CLIMATE CHANGE

Subjective Judgments by Climate Experts

M. GRANGER MORGAN
Carnegie Mellon University
Pittsburgh, PA 15213

DAVID W. KEITH
Harvard University
Cambridge, MA 02138

Structured interviews using "expert elicitation" methods drawn from decision analysis were conducted with 16 leading U.S. climate scientists. We obtained quantitative, probabilistic judgments about a number of key climate variables and about the nature of the climate system. We also obtained judgments about the relative contributions of various factors to the uncertainty in climate sensitivity. We found strong support for the importance of convection/ water vapor feedback and of cloud optical properties. A variety of questions were posed to elicit judgments about future research needs and the possible sources and magnitude of future surprises. The results reveal a rich diversity of expert opinion and, aside from climate sensitivity, a greater degree of disagreement than is often conveyed in scientific consensus documents. Research can make valuable contributions, but we interpret our results to mean that overall uncertainty about the geophysics of climate change is not likely to be reduced dramatically in the next few decades.

When scientific uncertainty limits analytic modeling, but decision makers cannot wait for better science, expert judgment can be used in the interim to inform policy analysis and choice. Approaches such as the "Delphi method" (*1, 2*) have been developed to obtain consensus summaries. Without using formal methods, scientific panels, such as those of the Intergovernmental Panel on Climate Change (IPCC) (3) or the National Research Council (NRC) (4) often strive to produce consensus summaries.

However, when uncertainty is high because of fundamentally different views about underlying physical processes, a consensus summary may not best serve policy analysis needs. An alternative approach, widely used in applied Bayesian decision analysis (*5–7*), formalizes and quantifies the judgment of individual experts through expert elicitation (8-11). Subsequent analysis of results allows conclusions about the importance of the range of expert opinions to the overall policy debate. Apparent deep disagreements can make little difference to the policy conclusions or be critically important (*11*).

We conducted detailed expert elicitations with 16 leading climate scientists as part of a Carnegie Mellon program of integrated assessment of climate change (12). Several limited elicitations on climate-related topics have been reported in the literature [*13* (for a critique see Stewart, T. R.; Glantz, M. H. *Climate Change* 1985, 7, 159–83), 14,151, but ours were more technically detailed. Because we were concerned about both model uncertainty and parameter uncertainty, we also developed several new question designs and response modes not used in previous elicitations.

The results reveal a rich diversity of expert opinion. Aside from estimates of climate sensitivity, defined here as the global average surface warming caused by a doubling of carbon dioxide concentrations, they indicate a greater disagreement than we believe is usually conveyed in scientific consensus documents. We are aware of one instance in which such a panel attempted to use subjective probability distributions to characterize uncertainties (*1*6, *1* 7). Although such techniques require time and care, we believe that more general use of these methods could improve communication significantly.

We find there is almost no agreement about the effect of climate change on policy-relevant factors such as changes in precipitation over land and various forms of interannual variability. However, experts agree that uncertainties in policy-relevant measures, such as climate sensitivity, could be reduced by about 20% over the next decade or two if research substantially improves understanding of convection/water vapor feedback or cloud optical properties. Research on oceanic convection and CO, exchange with terrestrial biota and the oceans could yield 10% reductions in uncertainty. Although encouraging, these findings suggest that dramatic reductions in uncertainty are unlikely for the next few decades.

Interviewing the experts

Judgments about uncertainty can be subject to biases, resulting from the cognitive heuristics people use when reasoning under uncertainty (*18*). In designing elicitation protocols, such effects must be minimized (8). We developed the interview protocol over a nine-month period. We first listed technical and policy questions that we believed were central to our integrated assessment. To assure systematic coverage and avoid the effect of cognitive heuristics called the availability bias (*18*), we constructed a number of influence diagrams. By iterating between policy concerns, the scientific literature, and trial questions administered to our colleagues, we made trade-offs to reduce the set of questions to something that could be completed in a day. We ran full-scale rehearsals with Mark Handel of MIT (now at NRC) and Filippo Giorgi at NCAR. Ann Henderson-Sellers of Macquarie University reviewed the draft materials. Originally, we planned to conduct two rounds of interviews: the first to develop questions that all experts found appropriate, the second to administer them. Resource constraints prevented this. We did conduct a simple mail survey of the experts that revealed several problems before we started the interviews. Modest mid-course refinements were made in the protocol, one after the third and the other after the sixth interview. Simple mail-back questions were used to get uniform coverage.

The protocol was produced as a workbook (19). The interviews began with a general introduction. In Part 2 we spent between 0.5 and 2.5 h in a free-wheeling technical discussion that began with a critique of a briefing paper we had previously distributed in an effort to clarify key concepts and definitions (*20*). We asked experts to provide us with a detailed commentary on the state of the field, particularly discussing issues they believed were getting too much or too little attention. To assist them, we provided an influence diagram of our construction as well as one from the literature (*21, 22*). No one chose to. make significant use of these diagrams.

Part 3 involved a series of questions about uncertain coefficient values and typically lasted between 1 and 1.5 h. Responses were elicited in the form of subjective probability distributions, using standard methods from the literature (8). A number of broader questions also were posed about time dynamics, interannual variability, and the nature and structure of the climate system.

Part 4 involved a card-sorting task to systematically identify and rank order the factors that contribute to uncertainty about climate sensitivity. This lasted about an hour.

Part 5 involved a two-stage process of building a zero-based U.S. climate research budget at a level of $1 billion per year for 15 years. This task, which involved the use of checkers on specially constructed game boards, typically required about an hour.

In the final part, we asked the experts to systematically identify and discuss surprises that might be uncovered by this research program. Then they quantitatively judged how these surprises might influence the mean and standard deviation of their future judgment of the uncertainty of the value of climate sensitivity. All were cooperative, completing the entire process. However, we frequently sensed, as the final question came to a close, that we had pushed to the limit of endurance. Several experts said they found the process exhausting but intellectually stimulating. Most spoke enthusiastically about the experience in our wrap-up conversation.

Because each set of expert judgments is offered as a single considered view, it generally is not appropriate to average across the results obtained from different experts (for an elaboration see section 7.7 in [*8*]). For this reason, we made no attempt to select a group that was "statistically representative" of the field. Rather, we sought to include at least one representative from most of the mainstream schools of thought. We gave serious consideration to the quality of scientific credentials; all experts could be characterized as "serious scientific players of the first rank." Fiscal and logistical constraints limited the number and precluded experts outside of the United States. Three experts declined to participate. The experts are listed in the sidebar. Numbers used to reference experts were assigned randomly.

All interviews were recorded and transcribed. Transcripts are between 20,000 and 56,000 words. In addition, the two authors, who jointly conducted all the interviews, took extensive notes. Finally, experts recorded specific responses in written form in the interview workbook Detailed results can be made available to researchers who wish to make use of specific findings. We report selected highlights. Further details will be reported in additional papers.

Parameter values and dynamics

Equilibrium temperature response to doubling the concentration of CO, (2 x [CO,]) has become a benchmark for assessing climate sensitivity. Although the applicability of such scenarios to practical estimates of climate impact is indirect and controversial, we began Part 3 with questions about a 2 x [CO,] scenario because all the experts would have considered this case. We asked the experts to provide a probability distribution for "the equilibrium change. . .in global average surface temperature" given "a doubling of CO, from pre-industrial levels. . ." assuming "anthropogenic aerosols and other [greenhouse gases] remain at current levels." We defined "equilibrium change" as elapsed time of about 200 years. IPCC's best estimate in 1990 was 2.5 K (3) and in 1992 it was "unlikely to lie outside the range 1.5 to 4.5 °C" (22).

The experts' probability distributions are summarized as box plots in Figure 1. We asked them to consider all possible outcomes. Experts 2 and 4 produced a separate distribution for a "state change" or a "surprise," respectively. Except for these two distributions and that of Expert 5, all the distributions are remarkably similar. Note that 5 of the 16 experts assess a small probability that a 2 x [CO,] scenario could lead to net cooling. These negative values typically were explained by arguments about the

FIGURE 1

Surface temperature change

Box plots of elicited probability distributions of climate sensitivity, the change in globally averaged surface temperature for a 2 x $[CO_2]$ forcing. Horizontal line denotes range from minimum to maximum assessed possible values. Vertical tick marks indicate locations of lower 5 and upper 95 percentiles. Box indicates interval spanned by 50% confidence interval. Solid dot is the mean and open dot is the median. The two columns of numbers on right side of the figure report values of mean and standard deviation of the distributions.

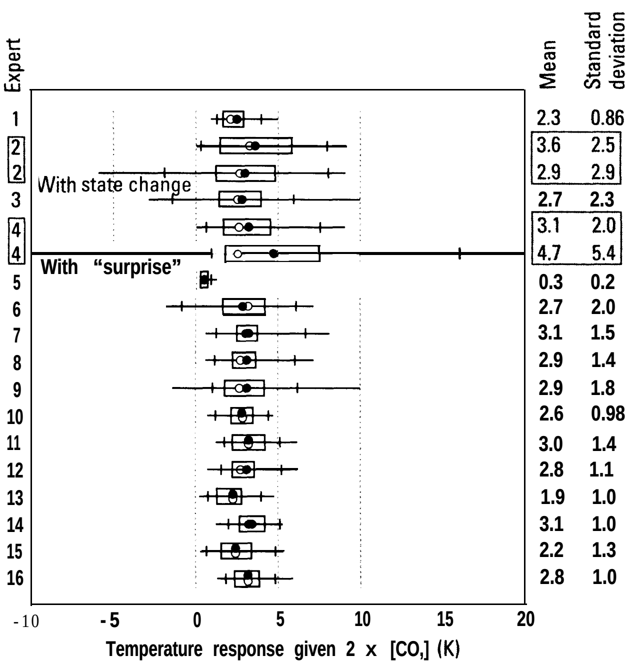
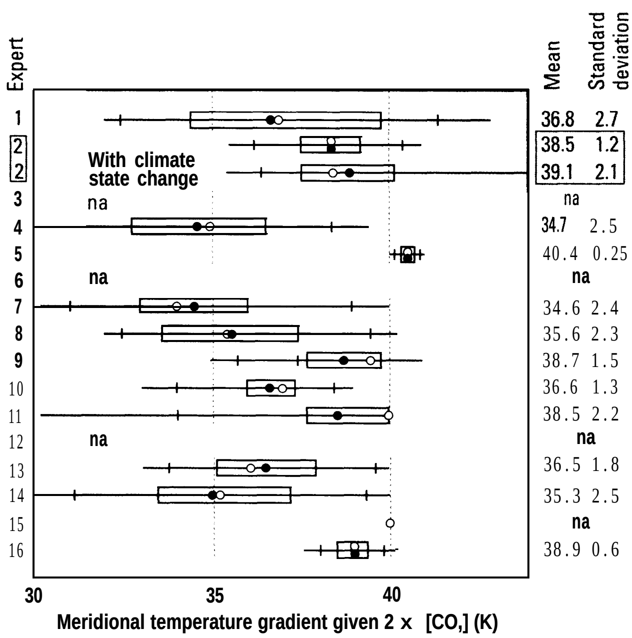


FIGURE 2

Meridional temperature gradient

Box plots of elicited expert subjective probability distributions for meridional temperature gradient (Equator to region above 70 °N) for a 2 x $[CO_2]$ forcing. The current value is 40 K. Display conventions are as in Figure 1.



disruption of North Atlantic deep-water formation.

We next asked experts for the "equilibrium value of the change in meridional temperature gradient" in the Northern Hemisphere. Responses to this question (Figure 2) show much greater variability than the previous question. Most experts expect polar regions to warm more than low-latitude regions and judged that the gradient would become more shallow. However, note that five experts assess a >5% chance that the gradient could steepen. Asked to explain the 10% chance he'd placed on the gradient exceeding 40 K, Expert 1 said he was thinking about a feedback that increased polar snow and ice cover.

We asked experts to estimate the impact of the same scenario on the "annual amount of precipitation over land" at several latitudes. The heavy dots and solid bars in Figure 3 summarize results (increases to the right, decreases to the left) from the 11 experts who answered. There was no clear consensus, even about the qualitative form of the latitudinal dependency. For example, four experts anticipated a sizable increase in equatorial precipitation relative to adjacent latitudes, whereas four clearly did not. We asked experts for a judgment about change in the interannual variability of precipitation. The four experts who responded (open dots and light solid bars in Figure 3) predicted an increase in interannual variability at intermediate latitudes. There was no consistency in the results at high or low latitudes.

In our pre-interview surveys, several experts told us to ask about variability. In addition to the precipitation question, we asked if the strength of an oscillatory behavior in the Pacific Ocean, known as El Niño-Southern Oscillation (ENSO), would be a good indicator of regional-scale interannual variability. All 16 answered yes, though several said it was a "useful but not sufficient indicator." We asked how the standard deviation of the ENSO sea surface temperature anomaly in the eastern equatorial Pacific might change from its current value of 0.6 K under 2 x $[CO_2]$. Ten experts could not answer this question and six added that they did not believe anyone else had the knowledge to answer. Those who did answer provided the following estimates of the ratio change in standard deviation: Expert 1, 1.1; Expert 5, ~ 1; Expert 7, 1.5; Expert 13, 1.1 to 1.5; Expert 14, >1; and Expert 16, <1 "with low confidence." In our wrap-up discussions, several experts returned to stress the importance of variability. For example, Expert 14 noted that we had not asked about possible changes in the frequency of long-lasting extreme events such as droughts. We queried, "Suppose we had asked you questions about extreme events and droughts. . ." and were told, "I would have given you an answer equally evasive as my answer about ENSO."

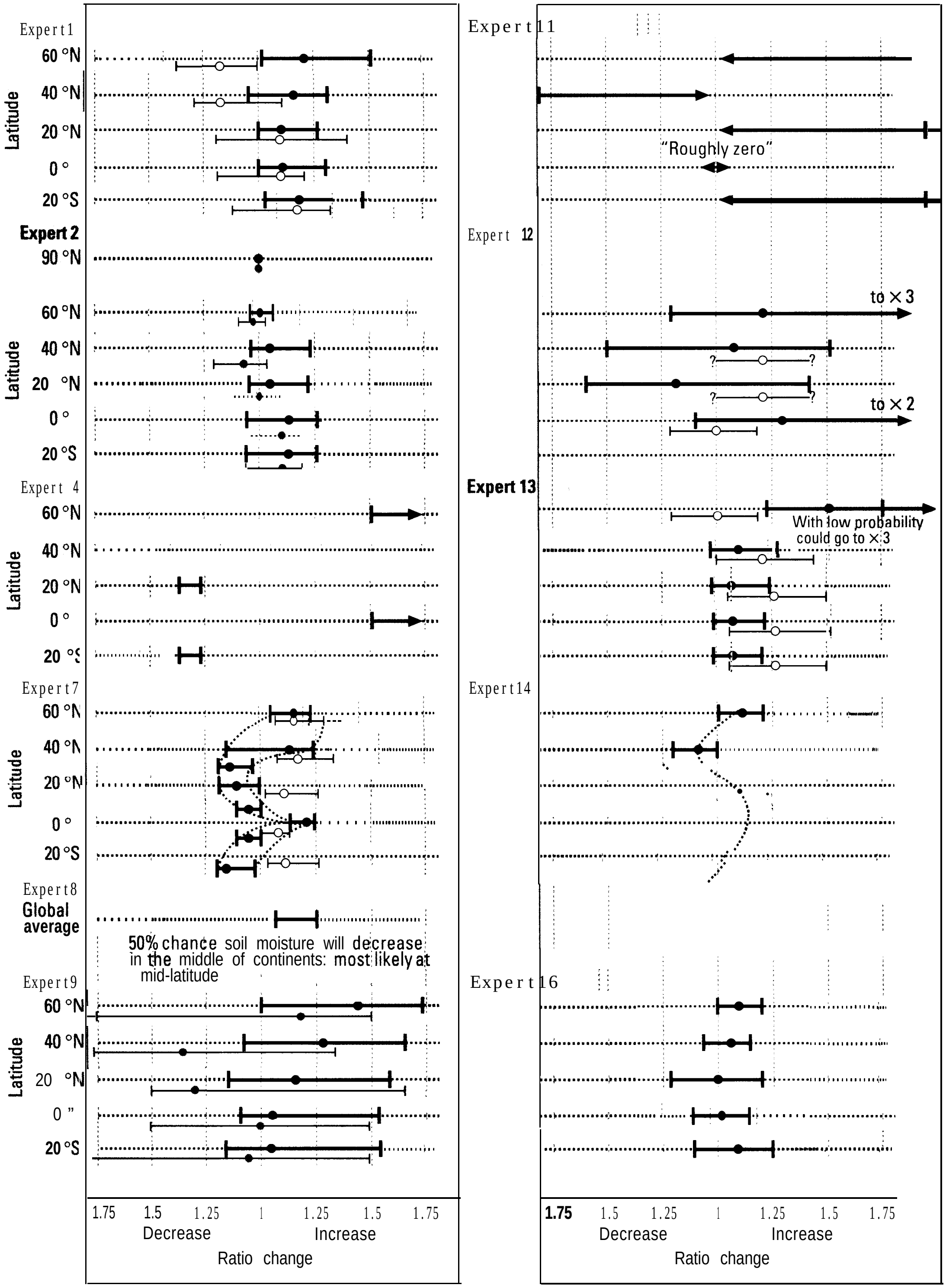
Anthropogenic aerosols generally are thought to cause cooling. Figure 4 summarizes responses to the question, "What is your current subjective probability distribution for the true value of the average contribution today of anthropogenic aerosols to radiative forcing in the Northern Hemisphere in W/m^2?" including both direct and indirect effects, reported as mean and 90% confidence interval. Of the 13 responses received, 4 of the means lie in the interval -1 to ≤ 0 and 9 lie in the interval -2 to ≤ – 1.

Our first question about climate sensitivity was

FIGURE 3

latitude differences

Solid dots and bold bars show "best estimates" and 90% confidence intervals for experts' judgments of "longitudinally averaged average annual precipitation over land" as a function of latitude for 2 x [CO_2]. Dashed lines were drawn by experts. Open dots and light bars show experts' estimated change in the standard deviation of interannual variability. Estimated changes in soil moisture are reported with shaded dots and light bars.



based on specifying a CO_2 concentration. However, because of incomplete understanding of biogeochemistry, there also is uncertainty about what concentration results from a given level of emissions. To obtain an estimate of this uncertainty, we asked experts to assume that the current mix of anthropogenic emissions was adjusted to a level judged most likely to result in the same amount of radiative forcing as 2 x [CO_2]. They estimated the associated uncertainty by editing a copy of the distribution they had provided for the 2 x [CO_2] scenario. The resulting means and standard deviations are reported in the left-hand part of Table 1. Question formulation constrained the resulting means to values similar to

TABLE 1

Responses of the 16 experts to various questions

Columns two to four report judgments about climate sensitivity of global average surface temperature, based on emissions rather than concentration. A value of 1 in column four means the expert believes uncertainty in going from emissions to concentration (i.e., uncertainty about biogeochemistry) is negligible compared with uncertainty in going from concentration to temperature change. The four center columns report responses to questions about multiple-climate states. The last column reports the chance that uncertainty about climate sensitivity will increase by more than 25% after a 15-year research program specified by the expert.

Climate sensitivity far changed emissions				Questions about multiple-climate states				Future uncertainty
Expert number	Mean	Standard deviation	Ratio of standard deviations: flux/conceit-tration	Are there multiple stable climate states?	Can forcing with CO_2 alone cause a state change?	Probability that 2 x CO_2 will cause a state change?	How much CO_2 to produce a 20% chance of state change?	Chance climate sensitivity uncertainty grows > 25% after a 15-year research program
1	2.2	1.2	1.4	yes	yes	< 0.05	8 × CO_2	10
2	2.9	4.4	1.5	yes	weak yes	0.1 to 0.2	na	18
3	2.6	2.3	1.0	yes	yes	0.2	na	30 (Note 4)
4	5.1	6.9	1.3	yes	yes	20.1	na	22
5	0.3	0.2	3.0	Note 1	no	implicit 0	na	30
6	2.7	2.0	1.0	yes	yes	0 to < 0.01	na	14
7	2.6	1.8	1.2	yes	weak no	implicit 0	na	20
8	3.2	2.1	1.5	yes	yes	0.05	4 × CO_2	25
9	3.1	3.2	1.8	yes	yes	0.01 to 0.1	na	12
10	2.7	1.2	1.2	yes	yes	very small	24 x CO_2	20
11	3.0	2.0	1.4	yes	no (Note 2)	na	na	40
12	3.3	1.5	1.4	yes	yes	0.5	na	16
13	2.2	1.4	1.4	yes	yes	0.25	na	12
14	3.0	1.1	1.1	no	no	na	na	18
15	2.9	1.8	1.4	yes	yes	0.02	10 × CO_2	14
16	3.0	1.4	1.4	yes	yes (Note 3)	na	na	8

na = Expert did not answer.
Note 1: Expert 5 views the climate system as a nonequilibrium system wandering through phase space.
Note 2: Expert 11 specified not at plausible CO_2 levels.
Note 3: Expert 16 observed that on time scales of hundreds to thousands of years significant state change could result from melting of Greenland ice cap and reduction in the thermohaline circulation.
Note 4: Expert 3 used a different response mode for this question. We gave a 30% of an increase by a factor of >2.5.

those in Figure 1. Experts broadened their distributions by between 0 and 80% with a mean increase of 30% (fourth column of Table 1).

In order to explore their beliefs about the dynamic response of the climate system, we gave the experts a linear CO, concentration ramp, which started at 353 ppm and reached a concentration of 560 ppm over the next 50 years. We then asked them to sketch the time response of the corresponding average surface temperature. We offered the option of stating a time constant and asymptotic value. Expert 10 gave us a mixture of two time constants.

To avoid an anchoring effect, we offered a choice between two different time resolutions. It is our impression that experts had firm views on this subject and that this precaution was unnecessary. Results are reported in Figure 5. Most experts displayed considerable consensus. Experts 2, 6, and 9 also gave confidence intervals (not shown), which in the case of Experts 2 and 6 included a small possibility of falling temperatures (by 2100 and 2150, respectively) after initial warming, because of a possible dramatic change in the large-scale thermohaline circulation.

In a series of questions we asked whether the climate system has other stable states; whether forcing from CO, alone could move the system to a new state in which it would remain for many years after the forcing was removed; and, if so, how much forcing with CO, would be required for a 20% chance of accomplishing this. Most believed that forcing by CO, alone could move the climate into another state (middle of Table 1). About half of the experts gave probabilities of > 5% and a quarter gave probabilities of > 20% that 2 x [CO,] could produce such a change. In describing a "new state," experts gave examples ranging from an end to the production of North Atlantic deep water to more extreme outcomes.

The experts generally agreed that predicting small-scale climate change by "downscaling" from the results of large-scale models was a tractable problem. Most mentioned two methods: nested dynamic models and statistical extrapolation from correlations between large and small scale in the current climate. We asked them to estimate the fraction of the uncertainty in small-scale climate prediction from uncertainties in predicting the large-scale response versus uncertainties in performing the downscaling. Most judged uncertainty in the large-scale sensitivity as the dominant factor in the total uncertainty, 9 of 10 ≥ 70%. In addition to three "don't knows," qualitative responses included words like "substantial" and "could be dominant."

Sources of uncertainty

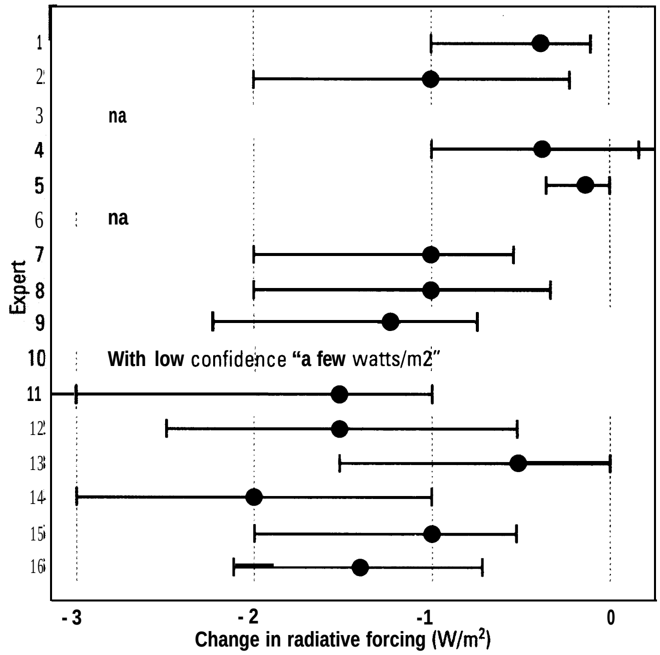
In Part 4, we explored the factors that contribute to uncertainty about climate sensitivity. We presented the experts with factors on two sets of cards. One set consisted of eight factors involving boundary conditions, including the biogeochemistry, that determine concentrations from anthropogenic emissions (denoted by BC in Table 2). The second set consisted of 14 factors that determine the climate sensitivity given a concentration (denoted by CS in Table 2). Experts reviewed our parsing of the problem and changed the wording on the cards until they were content that everything important had been included. They were then asked to rank order the two sets of factors in terms of contribution to uncertainty about climate sensitivity, assuming anthropogenic emissions were known. Finally, two cards having to do with the mechanics of modeling were introduced (denoted by MD in Table 2). Experts merged the three sets to obtain the top five factors contributing to their uncertainty. Table 2 lists the factors as they appeared on the cards, along with more detailed information for the top five. Because some experts ranked more than one factor at me same level, the second entry reports weighted ranks. Three different ordering procedures (number of mentions, weighted number of mentions, and weighted sum of number of mentions) yield the same top five factors across the set of all experts.

We asked how much uncertainty reduction could be achieved, in the 90% confidence interval, if research could completely eliminate the uncertainty associated with each of the five top-ranked items. This allowed a rough quantification of differences between the qualitatively ranked factors. The last column of Table 2 reports the mean of these estimates, which had a uniform distribution. Because the categories are not exclusive and the system is nonlinear, the sum of the ratio reductions may be >1.

FIGURE 4

Radiative forcing

Elicited expert subjective judgments of mean and 90% confidence interval on "the true value of the average contribution today of anthropogenic aerosols to radiative forcing in the Northern Hemisphere in W/m^2."

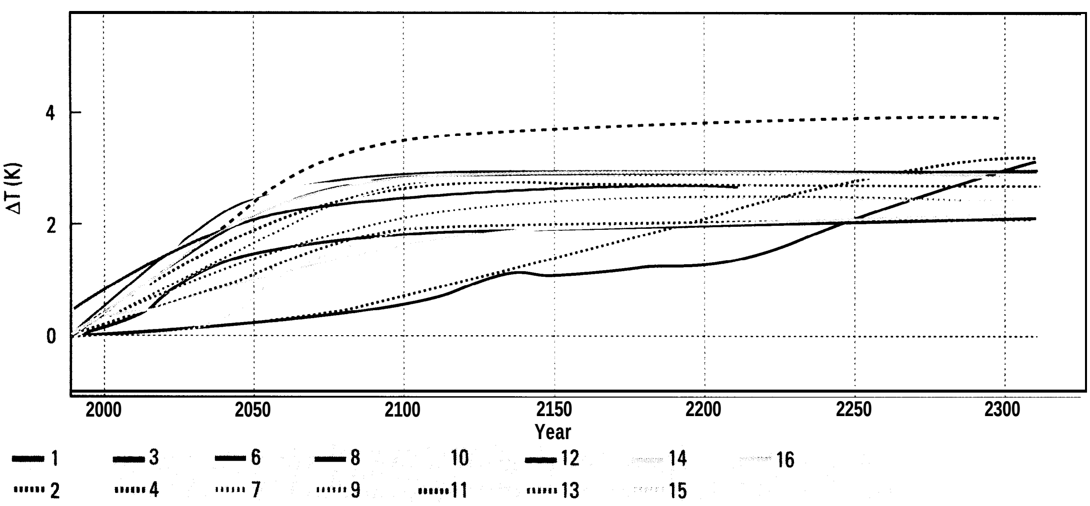


Part 5 addressed issues of research design and future research needs. We asked experts to discuss possible tensions between short-term prediction and other mission-oriented objectives and fundamental understanding. We had the experts "allocate a research budget of **$1** billion/year in climate science over a period of **15** years" to minimize uncertainty in a set of global-scale variables such as those considered in Part 3. To draw a sharp distinction be-

FIGURE 5

Temperature response

Summary of experts' judgments about climate system response to a linear concentration ramp that leads to a doubling of CO, over the next 50 years. Experts 2 and 6 also gave confidence intervals (not shown), which included some possibility of falling temperatures (by 2100 and 2150 respectively) after initial warming, because of the possibility of a dramatic change in the large-scale thermohaline circulation.



Experts interviewed in the study

Expert numbers used in reporting results do not correspond with either alphabetical order or interview order.

James Anderson, Harvard University
Robert Cess, State University of New York at Stony Brook
Robert Dickinson, University of Arizona
Lawrence Gates, Lawrence Livermore National Laboratories
William Holland, National Center for Atmospheric Research
Thomas Karl, National Climatic Data Center
Richard Lindzen, Massachusetts Institute of Technology
Michael MacCracken, U.S. Global Change Research Program
Syukuro Manabe, Geophysical Fluid Dynamics Laboratory
Ronald Prinn, Massachusetts Institute of Technology
Stephen Schneider, Stanford University
Peter Stone, Massachusetts Institute of Technology
Starley Thompson, National Center for Atmospheric Research
Warren Washington, National Center for Atmospheric Research
Tom Wigley, University Center for Atmospheric Research/National Center for Atmospheric Research
Carl Wunsch, Massachusetts Institute of Technology

tween a topic's importance and the marginal cost of making research progress, we posed the question in two parts. Experts first placed 50 chips on a specially designed game board indicating the relative importance of each of six categories. The game board was divided by "observational and data management" and "understanding and prediction" on one axis, and "radiative properties," "atmospheric and ocean dynamics," and "biogeochemical cycles" on the other. Then they received a much larger game board that preserved the six-part division but added two lower levels of disaggregation. Space precludes our providing more than a brief summary of the results. A draft manuscript that provides greater detail is available from the authors.

We found considerable diversity of opinion about the issues most needing research and the strategies most likely to yield improved understanding. However, a few generalizations are possible. Although there was little consensus on the importance of observational studies (16 to 76% of the importance weight) all experts allocated > 60% of their budget (mean = 71%) to observational and data management activities. Most allocated the largest amount to activities in atmospheric and oceanic dynamics (min = 28%, mean = 47%, max = 63% for observational and data management; and min = 27%, mean = 42%, max = 62% for understanding and prediction). Three experts allocated > 80% to observational activities. Of these, Experts 4 and 13 allocated the largest portion of their budget to studies of radiative properties (60% and 40%). Most allocated the smallest amount to observational studies of biogeochemical cycles (min = 9%, mean = 19%, max = 34%). Few of the experts work primarily in this area.

We found broad agreement that the research community does a poor job of using the results of process studies or field experiments on subjects such as cloud microphysics to improve large-scale models. Several argued this was due not only to the difficulty of such synthesis but also to systematic poor coordination within the atmospheric science community.

We found fairly wide agreement on the need to expand monitoring of key climate variables. Although most saw space-based observation as critical, many gave greater emphasis to smaller, cheaper, and more flexible platforms that allow more consistent coverage and more rapid programmatic adaptation. Low-cost, autonomous atmospheric and oceanic platforms received several mentions as deserving greater attention. The need for expanded oceanic and paleoclimatic studies and for expanded computational power also was emphasized frequently.

There was strong divergence of opinion about the relative importance of theory and modeling versus observational studies in reducing policy-relevant uncertainties. The extreme views may be caricatured as follows: "We have all the data needed to make substantial progress on reducing the uncertainty of climate prediction. We need to be smarter about using what we have, and we need more computational power" and "The problem is too difficult to solve in the next few decades before clearly detectable change may occur. Our best course is to build a strong and systematic climate monitoring program and interpret the results, using our best available models, to determine the climate sensitivity as change occurs." Observationalists tended to favor the second position, but surprisingly, some modelers strongly concurred.

On various occasions we asked about the utility of large "climate system" models that couple atmospheric and oceanic GCMs with models of the land surface, the carbon cycle, or atmospheric chemistry. Few experts believed that the results of such models should be trusted in a quantitative sense. However, many claimed that building such models was a healthy exercise because it encouraged scientists to consider related disciplines.

At the end of the budget discussion, we asked the experts to assume that $1 billion per year was being allocated to research in climate science and asked them to indicate what they considered a "socially appropriate" annual research budget for seven other areas. Average responses in millions of dollars per year were as follows: human emissions, 60; ecosystem response, 130; socioeconomic impacts and adaptation, 120; strategies for abatement, 630 (the highest 3 allocations were 800, 2000, and 5000; the average of the remaining 12 experts was 200); strategies for geoengineering, 47 (3 responses were 0); and integrated assessment, 38.

Exploring surprises

Having reminded the experts of their earlier mean and standard deviation judgements of climate sensitivity, we asked them to describe things that, if learned in the course of the research program they designed, would change the distribution of a repeated interview 20 years from now, after the research was completed. All were able to describe many such potential discoveries.

Several pursued a logic similar to that of Expert 4, who, when asked what could be learned "that would result in your concluding that your mean estimate of delta T was a good deal less than it currently is," responded "that would mean finding. . .either a new negative feedback or a feedback that was

TABLE 2

Uncertainty sources

Summary of results from the card-sorting task to determine sources of uncertainty about climate sensitivity. The first column produces the wording exactly as it appeared on the cards. The mnemonics BC (boundary conditions), CS (climate sensitivity), and MD (modeling details) denote the three groups in which the cards were originally presented before the final merge. In the body of the table the first value reported is the number of times the rank was assigned. In the second value, each expert's assignment of a rank to an entry is weighted inversely by the total number of entries that expert gave that same rank. The column labeled Weight is the weighted sum of the weighted ranks, where rank 1 is weighted 1, rank 2 is weighted 0.8, and so on, The last column reports the mean of the experts' estimates of how much uncertainty reduction there would be in their estimate of climate sensitivity, if all uncertainty about this factor were removed.

Wording on card	Rank 1	2	3	4	5	Weight	Mean % uncertainty reduction w/full info.
Cloud [distribution and] optical properties (including aerosol effects)-CS	9/7.7	1		1	3	9.20	3%
Convection/water vapor feedback (all processes transport water vertically *excluding* transport in the planetary boundary layer)-CS	6/3.53	3	2			7.13	35
Carbon dioxide exchange with terrestrial biota-BC	1/0.2	1/0.25	3/2.5	5/4.5	1	3.90	19
Carbon dioxide exchange with the oceans (including ocean biota)—BC		1/0.25	5/4	2/1.5	2	3.60	16
Oceanic convection (e.g., high-latitude production of deep water)-CS		3/2.33		1	1	2.46	22

The additional cards, listed in order of the weighted sum of the weighted ranks (in parentheses) were:

Spatial resolution of ocean models-MD (2.4)
Effect of flux correction on results of coupled models-MD (1.75)
Initial state of the ocean—CS (1.6)
Large-scale atmospheric dynamics (large-scale processes which transport heat, water, and momentum horizontally)—CS (1.43)
Large-scale oceanic dynamics--G (1.40)
Land surface interactions (hydrological properties including plant physiological response but *not* ecosystem change)—CS (1.35)
Ice-ocean feedback (salt pump)-CS (1.06)
Ice-af bedo feedback—CS (1.00)
"All ocean response and coupling".—CS (0.80) See note 2
Clear sky properties of aerosols—CS (0.08)
Solar ff ux variations-BC (0.80)
Ocean/atmosphere bulk transfer laws—CS (0.80)
Aerosol chemistry and physics (known emissions → aerosol densities, e.g., SO_2 to H_2SO_4)—BC (0.75)
Methane nonanthropogenic sources and *all* sinks (e.g., removal by OH)—BC(0.46)
Atmospheric boundary layer physics—CS (0.33)
Oceanic quasi-vertical mixing processes (heat and salt)—CS (0.15)
Atmospheric mixing processes (horizontal mixing, other than water, but *not* in the planetary boundary layer)—CS (0)
Other CHG nonanthropogenic sources and all sinks—BC (0)
Volcanic activity-BC (0)

Note 1: Expert 2 created a separate category "OH distribution and chemistry," which he ranked 1 along with four other categories. We have repotted it here under "methane nonanthropogenic sources and all sinks," although the basis of his concern was broader.
Note 2: Expert 8 created a new summary category called "all ocean response and couplings."
Note 3: Expert 11 split the distribution of clouds among "atmospheric boundary layer physics," "convection/water vapor feedback," and "large-scale atmospheric dynamics" rather than include it under "cloud optical properties."
Note 4: Expert 12 added "spatial resolution of atmospheric models" to "spatial resolution of oceanic models" and reported them together. We have reported this under spatial resolution of oceanic models.

much stronger than you thought." He discussed specific possibilities, focusing on clouds. Expert 5 noted that if new "paleo data were to demonstrate that average temperature. . .over the tropics was not stable. . .that would call for a revolution in the field" and elaborated some implications. Expert 2 focused on how the cycles for greenhouse gases might be affected, noting, for example, that if "the oxidizing capacity of the atmosphere becomes saturated. . .[that will] drive up the lifetime of methane."

Several experts emphasized discoveries that might be made about the ocean. Expert 6 said, "Suppose that the conveyer belt doesn't just simply turn off in the Atlantic, but it turns on in the Pacific; then it could be that the nature of the Pacific overturning could drive a lot more heat to high latitudes. . .so that somehow we could get a warmer climate." He focused particularly on switching between "a set of stable or metastable states."

We asked the experts to quantify their judgment by distributing 50 checkers (2% each) on a 6 x 7 cell game board in the space of mean values (μ) and standard deviations (σ) for their possible future estimates of climate sensitivity under forcing from the mix of all anthropogenic sources. For example, the center cell of the board encompassed the range of outcomes 0.75μ to 1.25μ and 0.750 to 1.250. By integrating the data over μ, we get the experts' subjective probability distributions of the change in their estimate of uncertainty in climate sensitivity attributable to the results of the research program that they view as most likely to reduce uncertainty. Results are summarized in the rightmost column of Table 1. Only five specifically said that uncertainty could decrease by more than 75% (mean probability estimate 0.09). Most expect a modest reduction in uncertainty (weighted mean reduction of 8%). They gave probabilities of between 0.08 and 0.40 (mean = 0.18) that, after the research was completed, they would be > 25% more uncertain than they are today.

These results, along with the estimates of uncertainty reduction reported in Table 2, strongly suggest that our ability to predict the gross character of climate change will improve slowly, even with well-designed research programs. These findings appear to be at odds with the consensus view expressed in the IPCC document (*3*), which predicts "substantial" reduction in uncertainty in the next 10–15 years.

Reasons for consensus

Where possible we checked for biases from question formulation by assuming that experts must be consistent in certain ways. For example, we assumed that it is inconsistent to believe that flux-to-concentration feedbacks introduce no extra uncertainty in climate sensitivity (Part 3) while asserting that removal of all uncertainty about a biogeochemical factor would reduce overall uncertainty (Part 4). We identified three experts who had expressed such views, and asked them to reconsider.

We can hypothesize four possible sources of the high degree of consensus observed among the experts in their judgments of climate sensitivity, and the much lower consensus about meridional temperature gradient and zonally averaged precipitation: varying degrees of familiarity with the relevant science, differences in the intrinsic scientific difficulty of the questions, conventional psychological anchoring, and differences in the cognitive difficulty of various response modes.

From the responses in Part 6, it is possible to construct an independent estimate of climate sensitivity. We combined normal distributions with means and standard deviations given by their position in the (μ,σ) space using the elicited weights. Because these judgments were made by systematically considering the relevant sources of uncertainty (the card-sorting task) and then designing a research program to address them, the result arguably constitutes a more carefully considered judgment than the initial holistic judgment of sensitivity. The results are almost identical. Had there been a strong psychological anchoring in the initial judgments that prevented experts from incorporating possible surprise in their answers, it should have been revealed here. The strong consistency between responses suggests that experts have robust beliefs about the uncertainty associated with climate sensitivity.

We hypothesize that the disagreement about meridional temperature gradient results primarily from the experts' unequal familiarity with the relevant science rather than from question formulation or intrinsic difficulty. Precipitation is known widely as an inherently difficult problem. We hypothesize that it is this difficulty that leads to the divergent responses we received.

Acknowledgments

This research was supported by the Electric Power Research Institute, the National Science Foundation, a travel grant from NOAA, and academic funds from Carnegie Mellon University. We thank the 16 experts listed for their time and thoughtful and cheerful cooperation. Thanks for advice or assistance go to Cliff Davidson, Hadi Dowlatabadi, Asma Fadra, Baruch Fischhoff, Paul Fischbeck, Filippo Giorgi, Ann Henderson-Sellers, Milind Kandlikar, Mark Handel, Lester Lave, Charles Linville, Anand Patwardhan, James Risbey, Elena Shewliakowa Siegel, Richard Sonnenblick, Ed Rubin, Mitchell Small, and Patti Steranchak.

References

(1) Dalkey, N. C. "The Delphi Method: An Experimental Study of Group Opinion"; Technical Report RM-5888-PR; The Rand Corporation, Santa Monica, CA, 1970.
(2) Linstone, H.; Turoff, M. *The Delphi Method: Techniques and Applications*; Addison-Wesley: Reading, MA, 1975.
(3) Intergovernmental Panel on Climate Change. *Climate Change: The IPCC Scientific Assessment*; Houghton, J. T.; Jenkins, G. E.; Ephraums, J. J., Eds.; Cambridge University Press: New York, 1990.
(4) "Policy Implications of Greenhouse Warming: Mitigation, Adaptation and the Science Base"; Report by the Panel on Policy Implications of Greenhouse Warming of the Committee on Science, Engineering and Public Policy; National Academy Press: Washington, DC, 1992.
(5) De Groot, M. *Optimal Stntisticnl Decision;* McGraw-Hill: New York, 1970.
(6) Watson, S. R.; Buede, D. M. *Decisiorl Synthesis: The Principles and Practice of Decisiorl;* Cambridge University Press: New York, 1987.
(7) von Winterfeldt, D.; Edwards, W. *Decisiorl Analysis and Behavioral Research*; Cambridge University Press: New York, 1986.
(8) Morgan, M. G.; Hem-ion, M. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*; Cambridge University Press: New York, 1990.
(9) Spetzler, C. S.; Staël von Holstein, C-A. S. *Manage. Sci.* **1975,** *22*(3), 340-52.
(10) Wallstein, T. S.; Whitfield, R. G. "Assessing the Risks to Young Children of Three Effects Associated with Elevated Blood Lead Levels"; ANL/AA-32; Argonne National Laboratory: Argonne, IL, 1986.
(11) Morgan, M. G. et al. *Risk Analysis* **1984,** *4*(3), *201-16.*
(12) "The Global Climate Change Integrated Assessment Program"; Department of Engineering and Public Policy; Carnegie Mellon University: Pittsburgh, PA, 1995.
(13) National Defense University. "Climate Change to the Year 2000: A Survey of Expert Opinion"; Report published by the National Defense University in cooperation with the U.S. Department of Agriculture, the Defense Advanced Research Projects Agency, the National Oceanic and Atmospheric Administration, and Institute for the Future, Washington, DC, February 1978.
(14) Stewart, T. R.; Mumpower, J. L.; Reagan-Cirincione, I? "Scientists Agreement and Disagreement about Global Climate Change: Evidence from Surveys"; Report for the Center for Policy Research, Nelson A. Rockefeller College of Public Affairs and Policy, State University of New York: Albany, NY, April 1992.
(15) Nordhaus, W. D. *Am. Sci.* **1994,** *82*(1), *45-5 1.*
(16) "Protection Against Depletion of Stratospheric Ozone by Chlorofluorocarbons"; Report by the Committee on Chemistry and Physics of Ozone Depletion and the Committee on Biological Effects of Increased Solar Ultraviolet Radiation, National Research Council; National Academy Press: Washington, DC, 1979.
(17) National Research Council. *Stratospheric Ozone Depletion by Halocarbons: Chemistry and Transport*; National Academy Press: Washington, DC, 1979.
(18) Kahneman, D.; Slovic, I?; Tversky, A. *Judgment Under Uncertainty: Heuristics and Biases*; Cambridge University Press: New York, 1986.
(19) Keith, D. "Briefing Notes for Elicitation of Expert Judgments About Uncertainties in Climate Prediction"; National Center for Atmospheric Research and Carnegie Mellon University: Pittsburgh, PA, April 1993.
(20) "Questionnaire and Elicitation Workbook for Climate Experts"; Department of Engineering and Public Policy; Carnegie Mellon University: Pittsburgh, PA, 1994.
(21) Robock, A. *Bull. Am. Meteorol. Soc.* **1985,** *66*(7), *786-87.*
(22) Kellog, W.; Schneider, S. *Science* **1974,** *186*(4170),1164–72.