When is it appropriate to combine expert judgments?

David W. Keith
Department of Chemistry, Harvard University,
12 Oxford St., Cambridge, MA.
Internet: keith@huarp.harvard.edu

Published in *Climatic Change*, **33**, 139-144 (1996).

A recent study by Titus and Narayanan (in this issue) is a remarkably ambitious attempt to perform an end-to-end evaluation of the probability of sea level rise. At various steps in their analysis the authors incorporate the elicited judgment of experts including, in some cases, both parameter and model uncertainty. They then average the expert judgments to produce a single-valued probability distribution for future sea level rise. The authors acknowledge the methodological difficulties associated with combining expert judgment, but argue implicitly that the end justifies the means.

Under what circumstances is it appropriate to combine expert judgments? In this editorial an answer is suggested by focusing on the question, "Who is the audience for the analysis and what do they need?", with the conclusion that it is rarely appropriate to combine divergent expert judgments.

In order to sharpen the question, consider following the exaggerated recipe for policy analysis. Policy analyses of climate change should aim to be comprehensive and analytic: comprehensive in the sense that they encompass as much as possible of the causal chain that connects human actions through their influence on global climate to human impacts, and analytic in the sense that they are probabilistic numerical estimates. A key barrier to producing comprehensive analyses is scientific uncertainty that prevents the use of physically based models to analyze many of the important links in the causal chain, e.g., climate sensitivity, the link between climate forcing and response. The combination of carefully elicited expert judgment, while troubled by some methodological difficulties, is a uniquely powerful method for producing in analytic form a credible "best current estimate" of scientific knowledge. Probabilistic estimates from many disciplines can be combined to produce a comprehensive analysis with a systematic representation of uncertainty. Such analyses are an ideal rational basis for public policy choices.

In my view this recipe for policy analysis is not only impractical, due to methodological difficulties, but is also, in many cases, undesirable even if those difficulties could be overcome.

The core methodological problem of combining expert opinion is based on a simple, unavoidable, truth: the fraction of experts who hold a given view is not proportional to the probability of that view being correct. Moreover, even if this were the case, the results of combined expert opinion might be meaningless if the expert's models were incommensurate. I.e., one may safely average model parameters, but not models.

We may define the task of combining expert opinion as "difficult" if different, plausibly correct, methods for the combining produce outcomes that differ enough to significantly affect the conclusions of the analysis. When combining expert opinion is difficult we have three choices: (i) resolve scientific disagreement by use of consensus building methods such as the Delphi method (Dalkey, 1967) or other procedural devices such as science courts (Kantrowitz, 1967), (ii) combine expert opinion using the best available method (for a survey of methods see Genest and Zidek, 1986), or (iii) punt.

If the result of consensus building is simply to arrive at a statement that all can accept without the experts changing their individually held views, then the results may have no more merit than if the views were weighted (perhaps equally) by an independent analyst. In this case the consensus building exercise has simply produced a combined opinion weighted by social persuasiveness. Effectiveness at influencing group dynamics has no known correlation with

accuracy of scientific viewpoint. Only if consensus building produces a real movement in the experts' personal views through active scientific discussion will the results be a clear improvement over a combination of expert judgment by an analyst. But this is simply a recipe for real scientific progress, and the goal of consensus building is to make better estimates in the face of unresolved scientific uncertainty.

Punting need not mean dropping the analysis altogether, but rather, may entail the adoption of alternative modes of analysis such as the design of adaptive strategies which are a robust response to the scientific uncertainty (Lempert et al, this issue), or the use of scenario analyses to bound the problem. Posing the question, "Who is the audience for our analysis and what do they need?", can resolve methodological questions about combining expert opinion by showing that in most cases the best choice is to punt.

Consider four common answers to the "audience" question. (i) A unitary rational actor with the power to set global policy, and the ability to judge aggregate global utility, who we may call the "benevolent world dictator"; (ii) research managers; (iii) decision makers with responsibility for long-range plans who have limited, but non-zero, influence over global decisions, e.g., national leaders; or, (iv) decision makers who must make long-range plans in a limited domain but who have no power to influence global decisions, e.g., coastal zone managers. We now consider each audience individually.

The world dictator needs policy analysis to answer "what if" questions that span the causal chain from a given policy choice such as an energy policy or an emissions trajectory to human impacts perhaps expressed through aggregated utility functions. Despite the absence of a world dictator, much policy analysis of climate change is implicitly framed in terms appropriate for this audience. One may argue (Morgan and Dowlatabadi, 1996) that the heterogeneity both of climate impacts, and of actors whose choices affect climate, is a fundamental part of the climate problem. Comprehensive analyses necessarily tend to ignore this heterogeneity and so may seriously misconstrue the problem. There are some legitimate reasons to frame analyses in these terms, for example, when it suits their interests in international negotiations, national leaders may appeal to arguments that are implicitly based on a "benevolent world dictator" viewpoint. However, it seems fair to argue that policy analysts too often adopt this framing and give insufficient consideration to its implicit limitations.

Research managers do not often require comprehensive analysis. They do not need to know what "the" answer is, but rather, what the leading sources of uncertainty are, what is the cost and time-scale for reducing uncertainty, and what is the potential for surprise. In general, they will not require the combining of expert judgment in difficult cases because the presence of diversity of opinion is itself precisely the information that they need.

National political leaders face a collective action problem in addition to the "what if" problem faced by the illusory world dictator. For example, they must know how their choice of a carbon dioxide emissions policy will influence globally aggregated emissions (a collective choice problem), in addition to what climate-mediated impact on national utility will result from a given emissions scenario. This added layer of uncertainty makes end-to-end quantitative policy analysis implausible (Ayres, 1984). Thus, for this audience comprehensive and analytic policy analyses are not directly useable, and so the best choice is to punt and to perform less comprehensive policy analyses that place less emphasis on the combined judgment of experts.

In the final category we find many decision makers who want analyses of the implications of climate change, e.g., water resource managers, urban planners, or state governments. However, for a variety of reason few of them have a legitimate use for comprehensive and analytic policy analyses. Typical reasons include: lack of a real need to plan on the time-scale of climate change; the problems in their domains are too complex for quantitative analysis; or, the aggregate impact of climate change is small enough that only a bounding analysis is appropriate.

There may be, however, a subset of decision makers who have a sensible use for policy analysis along the lines described in the opening "recipe." The problems that such decision mak-

ers face include the following characteristics: (i) intrinsic time-scales (e.g., capital turn-over rates) comparable to climate time-scale, (ii) impact of climate change significant compared to other factors, (iii) decisions made in an environment where aggregation is natural so that the decision problem looks simple and may be analyzed quantitatively. For audiences that meet these criteria it may be appropriate to combine expert opinion using the best available methods, and to perform a comprehensive analysis that begins with an estimate of the future course of anthropogenic climate forcing determined by socioeconomic factors. This is appropriate because no other method of analysis produces a better single-valued result, and a because a decision is required immediately.

A simple, and seemingly objective, method for combining expert judgments is to weight equally all experts in a statistically representative sample. In real-world cases this will generally be impossible. This is because it is prohibitively difficult to sample a sufficiently large number of experts while simultaneously ensuring that each elicitation of expert judgment is performed in sufficient depth to achieve an accurate representation of the experts' views on a subject with a real-world degree of complexity. For example, in a recent elicitation of expert judgments about climate change in which I was involved (Morgan and Keith, 1995), we were only able to sample 15 experts with an effort by our group that amounted to several person-years, while maintaining the minimum adequate level of detail in each elicitation. In my view, the minimum statistically representative sample size for this population on the question of climate sensitivity would be several times larger than our sample due to the difficulty of defining who is a expert.

In the absence of a statistically representative sample, the analyst must weigh expert opinion unequally, either by use of explicitly assigned weights, or implicitly by choice of experts. Thus, the results of comprehensive policy analysis based on combined expert opinion will be subjective, in the sense that different analysts will produce different answers. This is a consequence of the unavoidable absence of a objective basis for combining expert opinion.

Summary

Rather than seeking improved methodologies, difficulty in combining expert opinion should serve as a warning flag that causes us to seek alternative modes of policy analysis. These alternatives are usually more appropriate for the real audience for our analyses.

Policy analysis of climate change is too often framed in terms that amount to preparing the tools with which a benevolent world dictator could do cost-benefit analysis. This tends to over emphasize end-to-end analysis that must rely on the combined opinions of experts. This framing is unrealistic and encourages omission of important aspects of the climate problem such as its heterogeneity. Rejecting this framing in favor of alternate, less all encompassing, forms of policy analysis permits more robust results, and reduces the emphasis on combining expert opinion.

While the opinions expressed here are my own, I thank Hadi Dowlatabadi, M. Granger Morgan, Ted Parson, and James Risbey for their perceptive comments.

References

Ayres, R.U. (1984) "Limits and Possibilities of Large-Scale Long-Range Societal Models", *Technological Forecasting and Social Change*, v. 25, 297-308.

Dalkey, N.C. (1967) "Delphi" Rand Corporation paper P-3704, Rand Co., Santa Monica, CA.

Kantrowitz A, (1967) "Proposal for an Institution for Scientific Judgment", *Science* v. 156, 763-764.

Lempert R.J., Schlesinger, M.E., and Bankes, S.C., 1996. "When we don't know the costs or the benefits: adaptive strategies for abating climate change," *Climatic Change*, (this issue).

Morgan, M.G. and H. Dowlatabadi (1996) "Learning from Integrated Assessment of Climate Change", *Climatic Change*, in press.

Morgan, M.G. and D.W Keith (1995) "Subjective Judgments by Climate Experts", *Environmental Science and Technology*, v. 29, 468-476 A.

Titus, J.G. and V. Narayanan (1996) "The Probability of Sea Level Rise", *Climatic Change*, (this issue).