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 community, and it does not seem appropriate to us to address this problem here. However, we do take issue with several of Professor Shafer's other points.

First, we would argue that a joint distribution for E and H, in Shafer's notation, is unnecessary *a posteriori*. We believe that only $P(E \mid H)$ and $P(H)$ are required and that they may be elicited for all H and the given E. Thus, at least in principle, additional data (or some subset of E) would not be needed to develop a model to find $P(H \mid E)$. This leads us to reject the infinite conditioning regression described by Shafer, in what seems like an empirical Bayesian approach. We do recognize, however, that at this stage of development of the Bayesian theory, eliciting $P(E \mid H)$ in the supra Bayesian context still appears unrealistically difficult in practice, except in some extremely simple cases.

We disagree with Shafer about what it means for a theory to be "normative." The Bayesian theory, as it were, does not "force" a person to invoke Bayes' rule. Rather, it asserts that an individual will update his/her opinion in accordance with Bayes' rule *if* that individual's likelihood patterns conform to the axioms underlying subjective probability theory. Consequently, if a subject aspires to have such ideal belief patterns, he/she ought to use Bayes' theorem to update his/her opinion. The supra Bayesian theory which we surveyed in our paper is "normative" in the sense that it describes what is to be expected of those people who aspire to this ideal, even though their actual belief patterns may fall short of it. Descriptive accounts of these patterns are relevant only to the extent that they might guide individuals toward their ideals.

Finally, we reject Shafer's conclusion that groups do not differ from individuals in their treatment of probabilities, even though the identification analogy which he proposes is tempting. One important distinction is that individuals are capable of within-person comparisons of degrees of belief, whereas interpersonal comparisons may not be possible. French also comments on this issue.

Professor Winkler is right when he argues that axioms like independence preservation are stronger than their naive interpretation might suggest. But they are considerably weaker than the assumption of the pooling operator they imply! And where else would these axiomatic analyses begin? To strip off successive layers of glossy assumptions to discover the most primitive foundations of aggregation is the common goal of these axiomatic enquiries. While none of the researchers involved in this area would claim that a universal prescription for pooling is in hand, their work has shown the kinds of assumptions that underlie particular pooling recipes. This enables us to anticipate the manner in which paradoxes might arise when we reject a specific method of combination. If a linear opinion pool were used instead of a logarithmic formula, for example, this would lead to paradoxical conditional results and a potential need to rationalize the strange behavior that arithmetic averaging entails.

We do not support Professor Winkler's judgment that the form in which individual statements of uncertainty are cast is a red herring. Unless the issue about scaling of quantitative expressions of opinions is ignored, odds and probabilities are not equivalent. This issue is of great fundamental importance given the existing doubt about the legitimacy of interpersonal comparison of probabilities. Difficulties arise, *inter alia*, from the consideration of "irrelevant alternatives" and the fact that all probabilities are conditional upon individual states of information which vary (implicitly or explicitly) from person to person. These concerns are echoed and amplified in French's remarks about personal probabilities.

Both Morris and Hogarth emphasize the connection between the aggregation and the decision problems. We make the point in our introduction that the problem of consensus would often (but not always) be a group decision problem in practice. Some action would ultimately be recommended by the group, and this would depend both on personal preferences and degrees of belief of its members. The group's action might simply be the declaration of an aggregated opinion in the form of a probability distribution. In fact, one solution to this decision problem leads to the linear opinion pool under the assumption that utilities are intercomparable. Unfortunately, this is still largely unexplored territory.