ELSEVIER International Journal of Forecasting 10 (1994) 387–403

international journal of forecasting

What forecasts (seem to) mean

Baruch Fischhoff*

Department of Social and Decision Sciences, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

A forecast is just the set of probabilities attached to a set of future events. In order to understand a forecast, all one needs to do is to interpret those two bits of information. Unfortunately, there are pitfalls to communicating each element, so that the user of a forecast understands what its producer means. One source of potential problems is ambiguity regarding the event being predicted and what exactly is being said about it. Another is the difficulty of determining the relevance of the problem that the forecaster has solved for the problem that the user is facing. Problems can also arise out of epistemological and sociological issues of trust and context. A simple framework is offered for considering these communication problems and is then illustrated with a mixture of systematic data and anecdotal observation. The criticality of these different problems is considered, along with procedures that might reduce them.

Keywords: Forecasting; Ambiguity; Risk communication; Uncertainty; Subjective probability

1. Introduction

Several months ago, I was approached by a producer from the Weather Channel, a 24-hour-a-day cable television station devoted to weather forecasting, for help with a forthcoming special (Superstorm, 1993). The Weather Channel's staff had been shaken by the high death toll from the massive winter storm that hit the Eastern Seaboard of the United States in March 1993. From the Channel's perspective, its experts had seldom had such clear indicators of a major weather event, nor expressed themselves with such confidence. Nonetheless, hundreds of people still died, from exposure, heart attacks, falling tree limbs, road accidents, and the like. No one has counted the number of injuries, illnesses, and close calls. The special was intended to encourage surviving viewers to take future forecasts more seriously. The production process seemed designed to help the Channel itself figure out where it had failed.

Obviously, the value of forecasts comes from providing needed answers in a usable form. Somehow or other, the Weather Channel's 'superstorm' forecasts either were not answering the questions that viewers were asking or were not being understood as intended (assuming, of course, that the people who became casualties had heard the Channel's warnings or the relayed messages of others who had). Although the failures of these forecasts were particularly

* Tel: +1 412 268 3246; bf0r@ANDREW.CMU.EDU

SSDI 0169-2070(94)00542-K

dramatic and tragic, I suspect that most forecasters believe that they have seen clients suffer when their forecasts were ignored or misinterpreted.

Working with the Weather Channel got me to thinking about the general conditions under which the message of a forecast can get lost in the transmission. At first glance, the communication task seems straightforward. All that needs to be explained is the probability attached to some future event. That same basic challenge arises whether one is predicting weather, economic events, nuclear power plant performance, surgery outcomes, or market penetration.

There seem to be four basic ways in which one could fail to communicate either the probability or the event portion of a forecast. In each case, the problem may be either that we have nothing to say about that aspect of the forecast, because we have not done the needed research, or that we are not saying it well enough to be understood by our audience.

(1) Ambiguity: not saying clearly what events we are forecasting or how likely we think that they are.

(2) Irrelevance: not offering forecasts that address our clients' informational needs.

(3) Immodesty: not admitting the limits to our knowledge.

(4) Impoverishment: not addressing the broader context within which forecasts (and contingent decisions) are made.

The following sections examine each of these possibilities in turn, drawing on research and anecdotes concerning a variety of forecasting tasks. Each section looks separately at interpreting probabilities and events. The concluding section considers the risks and benefits of both ignoring and acknowledging these issues.

2. Ambiguity

It is hard to follow any news medium for very long without encountering a health or safety warning, such as 'don't drink and drive,' 'practice safe sex,' 'don't do drugs,' or 'avoid apples treated with Alar.' Each such warning carries an implicit forecast regarding the occurrence of some misfortune, should the ill-advised course of action be pursued. Unfortunately, these are typically not very precise forecasts. They fail to say exactly which actions should be avoided, which consequences can follow, or how likely that connection is. Recipients who hoped to act on these warning would have to guess what exact message was intended. Their guess could be wrong, if, for example, those who issued the warning had different linguistic norms or thresholds of concern (reflecting their aversion to particular outcomes and the costs of avoidance).

2.1. Probabilities

The ambiguity in qualitative expressions of probability is well documented, in this journal (Beyth-Marom, 1982) and elsewhere (Lichtenstein and Newman, 1967, and Wallsten et al., 1986). A given term such as 'likely' or 'rarely' can be interpreted as implying different probabilities to different people in a single context and different probabilities to the same person in different contexts. Such ambiguity has been found even within communities of professionals, such as physicians and intelligence officers (e.g., Beyth-Marom, 1982, and Merz et al. 1991). Fig. 1 shows typical results.

The criticality of such ambiguity depends, of course, on how the forecast is used. Sometimes, an inferred probability of 1% and 10% will lead to the same choice; sometimes not. The risk of letting consumers make their own interpretations presumably increases with the diversity of potential users and contingent decisions – in the sense of there being a greater chance of people making wrong guesses and then, as a result, making wrong choices.

There are some interesting studies examining the variability in people's interpretation of different verbal labels in circumstances (e.g., Cohen et al., 1958, Hamm, 1991, Holyoak and Glass, 1978, Poulton, 1989 and Zimmer, 1984). In theoretical terms, these studies have improved our understanding of how people extract meaning from context (psycholinguistics) and how they use numeric scales (psychophysics). In

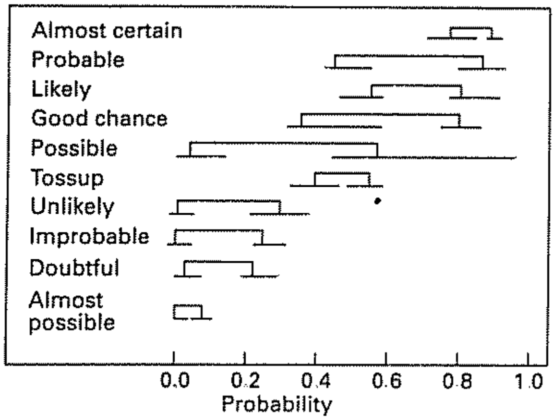

Fig. 1. Upper and lower probability limits for probability expressions (Upper bars show mean lower and upper probability limits provided by subjects for each verbal expression Lower left- and right-hand bars show the interquartile ranges provided for the lower and upper limits, respectively.) Source: Wallsten et al. (1986)

practical terms, these studies can show how to use the least ambiguous terms, making clear which risks are posed by the residual ambiguities. However, the sensible solution is just to be explicit. If one has, in fact, done the work needed to determine a probability, then give it. If recipients are uncomfortable with numbers, then let them substitute the verbal label that works for them.

2.2. *Events*

Much less research attention has been given to potential ambiguity in the events to which these probabilities (or verbal likelihood labels) are attached. Imagine, for example, a public service announcement saying, 'If you drink and drive, you will have a 5% chance of getting into an accident.' The number provides some information. However, its practical implications would be very different if it was referring to the risk from driving home following a single beer rather than from a lifetime of trips made by a heavy drinker.

Some years ago, Allan Murphy, who has worked hard to institutionalize the use of explicit probabilities in weather forecasting (e.g., Murphy and Winkler, 1984), found his accomplishments threatened by the claim that such forecasts confused lay people. Collecting behavioral data to discipline these speculations, Murphy et al. (1980; also Murphy and Brown, 1984) found that the confusion concerned the event and not the probability. Thus, participants in the initial study were divided fairly evenly over whether 70% referred to the area that would experience rain, the time that it would rain, or the chance of at least some rain somewhere.

The converse of this ambiguity emerges in surveys that ask respondents to estimate the likelihood of some event, whose exact identity they must guess. For example, the National Health Information Survey (1987) is a well-funded effort which is used, among other things, to assess the effectiveness of the AIDS information programs of the Centers for Disease Control. It asks questions such as, 'How likely is it that you can get AIDS from sharing plates and other eating utensils?' We presented this question to a class at an Ivy League college. After answering, they were asked what they thought had been meant by 'sharing plates and other eating utensils.' Table 1 shows the distribution of their choices among the options that we offered for this and a second question from the survey. Even within this fairly homogeneous sample, there was considerable disagreement over what event was meant (Fischhoff, 1989, and Linville et al., 1993).

Moreover, even if all respondents agreed about the interpretation of an event, one would still have to guess what that was. In Table 1, that guessing requires an understanding of the culinary habits of a population of young adults. Other bits of cross-cultural knowledge would be needed for other questions. Presumably, the problems that we face in interpreting others' forecasts have their counterparts in their attempts to interpret ours. Thus, similar research would be needed to anticipate what people would infer from the forecast that, 'according to the best available technical knowledge, the probability of HIV transmission is 3×10^{-11} for sharing plates and other eating utensils.'

Quadrel (1990) found even greater disagreement, using an open-ended approach to investi-

Table 1
Interpretations of AIDS risk questions

How did you interpret 'sharing plates, forks, or glasses with someone who has AIDS?'
- Sharing utensils during a meal (e.g., passing them around, eating off one another's plates) (81.8%)
- Using the same utensils after they have been washed (10.9%)
- I was uncertain about the interpretation (5.8%)
- Multiple interpretations (1.5%)

Did you interpret 'sharing plates, forks, or glasses with someone who has AIDS' as
- Occurring on a single occasion? (39.1%)
- Occurring on several occasions? (19.6%)
- Occurring routinely? (27.5%)
- I was uncertain about the interpretation (12.3%)
- Multiple interpretations (1.4%)

How did you interpret 'having sex with a person who has AIDS?'
- Having vaginal intercourse without a condom (72.5%)
- Having vaginal intercourse with a condom (4.3%)
- Having other kinds of sex (6.5%)
- I was uncertain about the interpretation (6.7%)
- Multiple interpretations (8.0%)

Did you interpret 'having sex with a person who has AIDS' as
- Occurring on a single occasion? (61.6%)
- Occurring on several occasions? (22.5%)
- Occurring on many occasions? (7.2%)
- I was uncertain about the interpretation (5.1%)
- Multiple interpretations (3.6%)

Entries are the percentage of subjects (in a sample of 135 students at an Ivy League College) who reported having inferred each definition of the phrase when they had answered a question about the risk that it entailed.

gate the same issue. She asked adolescents to think aloud as they assessed the probability of deliberately ambiguous events like 'getting into an accident after drinking and driving' or 'getting AIDS as the result of sex.' She found that they raised a wide variety of concerns, most of which are arguably relevant to an objective determination of risk. Each column in Table 2 shows the number of her subjects who thought that each of three aspects of the 'dose' of a risky event were relevant to the probability of a risky outcome. For seven of these nine events, they felt the need

Table 2
Intuitive interpretations of ambiguous event descriptions[a]

Dose	Drink and drive	Smoke and cancer	Cocaine and addiction	AIDS	Smoke and addiction	Cocaine and health	Marijuana	Alcohol and addiction	Pregnancy
Amount	49	52	39	2	40	31	32	41	5
Potency	15	13	1	0	12	5	7	8	0
Method	4	3	7	4	1	10	10	1	4

[a] Source: Quadrel (1990); Quadrel et al. (1994).
Entries are the number of subjects (in a sample of 61 adolescents) who spontaneously referred to each aspect of the dose involved in nine incompletely described risk events. For example, 49 raised the issue of how much 'drinking' or 'driving' was implied in a vague statement about that risk.

to make some assumption about the amount of exposure.[1] In a sense, these teens (many of whom came from treatment homes) were more sophisticated than the investigators who designed the surveys from which these ambiguous questions were taken or adopted. Presumably, such teens are similarly confused about the meanings of the ambiguous risk forecasts communicated to them in health classes and public service announcements.[2]

Competent professionals know exactly what they are talking about when they issue forecasts. However, they have an obligation to determine what their clients think they are talking about. Problems can arise even with commonly issued forecasts. There may be no natural way for recipients to realize that they have consistently misinterpreted a term like unemployment, bankruptcy, viewer, room, GNP, or race. Survey researchers sometimes call such ostensibly objective and common terms, *quasi-facts* (Turner and Martin, 1984). They have created some useful worked examples of how to improve the understanding of technical terms, and to assess their residual ambiguity (Bailar and Rothwell, 1984, and Turner et al., 1992). Other help can be found in the psychological literature on categorization (Murphy and Medin, 1985) and its applications to tasks like improving the accessibility of information in places like the Yellow Pages or computerized databases (Fischhoff et al. 1987, and Furnas et al., 1983). One design consideration is to decide when one might sacrifice clarity in order to use terms whose ambiguities are more readily apparent (Fischhoff and MacGregor, 1986; Fischhoff et al., 1987).

[1] The two exceptions were troubling for quite a different reason. Her subjects did not see the relevance of the amount of sex to the probability of AIDS and pregnancy (see also Linville et al., 1993)

[2] A minister doing pastoral work with AIDS sufferers recently told me about a patient who said, 'I thought that the risks to intravenous drug users didn't include me, since I only skin popped.' There is, of course, always some chance of wilful misinterpretation, as people construe evidence to fit the actions that they intended to take anyway. However, part of the expert's job is to be explicit enough to close such loopholes.

An exogenous barrier to communication arises when the technical community changes its definitions to reflect changes in law, theory, or methodology. As a result, it is now asking about or telling about a different event, even though the wording remains the same. For example, over time, there may be shifts in the meaning of terms like 'risk,' 'unemployed,' 'privacy,' 'electronics,' 'child abuse,' or 'electorate.' Those changes may proceed in different directions and at different paces for different groups. A subsequent section deals with the false immodesty in a forecast that fails to reveal that it reflects a best guess at both how an event should be defined and how likely it is to happen.

3. Irrelevance

Even if it is well understood, a forecast is of little use unless it provides information that potential consumers need. Indeed, it is the desire to be more relevant that motivates most revisions in the terms, models, and analytical procedures of forecasts. For example, serious attempts are now being made to include environmental effects in national accounts and, hence, in forecasts of the effects of various public policies and natural processes. The results should be more relevant to those concerned with the health of nations, even if they will need to be educated about the nuances of a new set of measures (Costanza, 1991).

3.1. Probabilities

Whatever the event, users need not only a best guess at how it will turn out, but also some indication of the confidence that can be placed in that estimate. Having more confidence in a forecast allows one to take more decisive action, to curtail information collection, to plan for a narrower range of possible contingencies, and to invest less in vigilance for surprises.

When forecasters fail to express their uncertainty, then their clients must guess. If clients guess wrongly, then they risk unduly aggressive or cautious action. The former might happen if they assumed that estimates lacking any qualification needed no qualification. The latter might happen if clients refused to trust experts who gave unqualified predictions. In either case, the errors would reflect a failure to provide the kind of estimate most relevant to the client's purposes (Fischhoff et al. 1978, Krysztofowicz, 1983, and Morgan and Henrion, 1990).

When forecasting continuous events (e.g., interest rates, market share), a familiar convention is to offer a probability distribution over possible values (and not just the mean, median, or mode). A less ambitious expression of uncertainty is that of confidence intervals, bounding the set of possible values with some degree of certainty. With discrete events (will/will not happen), the associated probability is sometimes held to incorporate all relevant beliefs. The less the uncertainty, the closer the probability will be to 0 or 1. Others argue that intermediate probabilities can be held with different degrees of confidence, the knowledge of which can improve contingent decisions. Whatever the logical status of such 'second-order probabilities,' analysts may owe clients whatever insight they have about the stability of their knowledge (Gärdenfors and Sahlin, 1982 and 1989, Morgan and Henrion, 1990, Shafer, 1975, and von Winterfeldt and Edwards, 1986).

The definitiveness of an estimate depends on the extent of the knowledge upon which it is based. An estimate could, for example, reflect the data in a single study, the data from all similar studies, or the data and observations from all sources available to the forecaster. Such summaries might be based on classical statistics, meta-analysis, and Bayesian statistics, respectively. Each strategy can be followed competently. However, users will be misled if they assume greater or less comprehensiveness than is warranted.

For example, Fortney (1988) reported the results of a meta-analysis on all available studies of the health effects of oral contraceptives. She concluded, with great confidence, that a non-smoking woman who used the Pill throughout her reproductive career would do something between increasing her life expectancy by 4 days and decreasing it by 80 days. In addition, she was able to say that it was highly unlikely that this forecast would change because the existing data base was so great that no conceivable study could materially change the conclusions. The value of such an explicit expression of uncertainty can be compared to the implicit claims of definitiveness accompanying the typical newspaper account of the hot new medical study.

3.2. Events

It is a conversational norm to tell people things that they need to know (Grice, 1975). As a result, the issuing of forecasts carries an implication of relevance. Nonetheless, it is easy enough to think of processes leading to forecasts that violate this expectation. For example, government agencies may routinely issue forecasts of economic variables having great relevance to some members of its audience, but none to others (depending on the decisions that each faces). Sometimes, maintaining continuity in a time series may mean staying committed to questions and forecasts of decreasing relevance. Authorities may present laundry lists of potential side effects, predicting what might happen as the result of medical procedures, with little guidance as to what really matters. Disproportionate effort may go into refining estimates of the readily calculable parts of a problem, even when those constitute a relatively small portion of it. Forceful analysts may seize center stage for the results of whatever it is that they compute.

The work can be competently executed, qualified, and communicated, yet still miss the mark if it fails to address recipients' concerns. If this irrelevance, or marginal relevance, is recognized, then recipients may resent the misappropriation of their time and resources. If it is not recognized, then recipients may misidentify the nature of their concerns, assuming that a result must be important if an apparent expert brings it to their attention.

Relevance is defined most clearly within the context of specific decisions. One can then use techniques like value-of-information analysis to determine how much of a practical difference knowing a forecast makes. Even posing the question can be informative. For example, public health officials in several countries have tried to promote testing for radon by showing forecasts of the risks that it poses to individuals and to society as a whole. A little thought shows that the homeowners also need an estimate of the probability that they can reduce the risk for an investment within their budget (Evans et al., 1988, and Svenson and Fischhoff, 1985). However, institutional priorities and inertia have led to much less being spend on improving estimates of remediation efficacy.

Merz (1991) and Merz et al. (1993) performed a formal analysis of the relevance of risk estimates for a dozen side effects of carotid endarterechtomy, a surgical procedure that reduces the risk of a stroke by scraping plaque from the major artery leading to the brain. Unfortunately, things can go wrong, including the general risks of surgery. Merz simulated the impact of learning about the probabilities of these side effects for a population of hypothetical patients (differing in their physical states and preferences). He found that knowing about a few of these side effects would change the decisions of a significant fraction of patients. Most of the side effects, however, were relevant to only a minute portion of patients. He argued that, while nothing should be hidden from patients, physicians should concentrate on communicating the few critical forecasts. Between the time that Merz submitted his dissertation and its defense, the results of a major clinical trial were released. Incorporating them in his model made little difference to its conclusions. That is, from this perspective, the trial had little practical importance, whatever it might have contributed to the understanding of fundamental physiological processes. It would take a separate analysis to determine whether, with foresight, the trial had any reasonable chance of producing decision-relevant information.

In other cases, it may take a major intellectual effort to discern the logical structure of the decisions that forecasts might serve. At present, the US spends something of the order of $1–2 billion dollars annually to forecast the effects of the build-up of greenhouse gases (Office of Technology Assessment, 1993). Much more of that research is focused on predicting climate, through large-scale atmospheric and oceanographic models, than on predicting social and economic effects (or on the impacts of possible remedial strategies). It is hard to find any reasoned basis for this allocation, beyond institutional politics. The observer of this research might reasonably conclude that we need to understand climate before we can start to think seriously about related actions. The counterargument is that, without some serious analysis of the relevance of forecasts, it is hard to know if forecasting resources are being invested wisely and if their conclusions are being given the proper weight (Dowlatabadi and Morgan, 1993, and Rubin et al., 1992). Waiting for excellent information can mean paralysis, in effect denying the need to face the kinds of gambles being taken. Figuring out what events to forecast may be as interesting and important as actually doing the forecasting.[3]

4. Immodesty

Unless forecasters say how confident they are in forecasts, recipients are left to guess. However, even when confidence or lack of it is stated, recipients may still be left guessing whether the forecasters have overstated or understated how much they know. Like other possible biases, these, too, can have both motivational and cognitive roots. On the motivational side, the dominant incentive sometimes leads one to exaggerate one's confidence, for example, when business or attention goes to those who appear most knowledgeable. At other times, it pays to hedge aggressively, so as to avoid public commitment and accountability. The research

[3] A problem with analogous complexity is prioritizing the testing of chemicals for toxicity from the very large matrix of possibilities: 50 000 or so chemicals, various toxic endpoints and several tests for each, which could be conducted some number of times (National Research Council, 1983).

literature suggests cognitive processes hampering people's ability to evaluate themselves.

4.1. Probabilities

As mentioned, uncertainty about forecasts is typically expressed in terms of the probabilities of discrete events occurring or subjective probability distributions over possible values of continuous quantities. Many studies, some appearing in this journal, have shown similar patterns: individuals have some, imperfect, understanding of how much they know. The most commonly observed overall tendency has been over-confidence, which increases as people's knowledge decreases and shifts to under-confidence when they know a lot (Lichtenstein and Fischhoff, 1977, Lichtenstein, et al., 1982, and Yates, 1989).

Fig. 2 shows an early representation of this pattern. It shows the proportion of correct predictions among those statements assigned each probability of being correct. Over-confidence is reflected in points below the diagonal. One interpretation of the overall trend is that people enter a task expecting some intermediate level of difficulty and adjust insufficiently when tasks prove harder or easier. One interpretation of the flatness of each curve is that people feel that they can (and, perhaps, should) use the entire response scale, but are unable to make the necessary distinctions between different levels of knowledge.[4]

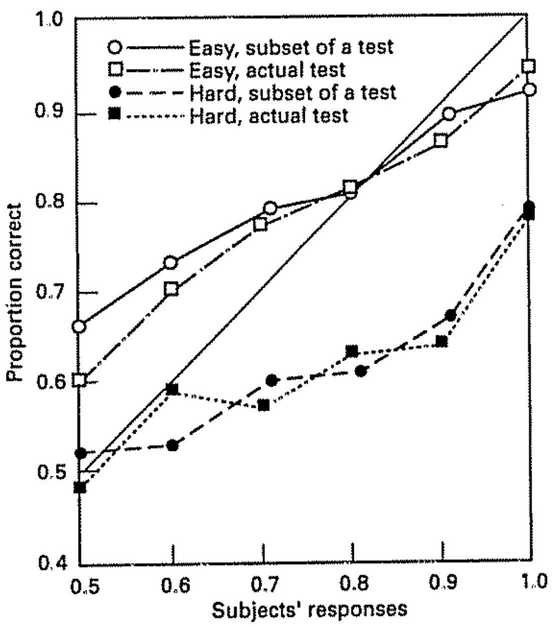

Fig. 2. Calibration curves for tasks varying in difficulty level. Statements assigned each probability of being correct (0.5, 0.6, . . . , 1.0) are grouped. Each point on each curve represents the proportion of statements assigned the indicated probability that were, in fact, correct. Source: Lichtenstein and Fischhoff (1977).

The critical question for the consumers of forecasts is whether similar patterns of behavior will be observed in the kinds of experts who produce them. Ayton (1992) provides a recent critical review of this topic, finding, it would seem, enough evidence of inappropriate expert confidence (mainly over-confidence) to justify caution. For practical purposes, as long as there is some significant and unpredictable tendency for expert miscalibration, users are left guessing about how much trust to place in forecasts.

After a review conducted 10 years earlier (Fischhoff, 1982), I concluded that confidence assessment could be viewed as a complex skill, whose acquisition requires the proper conditions for learning: prompt, unambiguous feedback with rewards for performance (and not, say, for bragging or hedging). The excellent performance of weather forecasters assessing the probability of precipitation (Murphy and Winkler, 1984) fits this pattern, as does that of expert bridge players predicting whether they will make a contract (Keren, 1987).

Figs. 3–5 present several expressions of these potential problems. Fig. 3 shows (something akin to) subjective confidence intervals, used by particle physicists to express confidence in estimates

[4] Yates (1989) offers an insightful summary of the research results and methods. One topic attracting some attention currently is an apparent discrepancy between the confidence expressed in a set of forecasts and in the individual members of a set (Sniezek and Buckley, 1991, Sniezek et al., 1990). As this topic evolves, the critical question for forecasters is which task is more pertinent. That is, are decisions made on the basis of confidence in individual forecasts or in set of forecasts?

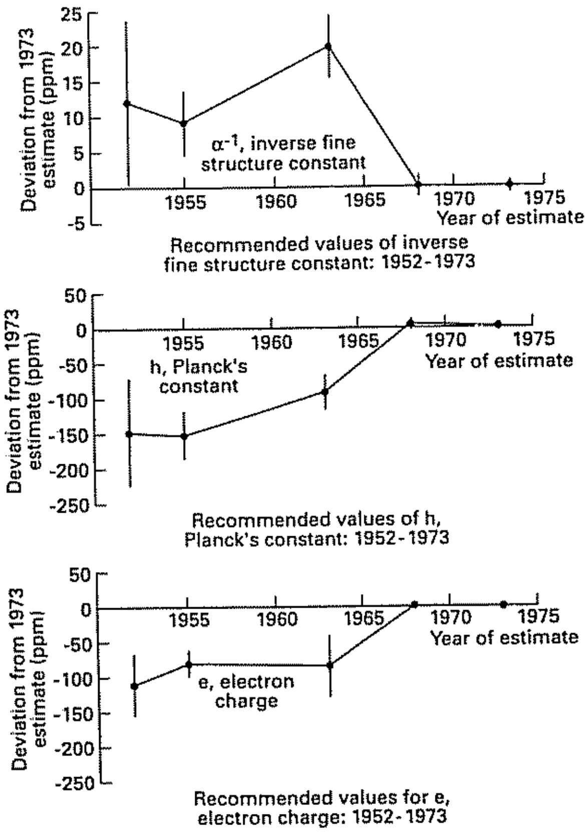

Fig. 3 Confidence intervals surrounding estimated values of physical constants. Source: Henrion and Fischhoff (1986).

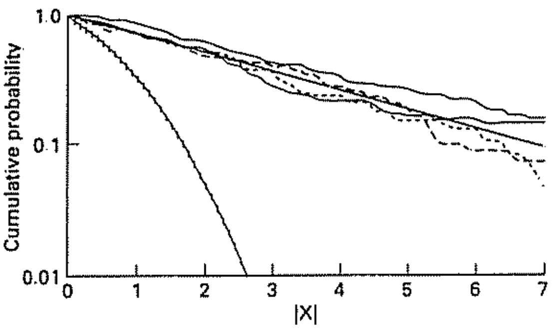

Fig. 4. Annual energy outlook forecasts. The plots depict the cumulative probability that new measurements, a, will be at least $|x|$ standard deviations d away from the old results A; $x=(a-A)/d$. 1983 to 1990 (heavy dotted line); 1985 to 1990 (dashed line); 1987 to 1990 (solid line), aggregated forecasts for all three forecast years (heavy dashed line); exponential distribution, $e^{-|x|/3}$, corresponding to $u=3.4$ (heavy solid line); gaussian (thin solid line with vertical bars). Source: Shlyakhter et al. (1994).

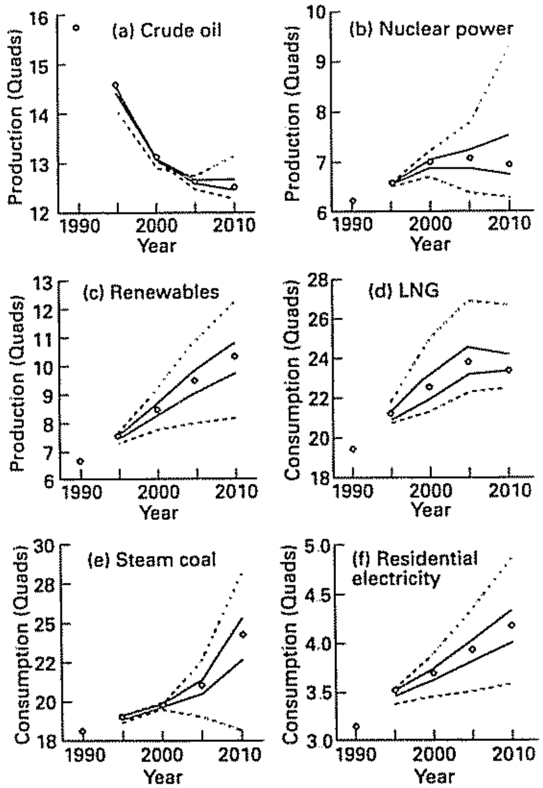

Fig. 5. Confidence intervals for six production and consumption sectors. In each, the diamonds and solid confidence limits are from the Annual Energy Outlook of the US Department of Energy (1992). Dashed confidence intervals are adjusted to reflect over confidence in historical projections. Source: Shlyakhter et al. (1994).

of physical constants (Henrion and Fischhoff, 1986). These intervals narrow over time, as scientific knowledge improves. However, current best guesses (and the ranges of those guesses) are typically outside the previous ranges of reasonable possibilities. Kammen, Shlyakhter, and their co-workers (Kammen and Marino, 1993, Shlyakhter and Kammen, 1992 and 1993; Shlyakhter et al., 1993, 1994) have demonstrated similar patterns in a reanalysis of these data, as well as in analyses of other estimates and forecasts, drawn from physics, climatology, demographics, and energy policy. Fig. 4 shows their evaluation of US government Annual Energy Outlook forecasts. In it, forecasts are characterized in terms of the absolute value of the difference between the forecast value and a subsequently determined 'exact value,' divided by the standard deviation of the forecasts. If estimates were drawn from a gaussian distribution, then one would expect these scores to fall along the leftmost curve in the figure. In three sets of particle data, these authors found a much higher rate of large deviations. Fig. 5 translates these results into revised confidence intervals (dashed lines) around best-guess forecasts (squares) for several major sets of data, by adjusting forecaster-provided forecasts (solid lines) in keeping with the over-confidence demonstrated in analyses like that in Fig. 4.

4.2. Events

As noted above, one practical challenge to the communication of forecasts, or even descriptions of current and historical situations, is change in the definition of terms. Because these issues can apply to analyses of current, historical, or future states, I will use the term 'analyst' rather than 'forecaster'. As mentioned, for theoretical, political,[5] or methodological reasons, the consumers of analyses may suddenly find themselves facing a new definition of 'money supply' 'unemployment,' 'gross domestic product,' 'health effect,' or 'manslaughter,' etc. In addition to the technical problems of ensuring that recipients understand what the experts are currently talking about, there is the conceptual problem of whether the experts know what they are talking about. Even if a definition is revised because of indisputable scientific advances in theory or measurement, the change is an admission that the preceding definition was flawed. If recipients of a previous analysis were unaware of that conceptual instability, then they would have placed undue confidence in it. They should also be asking about the stability of the current definition.

In attempting to assess the confidence that definitions deserve, it may be useful to distinguish between situations in which the analysts are working with the best definition they can manage and situations in which the analysts have chosen one out of a set of competing definitions. In the former situation, one needs an epistemological appraisal of the state of the science, a topic pursued in the following section. In the latter situation, if the full definition is given and one has precise informational needs, then one can assess the relevance of the definition that was used (the topic of the previous section). Without that precision, one must guess at how much confidence to place in analysts' implicit or explicit claims of being on the right topic.

One part of that assessment is an examination of the analysts' incentive structure. Namely, are they motivated to exaggerate the pertinence of their focal variable? Does the nature of their business require them to sell the same analysis to multiple users? Do they have a reasoned basis for defending their chosen definitions (beyond, say, analytical convenience or convention)? Are they afraid of challenging their clients to be more precise about their needs?

For example, there are many ways to define 'risk', such as that arising from a new technology. Some analysts compute expected changes in life expectancy, while others compute expected rate of premature deaths. In addition to the technical differences in the computation, there are ethical differences as well (Crouch and Wilson, 1981, and Fischhoff et al. 1984). The latter measure treats a death as a death, regardless of who experiences it, whereas the former measure places extra value on deaths among the young (because more days will be lost with each death). Unless these issues are explained, the recipient is confidently being led off target.

Another form of possibly unwitting over-confidence in event definitions occurs when the analyst is free to try alternative model specifications until some seemingly reasonable results arise. For example, regression techniques are often used to value particular features of goods, as a function of the preferences 'revealed' in market actions. Analysts often have considerable freedom to try different models, including different versions of the dependent variable, each of

[5] An example of a politically motivated change would be tightening the definition of 'actively seeking employment' in order to reduce the apparent unemployment rate

which has enough intuitive appeal to allow it to be justified, should it produce acceptable results (Bentkover et al., 1985, and Fischhoff and Furby, 1988). Under those circumstances, the recipient of the analysis does not know how many degrees of freedom have been 'used up,' prior to settling on the version that was delivered.

Finally, predicted events may be so complex that it is hard to know which parts of them to take seriously. After the accident at Three Mile Island, there was some debate over whether that event sequence had been included in the elaborate probabilistic analyses conducted in the preceding decade. Under those circumstances, it is unclear whether the analysts are predicting anything at all.

5. Impoverishment

Many of these issues arise, in part, because of the attempt to present (and use) forecasts in isolation. At times, that is a workable arrangement. At other times, though, one needs to understand the nature of the enterprise that produced the forecasts. Without such an epistemological background, even explicit statements of uncertainty may represent misplaced imprecision.

5.1. Probabilities

One fundamental question concerns the kind of evidence that is allowed. At one extreme, probabilities are restricted to estimates of the relative frequency of an event with recurring opportunities to occur. At the other extreme, probabilities can summarize evidence from whatever (potentially diverse) sources seem relevant. Many forecasts lie somewhere along this frequentist-subjectivist continuum. Even if restricted to historical data, they incorporate an element of judgment, at least in the selection and application of statistical models. There are (long-standing) arguments for these different positions. Recipients need to know which view prevailed. Knowing what kind of probability they are facing is essential for judging how inclusive and reliable the analytical process has been (Gärdenfors and Sahlin, 1989, and Shafer and Tversky, 1985).

Providing such a description is far from trivial, requiring a thicker description than even second-order probabilities. In order to meed such needs, Funtowicz and Ravetz (1990) have developed an alternative notation, describing what they call the 'pedigree' of numerical estimates. Table 3 shows the three components of pedigree: the quality of the theoretical model generating the estimate, the quality of the data used, and the degree of social acceptance for the procedure. Funtowicz and Ravetz are philosophers of science, who derived the scheme from normative considerations, regarding what people ought to know about expert-generated estimates. Although the scheme has been applied to a variety of analyses, its interpretation by lay recipients has yet to be explored.

5.2. Events

Analysts work with abstractions of real-world situations. They should not be held responsible for the intrusion of issues outside their domain – providing they have circumscribed the limits to analysis. Thus, the consumers of probabilistic risk analyses need to know if the failure rates are being predicted on the basis of models that neglect human sources of error (such as operator, maintenance, or designer oversight), or treat these sources of error with procedures derived from hard-surface engineering. The consumers of flood forecasts need to know whether an attempt has been made to consider the cumulative effects of changes in land use (e.g., paving, flood control). The consumers of economic forecasts need to know whether the possibilities of major disruptions (e.g., wars) have been considered, as well as whether the model distinguishes between good and bad expenditures (e.g., waste disposal).

Circumscribing a model in this way requires self-reflection and candor on the analyst's part. It may mean overcoming the cognitive tendency to view what is customarily modelled as being what

Table 3
The pedigree of numerical estimates

Score	Source of uncertainty		
	Theoretical (Quality of model)	Empirical (Quality of data)	Social (Degree of acceptance)
4	Established theory • many validation tests • causal mechanisms understood	Experimental data • statistically valid samples • controlled experiments	Total • all but cranks
3	Theoretical model • few validation tests • causal mechanisms hypothesized	Historical/field data • some direct measurements • uncontrolled experiments	High • all but rebels
2	Computational model • engineering approximations • causal mechanisms approximated	Calculated data • indirect measurements • handbook estimates	Medium • competing schools
1	Statistical processing • simple correlations • no causal mechanisms	Educated guesses • very indirect approximations • 'rule of thumb' estimates	Low • embryonic field
0	Definitions/assertions	Pure guesses	None

Source: Funtowicz and Ravetz (1990).

really matters, as well as the motivational tendency to deny the poverty of one's reach. However, it is only by knowing these facts that consumers can create their own assessment of pedigree, anticipate the direction of systematic biases, predict the rate of forecasting progress, judge the meaning of agreement among forecasters, and know where else to go for help. At this level, we know very little about what people take away from forecasts. Wagenaar et al. (1985) and others (Berry, 1986, Hyatt et al., 1978, and cited in Ayton, 1988) found that people remember rather little of broadcast weather forecasts, even when unusual steps are taken to ensure comprehensibility. Perhaps people's attention and commitment wander when they have deep questions regarding the nature of forecasts, but do not know where to begin asking questions or getting answers.

6. (Where) did the Weather Channel go wrong?

To be useful, a conceptual scheme should at least be able to shed some light on past experience. Table 4 applies one such scheme heuristically to the experience of the Weather Channel with the storm of March 1993, although

Table 4
Where did the Weather Channel go wrong?

Threats to communicating forecasts	Description of	
	Probabilities	Events
Ambiguity	Certain to happen	Clear on weather
Irrelevance	Certain	Just weather
Immodesty	You can really trust us this time	Ignored resolution (time, space)
Impoverishment	Gave sources of confidence ignored historical record	Ignored ties of weather to life

a more detailed review of the tapes and assessment of viewers' interpretations would be needed to confirm these speculations.

Ambiguity does not seem to have been the problem. Both the probability and the event were clearly stated: a very large storm, with unusual winds and precipitation, was certain to occur over a large (and specified) area of the eastern United States. The probability side of the forecasts holds up fairly well on the remaining criteria. The Channel said explicitly how much confidence could be placed in them (total), as well as giving an indication of whether those assertions could be trusted. There were even some efforts to explain the sources of their confidence, by showing the quality of their observational data, the strength of the underlying theory, and the convergence of unstoppable weather phenomena. If this aspect of the forecast is to be faulted, it might be for failing to reconcile the proclaimed trustworthiness of this forecast with failures in the past. Thus, the Weather Channel's ability to make the case for this forecast may have been limited by the absence of a context for understanding the successes and failures of previous forecasts. Even if the Weather Channel changed its practices, it would still have to face the confusion created by the less enlightened practices of other weather forecasters.

It is on the event side of the ledger than things seem to unravel. The Channel predicted aggregate weather, with considerable accuracy. However, the fate of individuals is determined by much finer phenomena, such as the weight of snow on electric lines, the traction on icy roads, and the melt rate on the sidewalks that they might be tempted to shovel. Thus, the Weather Channel was not predicting the event that really mattered to recipients. Of course, it would be impossible for a general service forecaster to make (or deliver) driveway-by-driveway predictions. Indeed, there is money to be made customizing weather forecasts for specific clients. Nonetheless, ordinary recipients may need extra help to understand the limits to a general forecast's resolution.

As a matter of enrichment, one could explain why the details matter, thereby creating the potential links between weather and life. These connections would, in effect, be classes of mini-forecasts, of much greater specificity and much less confidence than the synoptic forecasts. For example, one might say, 'Such storms can produce uncommonly heavy snow. Shovelling such snow can, in turn, produce cardiac arrest among individuals with previously undiagnosed heart problems.' or 'Such storms typically produce scattered collapses of transportation systems, leaving people stranded for extended periods of time, possibly without proper protection against exposure.' The deliberate tentativeness of such mini-forecasts may help to convince recipients of the uncertainties that they face. The critical general messages may be that 'storms like this are so far outside everyday experience that intuitions cannot be trusted,' and 'this means you.'

For such messages to work, recipients need an understanding not just of the weather, but of its role in their lives. For example, recipients might just not be able to imagine how one's heart could fail from a little shovelling. Improving people's mental model of their own physiology has increased adherence to medical regimes (Leventhal and Cameron, 1987). Perhaps the Weather Channel should team up with the Medicine Channel, in order to fulfil their mutual commitments to listeners.

Conclusion

An error matrix

Ideally, all forecasts would be entirely transparent, so that they mean (to the forecaster) what they seem to mean (to the recipient). When problems remain, the critical question is whether they are recognized or ignored. Table 5 suggests the consequences of pitfalls being acknowledged or ignored by the forecaster, as a function of whether the recipient is aware or ignorant of the possibility.

If both are aware of a problem (Cell A), then they can form a partnership, trying to improve

Table 5
Risks of (mis)communication

Client	Forecaster	
	Addresses pitfall	Ignores pitfall
Recognizes pitfall	*A* Partnership; mutual frustration with reality (for making forecasting hard)	*B* Lost business; frustration with forecasters (for avoiding obvious problems)
Ignorant of pitfall	*C* Education of client; frustration with forecasts (for failing to meet needs)	*D* Folie a deux frustration with future (for which one is unprepared)

communication and sharing frustration with a reality that poses complexity, ambiguity, and uncertainty. If forecasters are aware of problems that escape the attention of clients (Cell C), then there is, in principle, an opportunity for education – assuming that clients are willing to deal with those problems. If clients are unwilling, then they may find the forecasts quite frustrating, full of qualifications and explanations without clear justification. The resulting confusion might, however, at least keep them from relying too heavily on the forecasts, which they realize that they do not really understand.

If forecasters fail to deal with problems with potential clients recognize (Cell B), then they may lose respect and even business for failing to level with those clients. The frequency of such client sophistication might determine how willing forecasters are to take such gambles.[6] If no one recognizes the potential pitfalls (Cell D), then there is joint folly (or what psychoanalysts sometimes call a folie a deux). When both forecaster and client exaggerate the quality of forecasts, the client will often win the race to the poorhouse, discovering empirically the price of undue faith in forecasts.

[6] The legendary, and unacknowledged, ambiguity and immodesty of the annual forecasts that appear in supermarket tabloids suggests a perception of clients who are either very unsophisticated or very forgiving

A cautionary anecdote

Some years ago, I was invited to a meeting of academics and senior US military officers, primarily from the Navy. I was eager for the opportunity to try out a conceptualization of command and control that I had been developing with some colleagues (Fischhoff and Johnson, 1990, Lanir et al., 1988). Briefly, it asked how the strengths and weaknesses of individual decision making are enhanced or complicated by being embedded in organizations of increasing complexity.

The following speaker, a rear admiral, paid me the great honor of throwing away his prepared talk to see what he could make of our theory. I was immediately eager to see what he made of our attempt to integrate individual and organizational perspectives. He began, 'Professor Fischhoff has said that decisions have three components: options, outcomes, and uncertainties. Let me see if that fits any of the decisions that I have made.' At one level, he proceeded to show the opaqueness of some terms that are at the center of our enterprise. At another level, he showed the magnitude of the potential gaps between the conceptualizations of practitioners and those who have adopted the analytical turn of mind. Although he was able to make some sense out of our basic conceptual scheme, the particular elaboration that I had presented required an unrealistic degree of fluency.

A sketchy workplan

Unless we have the chance to interact directly with the recipients of our messages, it is hard to know what they make of what we say. If this seems like a worthy topic for systematic study, then we are fortunate in having some seemingly useful procedures for beginning the work. Some could be applied almost immediately by practitioners. For example, there is no good reason (other than the effort and resources demanded) why performance records should not be created and publicized, thereby calibrating the confidence placed in forecasts, or why think-aloud protocols should not be used to anticipate recipients' interpretations of forecasts. A modest investment in transfer of technology should be able to make these research techniques available for many applications. If not, then there is something to be learned about the technique by figuring out how to make this step.

In other cases, though, the work has hardly begun. As one goes down and to the right in Tables like 4, the communication task involves communicating the forecaster's worldview to the consumer. That requires both empirical work regarding lay epistemology and theoretical work regarding the nature of forecasting. It would not be the first time that taking teaching seriously helped to clarify one's thinking.

Acknowledgements

My thanks to the organizers of the Thirteenth Annual International Symposium on Forecasting for the invitation to present an earlier version of this paper as a plenary address. Daniel Kammen, Spyros Makridakis and Alexander Shlyakhter provided valuable editorial and substantive comments. The research reported here was sponsored in part by the National Science Foundation and the National Institute of Alcohol Abuse and Alcoholism. The comments expressed are my own. Please address correspondence to Baruch Fischhoff, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.

References

Ayton, P., 1992, On the competence and incompetence of experts, in: G Wright and F. Bolger, eds., *Expertise and Decision Support* (Plenum, New York), 77–105.

Ayton, P., 1988 Perceptions of broadcast weather forecasts, *Weather*, 43, 193–197.

Bailar, B A. and N D. Rothwell, 1984, Measuring employment and unemployment. in: C F Turner and E., Martin, eds., *Surveying Subjective Phenomena* (Russell Sage Foundation, New York) 129–142.

Bentkover, J D , V T Covello and J. Mumpower, eds., 1985, *Benefits Assessment: The State of the Art* (D. Reidel, Dordrecht)

Berry, C., 1986 What's wrong with the weather? *European Broadcasting Union Review*, 37, 34–37

Beyth-Marom, R., 1982, How probable is probable? Numerical translation of verbal probability expressions, *Journal of Forecasting*, 1, 257–269.

Cohen, J., E.S. Dearnley and C.E.M. Hansel, 1958, A quantitative study of meaning, *British Journal of Educational Psychology*, 28, 141–148.

Costanza, R., ed., 1991, *Ecological Economics* (Columbia University Press. New York).

Crouch, E A C. and R Wilson, 1981, *Risk/Benefit Analysis* (Ballinger, Cambridge, MA).

Dowlatabadi, H and G. Morgan, 1993, A model framework for integrated studies of climate change, *Energy Policy*, 21, 209–221

Evans, J.S , N.C Hawkins and J.D Graham, 1988, The value of monitoring for radon in the home, *Journal of the Air Pollution Control Association*, 38, 138–145

Fischhoff, B . 1982, Debiasing, in: D. Kahneman, P. Slovic and A Tversky, eds , *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, New York) 422–444

Fischhoff, B., 1989, Making decisions about AIDS, in: V. Mays, G Albee and S Schneider, eds , *Primary Prevention of AIDS* (Sage, Newbury Park, CA). 168–205.

Fischhoff, B and L Furby. 1988. Measuring values: A conceptual framework for interpreting transactions, *Journal of Risk and Uncertainty*, 1, 147–184.

Fischhoff, B and S Johnson. 1990. The possibility of distributed decision making: Appendix to Distributed Decision Making: Workshop Report (National Academy Press, Washington. DC). pp 25–58

Fischhoff, B and D. MacGregor, 1986, Calibrating databases, *Journal of American Society for Information Sciences*, 37, 222–233.

Fischhoff, B., D MacGregor and L. Blackshaw, 1987, Creating categories for databases, *International Journal of Man-Machine Systems*, 27, 33–63

Fischhoff, B., P. Slovic and S. Lichtenstein. 1978, Fault trees: Sensitivity of assessed failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 330–344

Fischhoff, B., S. Watson and C. Hope, 1984, Defining risk, *Policy Sciences*, 17, 123–139.

Fortney, J., 1988, Contraception: A life long perspective, in: *Dying for love*, (National Council for International Health, Washington, DC).

Funtowicz, S.O. and J.R. Ravetz, 1990, *Uncertainty and Quality in Science for Policy* (Kluwer, Boston).

Furnas, G.W., T.K. Landauer, L.M. Gomez, and S.T. Dumais, 1983, Statistical semantics: Analysis of the potential performance of key-word information systems, *Bell System Technical Journal*, 62, 1753–1806.

Gärdenfors, P. and N.E. Sahlin, 1982, Unreliable probabilities, risk taking and decision making, *Synthese*, 53, 361–386.

Gärdenfors, P. and N.E. Sahlin, eds., 1989, *Decision, probability, and utility*, (Cambridge University Press, Cambridge).

Grice, H.P., 1975, Logic and conversation, in: P. Cole and J.L. Morgan, eds., *Syntax and Semantics, Vol. 3, Speech acts*, (Academic Press, New York), 41–48.

Hamm, R.M., 1991, Selection of verbal probabilities, *Organizational Behavior and Human Decision Processes*, 48, 193, 223.

Henrion, M. and B. Fischhoff, 1986, Assessing uncertainty in physical constants, *American Journal of Physics*, 54, 791–798.

Holyoak, K.J. and A.L. Glass, 1978, Recognition confusion among quantifiers, *Journal of Verbal Learning and Verbal Behavior*, 17, 249–264.

Hyatt, D., K. Riley and N. Sederstrom, 1978, Recall of television weather reports, *Journalism Quarterly*, 55, 306–310.

Kammen, D.M. and B.D. Marino, 1993, On the origin and magnitude of pre-industrial anthropogenic CO_2 and CH_4 emissions, *Chemosphere*, 26, 69–85.

Keren, G., 1987, Facing uncertainty in the game of bridge: A calibration study, *Organizational Behavior and Human Decision Processes*, 39, 98–114.

Krysztofowicz, R., 1983, Why should a forecaster and a decision maker use Bayes' Theorem, *Water Resources Research*, 19, 327–336.

Lanir, Z., B. Fischhoff and S. Johnson, 1988, Military risk taking: C^3I and the cognitive functions of boldness in war, *Journal of Strategic Studies*, 11, 96–114.

Leventhal, H. and L. Cameron, 1987, Behavioral theories and the problem of compliance, *Patient Education and Counseling*, 10, 117–138.

Lichtenstein, S. and B. Fischhoff, 1977, Do those who know more also know more about how much they know? The calibration of probability judgments, *Organizational Behavior and Human Performance*, 20, 159–183.

Lichtenstein, S., B. Fischhoff and L.D. Phillips, 1982, Calibration of probabilities: State of the art to 1980, in: D. Kahneman, P. Slovic and A. Tversky, eds., *Judgment under uncertainty: Heuristics and biases* (Cambridge University Press, New York), 306–334.

Lichtenstein, S. and J.R. Newman, 1967, Empirical scaling of common verbal phrases associated with numerical probabilities, *Psychonomic Science*, 9, 563–564.

Linville, P.W., G.W. Fischer and B. Fischhoff, 1993, Perceived risk and decision making involving AIDS, in: J.B. Pryor and G.D. Reeder, eds., *The social psychology of HIV infection*, (Erlbaum, Hillsdale, NJ), 5–38.

Merz, J.F., 1991, Toward a standard of disclosure for medical informed consent: Development and demonstration of a decision – analytic methodology. Ph.D. dissertation, Carnegie Mellon University.

Merz, J., M. Druzdzel and D.J. Mazur, 1991, Verbal expressions of probability in informed consent litigation, *Medical Decision Making*, 11, 273–281.

Merz, J., B. Fischhoff, D.J. Mazur and P.S. Fischbeck, 1993, Decision-analytic approach to developing standards of disclosure for medical informed consent, *Journal of Toxics and Liability*, 15, 191–215.

Morgan, M.G. and M. Henrion, 1990, *Uncertainty*, (Cambridge University Press, New York).

Murphy, A.H., S. Lichtenstein, B. Fischhoff and R.L. Winkler, 1980, Misinterpretations of precipitation probability forecasts, *Bulletin of the American Meteorological Society*, 61, 695–701.

Murphy, A. and B. Brown, 1984, Comparable evaluation of subjective weather forecasts in the United States, *Journal of Forecasting*, 3, 369–393.

Murphy, A. and R. Winkler, 1984, Probability of precipitation forecasts, *Journal of the American Statistical Association*, 79, 391–400.

Murphy, G.L. and D.L. Medin, 1985, The role of theories in conceptual coherence, *Psychological Review*, 92, 289–316.

National Health Information Survey, 1987, Knowledge and attitudes about AIDS: Data. National Center for Health Statistics. Advance Data, No. 146.

National Research Council, 1983, *Priorities for chemical toxicity testing*, (National Research Council, Washington, DC).

Office of Technology Assessment, 1993, *Preparing for an uncertain climate (2 vols.)*, (Office of Technology Assessment, Washington, DC).

Poulton, E.C., 1989, *Bias in quantifying judgment*, (Lawrence Erlbaum, Hillsdale, NJ).

Quadrel, M.J., 1990, Elicitation of adolescents' risk perceptions: Qualitative and quantitative dimensions. Ph.D. dissertation, Carnegie Mellon University.

Quadrel, M.J., B. Fischhoff and C. Palmgren, 1994, Adolescents' definitions of risk behaviors, Unpublished work.

Rubin, E.S., L.B. Lave and M.G. Morgan, 1992, Keeping climate change relevant, *Issues in Science and Technology*, (Winter). 47–55.

Shafer, G., 1975, *A mathematical theory of evidence*, (Princeton University Press, Princeton, NJ).

Shafer, G. and A. Tversky, 1985, Languages and designs for probability judgment, *Cognitive Science*, 9, 309–339.

Shlyakhter, A.I. and D.M. Kammen, 1992, Sea-level rise or fall? *Nature*, 357, 25.

Shlyakhter, A.I. and D.M. Kammen, 1993, Uncertainties in modeling low probability/high consequence events: Application to population projections and models of sea-level rise, *Second International Symposium on Uncertainty Modeling and Analysis*, (IEEE Computer Society Press, Washington, DC), 246–253.

Shlyakhter, A.I., D.M. Kammen, C.L. Broido and R. Wilson, 1994, Quantifying the credibility of energy projections from trends in past data: The US energy sector, *Energy Policy*, 22, 119–131.

Shlyakhter, A.I., I.A. Shlyakhter, C.L. Broido and R. Wilson, 1993, Estimating uncertainty in physical measurements. Observational and environmental studies: Lessons from trends in nuclear data, *Second International Symposium on Uncertainty Modeling and Analysis*, (IEEE Computer Society Press, Washington, DC), 310–317.

Sniezek, J.A. and T. Buckley, 1991, Confidence depends on level of aggregation, *Journal of Behavioral Decision Making*, 4, 263–272.

Sniezek, J.A., P.W. Paese and F.S. Switzer, 1990, The effect of choosing on confidence in choice, *Organizational Behavior and Human Decision Processes*, 46, 264–282.

Superstorm, 1993, (1993), Atlanta, GA: The Weather Channel.

Svenson, O. and B. Fischhoff, 1985, Levels of environmental decisions, *Journal of Environmental Psychology*, 5, 55–68.

Turner, C., J.T. Lesser, J.T. and J.C. Gfroerer, 1992, *Survey measurement of drug use*, (National Institute of Drug Abuse: Research Triangle Park, NC).

Turner, C.F. and E. Martin, E., eds., 1984, *Surveying subjective phenomena*, (Russell Sage Foundation, New York).

US Department of Energy, 1992, *Annual Energy Outlook, with Projections to 2010*, (DOE/EIA-0383(92)). Washington, DC, The Department.

Wagenaar, W.A., R. Schreuder and A.H.C. van der Heijden, 1985, Do TV pictures help people to remember the weather? *Ergonomics*, 28, 765–772.

Wallsten, T., D.V. Budescu, A. Rapoport, R. Zwick and B. Forsyth, 1986, Measuring the vague meanings of probability terms, *Journal of Experimental Psychology: General*, 115, 348–365.

von Winterfeldt, D. and W. Edwards, 1986, *Decision analysis and behavioral research*, (Cambridge University Press, New York).

Yates, J.F., 1989, *Judgment and decision making*, (Wiley, Chichester).

Zimmer, A.C., 1984, A model for the interpretation of verbal predictions, *International Journal of Man-Machine Systems*, 20, 121–134.

Biographies: Baruch Fischhoff is Professor of Social and Decision Sciences and of Engineering and Public Policy at Carnegie Mellon University. He holds a B.S. in mathematics from Wayne State University and a MA and Ph.D. in psychology from the Hebrew University of Jerusalem. He is recipient of the American Psychological Association's Early Career Awards for Distinguished Scientific Contribution to Psychology (1980) and for Contributions to Psychology in the Public Interest (1991). He is a Fellow of the Society for Risk Analysis, as well as recipient of its Distinguished Achievement Award (1991). He is a member of the Institute of Medicine of the National Academy of Sciences. His current research includes risk communication, adolescent decision making, evaluation of environmental damages, and insurance-related behavior. He serves on the editorial boards of several journals, including *Journal of Risk and Uncertainty, Accident Analysis, and Prevention, Cognitive Psychology,* and the *International Journal of Forecasting*.