurrence of a single event. Further assume that the decision facing the group will be assessed by reference to a group utility which depends only on the utilities of the individuals. By varying the gambles offered to the group, West shows that both the group utilities and beliefs must be derived via weighted geometric means without normalization constants. The author does not view this as an impossibility theorem, but rather suggests that "the group belief is *not* probability in any case other than that of complete agreement" (among the members of the group) (p. 604).

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

BU; D Often quoted review paper which is a clear forerunner of the Bayesian approach to aggregation of opinions. The so-called natural conjugate approach is suggested in which each expert's opinion is deemed as "sample evidence" which is incorporated into a decision maker's prior by successive applications of Bayes' theorem. Numerical examples are included.

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

BU; D A consensus model which allows for stochastic dependence between individuals is developed. When a certain location-invariance assumption holds, the decision maker's posterior is proportional to the product of his/her prior with a function of the individuals' "errors of estimation." Under normality, the model yields tractable formulas which are sensitive to the degree of dependence between the members of the group. The impact of dependence among the experts is further studied in Clemen and Winkler (1985), and Agnew (1985) derives formulas for posterior densities of the same normal model to accommodate situations in which the decision maker sought information on two or more random quantities.

WINKLER, R. L. and CUMMINGS, L. L. (1972). On the choice of a consensus distribution in Bayesian analysis. *Org. Behav. Hum. Perform.* **7** 63–76.

Example of an experimental study on the problem of pooling opinions. The results of an experiment are reported in which subjects were asked to subjectively generate a consensus distribution or to choose one from a fixed number of alternatives determined using different linear or logarithmic opinion pools. The weighted average methods are seen to have been most frequently used.

ZECKHAUSER, R. (1968). Group decision and allocation. Discussion paper No. 51, Harvard Institute of Economic Research, Cambridge, MA.

Cited by Raiffa (1968) and Bacharach (1975).

Comment

Glenn Shafer

This is a valuable review article. The annotated bibliography is a useful guide to the literature on the combination of beliefs, and the body of the article puts us in a position to assess the accomplishments and the direction of this literature as a whole, without undue emphasis on the ambitions and limitations of particular contributions.

My own view is that most of this literature is flawed by adherence to one of more of the following fallacies: (1) The Conditional Probability Fallacy: A Bayesian analysis of a problem always takes evidence into account by conditioning a probability distribution on that evidence. (2) The Fallacy of the Coherent Individual: Formation of opinion by a group is fundamentally different from formation of opinion by an individual. (3) The Fallacy of Normalcy: Use of the Bayesian paradigm is normative for an individual. My contribution to the discussion will concentrate on these fallacies.

1. THE CONDITIONAL PROBABILITY FALLACY

My interest was engaged when I read in the introduction that the problem of pooling knowledge would be covered in Section 4. But when I turned to Section 4, I found it entitled "The Supra Bayesian Approach." It appears to be taken for granted in the literature surveyed that the only Bayesian way to pool the knowledge represented by the opinions of several different people is to condition a "supra Bayesian's" probability distribution on these opinions. This is a special case of the conditional probability fallacy.

Like most fallacies, the conditional probability fallacy survives not because of persuasive arguments in its favor but because it so often goes unnoticed. It is a habit of thought resulting from our familiarity with the picture of statistical experimentation associated with parametric inference. The evidence that appears explicitly in this picture is a statistical observation, the result of a statistical experiment. The hypothesis (or parameter) space and the evidence (or observation) space are specified when the experiment is set up, before the observation is made. We are supposed to have a joint probability distribution over the Cartesian product of the hypothesis and evidence spaces, and once we get the evidence (i.e., make the observation) we are supposed to condition this probability distribution on it. Often we are told to do this in a way that uses Bayes's theorem; we must specify the joint distribution by specifying $P(H)$ and $P(E \mid H)$ for each element H of the hypothesis space and each element E of the evidence space, and then we must use Bayes's theorem to calculate $P(H \mid E)$.

In fact, a Bayesian analysis cannot take all the evidence into account by conditioning even in the case of a genuine statistical experiment. If the analysis is to be convincing, there must be some evidence that is used directly as evidence for the numbers $P(H)$ and $P(E \mid H)$. Since this evidence is not part of E, it is not taken into account by conditioning. This point is often overlooked because many advocates of the Bayesian paradigm do not want to acknowledge that the usefulness of Bayes's theorem in a particular problem depends on the existence and quality of the evidence for $P(H)$ and $P(E \mid H)$.

Our habituation to the picture of statistical experimentation also has a more subtle effect. Even after we admit that the evidence E on which we condition is not all our evidence, we tend to assume that E has been singled out from our other evidence by nature, not by ourselves. It is our "new evidence," the evidence that we just got from our experiment, and so it is easily distinguished from our "background information."

It is important to recognize that in the case of everyday evidence, at least, it is usually not true that E is singled out for us. We ourselves, when designing a Bayesian analysis, must decide which part of our evidence we will use to construct a probability distri-

Glenn Shafer is a Professor in the School of Business, University of Kansas, Lawrence, KS 66045.