# Comment

**Robin M. Hogarth**

As the review by Genest and Zidek shows, there is now a burgeoning literature on issues concerning the combining of probability distributions. Moreover, readers will be grateful to them for providing such a comprehensive overview and guide to this literature. In the following comments, I wish to emphasize three points. These relate to (1) whether it is reasonable to expect group opinion to act like the Bayesian model; (2) the importance of determining the commonalities and differences between distributions that are being combined; and (3) how considering the specific decision context can, from a practical viewpoint, often simplify the problem of combining distributions.

(1) On considering whether one should expect group opinion to conform to the Bayesian model, it is first important to consider what one means by group opinion and the purposes for which this has been elicited. I distinguish three forms of "group opinion." The first occurs when an *individual* decision maker assesses a distribution using, as inputs, distributions from other sources, e.g., experts, forecasts from models, and so on. In this case, the combined distribution becomes the opinion of the individual. A second case involves a group of people whose members wish to combine their probability distributions in order to make a particular decision. Examples of this kind of situation could involve business partners or even a married couple or family. A third case concerns a group of people who wish to express an opinion about some issue in the form of a probability distribution in order to communicate this as information for other people. An example would be professional groups (e.g., physicians) providing information to the general public (e.g., about health risks).

From a technical viewpoint, one could treat all these situations identically, e.g., by adopting a "supra Bayesian" model. However, in my view the three situations are conceptually quite different. The first and third situations are extreme cases. In the first, there seems little doubt that one would want the Bayesian model to apply. For example, one would expect the individual to update his or her "consensus" distribution in the light of new evidence according to Bayes' theorem. In the third case, however, all one really wants is a consensus of opinion at a particular point in time. It would seem strange to require the professional association to follow *all* the dictates of the Bayesian model with respect to that opinion. The second case is more complicated. On the one hand, it is possible to treat this like the third case and simply work on assessing a distribution for a particular problem. On the other hand, many people (myself included) feel that one should be able to apply the Bayesian theory to multiple party decision making. What distinguishes these situations? In my view, a critical variable is the extent to which the multiple parties resemble an individual decision maker. However, since resemblance has many dimensions, let me suggest two criteria: (a) whether group members are involved in a stable long-term relationship; and (b) whether they have similar or even identical interests. For example, members of professional associations may have some common interests but their relationships are inherently unstable (e.g., membership is changing constantly). Married couples, or even close business partners, on the other hand, have relationships that can be more easily assimilated to the notion of a single Bayesian decision maker. Without providing an answer, I am suggesting that we think more carefully about what kinds of groups could or should be thought of as Bayesian decision makers. It is not clear to me that an all purpose solution exists for this problem, nor that it would be desirable.

(2) In a recent paper, Clemen and Winkler (1985) show the deleterious effects on the combination process of assuming that distributions are independent when in fact they are not. Conceptually, it is possible to distinguish two types of dependence in aggregation. One is the common notion that two or more distributions may be based on the same data and thus double counting occurs if the data considered in each distribution are treated as independent in the aggregation process. The second is what psychologists term "common method variance." This refers to the possibility that two or more distributions being aggregated result from analyses based on a common method. To illustrate, consider the following thought experiment. Would you have more or less faith in a medical diagnosis if there was agreement between two experts using different methods of diagnosis as opposed to agreement between two experts using the same method?

From a practical viewpoint, what can be done to handle problems of redundancy among distributions? Much prescriptive and descriptive work in decision making attests to the value of decomposing decision

*Robin M. Hogarth is Professor of Behavioral Science and Director of the Center for Decision Research at the Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, IL 60637.*

problems and judgments that are used as inputs to decisions. In this spirit, I believe it is appropriate to identify the components on which the probability distributions of individuals are based, prior to considering questions of aggregation. I suggest that four sources of difference are important to the aggregation process. These center on: (a) the model underlying a person's distribution; (b) differential weights given to different variables within similar models; (c) different data; and (d) differential use of expressions of confidence.

(a) It is important to recognize that, even in the most subjective of situations, judgments of probability are based on some underlying model. This is not to state that such models are necessarily well specified. For example, whereas in some cases one can model a person's judgments by, say, a normal distribution with known parameters, in other cases the sophistication of the model may be no more than a rough statement that two variables have some qualitative relationship. However, if judgment has any validity, it must be generated by a systematic process involving some understanding (i.e., model) of the situation. In my view it is critical to the aggregation process to understand the models on which opinions are based.

(b) In applied settings, people may have quite similar causal models on which they base their distributions. For example, in economic forecasting these models could be based in economic theory. However, it could well be the case that different experts place different weights on the various variables and that this contributes to different predictions.

(c) On what data are predictions based? As noted above, common data are the most obvious source of redundancy between distributions. The extent of overlap, therefore, needs to be determined.

(d) When individual distributions are based on formal statistical models (e.g., the normal distribution), the variance of the distribution provides a precise means of assessing the uncertainty implicit in predictions. However, when distributions are based on subjective judgments, the uncertainty is that expressed by the assessor. Moreover, individuals can differ considerably in their ability to state probabilistic opinions that accurately reflect subsequent events. These individual differences need to be examined.

(3) In their review, Genest and Zidek chose not to consider issues concerning the combining of distributions within the context of decision problems. Whereas this was probably appropriate for their paper, I feel that from a practical viewpoint it is important to try to exploit characteristics of particular problems. Specifically, certain strategies that are commonly used in univariate assessment can also be applied when faced with several distributions. For example, in many problems optimal actions do not depend greatly on the precise distribution employed. Thus, prior to attempting the difficult task of combining distributions, much can be gained by sensitivity analyses designed to test the robustness of relative preferences for different actions implied by various hypothetical distributions. As in univariate analysis, it is smart only to spend time on cases where differences are important. Second, there are also many problems for which one does not need to have a complete distribution but where the optimal action depends solely on a certain fractile of the distribution. Once again, one need only go through the combination process to the extent that this is necessary.

A further class of decisions concerns situations where, although it seems that a single answer or distribution is required, the different parties in fact possess expertise about different aspects of the problem. In this case, the trick is (1) to structure the problem so that the nature of different expertise in the group is apparent and agreed upon by the parties, (2) to determine a model that decomposes the problem into appropriate components, (3) to elicit distributions from each of the members in their areas of distinctive competence, and (4) then aggregate within the structure of a general model on which all can agree. For an instructive example of this approach in the area of risk analysis, see Hammond et al. (1984). I am not, of course, suggesting that problems of combining distributions can always be avoided because of the inherent structure of decision problems. What I am suggesting is that in many cases, rather than being overconcerned about the intricacies and difficulties of pooling formulas and the like, one may still be able to restructure the group assessment task in a manner that uses the information available in the group.

To summarize, the review by Genest and Zidek is a welcome addition to the statistical literature in that it encapsulates many of the issues inherent in combining probability distributions. In this comment, I have tried to emphasize three points by way of complementing their paper. These have focused on what kinds of groups can be assimilated to single Bayesians, sources of redundancy and differences between distributions being combined, and seeking to avoid the problems of combining distributions by exploiting the characteristics of decision problems. Finally, and by way of a footnote, I cannot help but note that Genest and Zidek use the expression "Bayesianity" several times in their paper. I can understand, and even empathize, with the notion that in the recent past Bayesian statisticians have had to adopt an almost evangelical attitude in having their views accepted. However, perhaps even the most fervent believers might agree that this expression is going a bit far!

**ACKNOWLEDGMENT**

This research was supported by Contract N00014-84-C-0018 from the Office of Naval Research.

## ADDITIONAL REFERENCE

HAMMOND, K. R., ANDERSON, B. F., SUTHERLAND, J. and MARVIN, B. (1984). Improving scientists' judgments of risk. *Risk Anal.* **4** 69–78.

# Rejoinder

**Christian Genest and James V. Zidek**

We are grateful to the editor, Professor Morris H. DeGroot, for taking an active interest in our manuscript and for organizing the discussion. We are sure that workers in this area will be equally grateful to all the discussants for sharing their thoughts and shedding some additional light on the murky, multifaceted problem of aggregating expert judgments. Their comments add welcome dimensions to our survey and demonstrate that there is yet no consensus about how to reach a consensus.

The most serious difference of opinion occurs between Professor Shafer, who argues against the Bayesian paradigm for groups of all sizes including $n = 1$, and the remaining discussants who support and focus on the Bayesian approach with qualification. Among the Bayesians, Drs. Winkler, Morris, and Hogarth regard an elicited subjective probability as a measurement or as information which can be aggregated through a suitable supra Bayesian approach. In fact, all three appear to favor this approach even in situations where no natural choice exists for the supra Bayesian. Professor French, on the other hand, sides with de Finetti's completely personalistic view of probability. The latter suggests that interpersonal comparability of probabilities may not be possible, in which case it is not clear whether the supra Bayesian aggregation of elicited opinions would ever be meaningful to those who support this viewpoint.

In our opinion, Bayesian methods provide the sole normatively acceptable answer to the aggregation problem when the group reports to a third party. The Bayesian solution may not be so useful, however, in situations where the group as a whole is seeking a consensus or wishes to summarize its opinions for the benefit of others "at the end of the day." We regard this problem as one of fundamental importance in this area. Solving it would bring us one step closer to finding a middle ground better suited to modern science, between the classical notion of objectivity through replicability and the frequency theory of statistics on the one hand, and the entirely subjective theory of the Bayesians on the other. This issue is recognized by French, but he fails to see the relevance of axiomatic approaches to its resolution. Interestingly, the search for consensus, in its epistemological sense of unanimous agreement, has aroused a great deal of interest in philosophy, but it has been largely ignored by statisticians, who tend to take an "operations research" view of the whole subject matter. In the discussion, only Winkler is willing to admit the need to tentatively consider axiom-based formulas, and he does so only because "the modeling approach may be difficult to apply in actual situations."

The challenge of the theory as it stands is that it is not always clear when a given situation calls for compromise, summarization, or consensusalization. In this regard, Hogarth's recommendation that combination of opinion could be guided by the decision context would seem to be a useful observation. Confusion in the objectives of the theory derives in part from the context-dependent meaning of such words as "consensus" or "opinion pool." We are only beginning to recognize that there is more than one consensus problem. It is not surprising, therefore, that there should be "no single combining procedure for all seasons," as Winkler put it. What is surprising, however, is to see French and Morris seize on the supra Bayesian paradigm as a way of specifying the objectives of any problem. Morris goes even further in suggesting that we adopt this point of view as a way of evaluating the relative merits of prospective pooling methods or formulas. Although it would be legitimate for an individual to evaluate a group procedure on these grounds, this consideration would be irrelevant to the value of the procedure for the group as a whole.

Let us now turn to some of the more technical issues which were raised in the discussion. We begin with Professor Shafer's criticism, which focuses on the deficiencies of the Bayesian (and hence the supra Bayesian) approach. In our paper, we acknowledge that the supra Bayesian approach will inherit all the criticisms of the Bayesian philosophy. The value of Bayesian versus nonBayesian statistics has been and is still the object of a vigorous debate in the statistical