Lay Foibles and Expert Fables in Judgments About Risk

BARUCH FISCHHOFF, PAUL SLOVIC, and SARAH LICHTENSTEIN*

Public perceptions of risk are a focal point of many debates about the management of hazardous technologies. Different views about what the public knows and wants often lead to quite different beliefs about what policies should be adopted and even about how society's policy-making processes should be structured. Often these views about the public are based on speculation or anecdotal observation. In the interests of having better informed debates, the present paper reviews existing empirical evidence about public risk perceptions. In doing so, it reaches a number of interim conclusions and draws forth their implications for the respective roles of technical experts and lay people in technology management.

KEY WORDS: Risk assessment; Risk perception; Subjective probability; Judgment; Technology management.

INTRODUCTION

Discussions about managing technological hazards are often heavily speckled with assertions about what the public wants, knows, and is capable of understanding. The present paper reviews the empirical evidence about lay people's attitudes and behavior regarding risk. A look at the evidence seems useful because these discussants and commentators often act as though what they say goes regarding descriptions of behavior. Chemists, climatologists, physicists and others, who might tread very gently beyond the available data in their own areas of specialization, at times seem to feel no restraint in opining "This is what lay people think about the risks of nuclear power," "This is the sort of information the public needs to put risks into proper perspective," "This is how the public will react to stricter safety measures."

Like speculations about chemical reactions, speculations about human behavior need to be disciplined by fact. Since they make important statements about people and their capabilities, failure to validate such speculation may mean arrogating to oneself considerable political power. Such happens, for example, when one says that people are so poorly informed (and uneducable) that they require paternalistic institutions to defend them and, furthermore, that they might be better off surrendering some political rights to technical experts. It also happens, at the other extreme, when one claims that people are so well informed (and offered such freedom of choice) that one needn't ask them anything at all about their desires; to know what they want, one need only observe their behavior in the market place. It also happens when we assume that people are consummate hedonists, rational to the extreme in their consumer behaviour but totally uncomprehending of broader economic issues, so that we can impose effective fiscal policies on them without being second-guessed.

One reason for the survival of such simplistic and contradictory positions is political convenience. Some people want the lay public to participate actively in hazard management decisions, and need to be able to describe the public as competent; others need an incompetent public to legitimate an expert elite. A second reason is theoretical convenience. It is hard to build models of people who are sometimes wise and sometimes foolish, sometimes risk seeking and sometimes risk averse. A third reason is that one can so effortlessly speculate about human nature and even produce a bit of supporting anecdotal information. Indeed, good social theory may be so rare because poor social theory is so easy.

These reasons may also contribute to the fact that the record of systematic and empirical evidence is not as complete as one would like. However, enough seems known to make five nontrivial assertions.

ASSERTION 1. EXAMINING BEHAVIOR SYSTEMATICALLY YIELDS SOME SURPRISES

Social scientists often find themselves in a no-win situation. If they describe their work in technical jargon, no one wants to listen. If they use plain language, no one feels a need to listen. Listeners feel that they "knew it all along" and that the social scientist was just "affirming the obvious" or "validating common sense." One possible antidote to this feeling is to point out the evidence showing that in hindsight, people exaggerate how much they could have known in foresight, leading them to discount the informativeness of scientific reports (Slovic and Fischhoff 1977). A second is to note that common sense often makes contradictory predictions (e.g., two heads are better than one vs. too many cooks spoil the broth; absence makes the heart

*Baruch Fischhoff, Paul Slovic, and Sarah Lichtenstein are Research Associates at Decision Research, a branch of Perceptronics, 1201 Oak, Eugene, OR 97401. This research was supported by the National Science Foundation under Grant PRA 79-11934 to Clark University under subcontract to Perceptronics, Inc. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. This paper is an abridgement of one that appeared under the same title in T. O'Riordan and R.K. Turner (eds.), *Progress in Resource Management and Environment Planning, Vol. 3,* Chicester: John Wiley, 1981. Our thanks to Ian Nimmo-Smith, Timothy O'Riordan, and Miron Straf for comments on previous drafts.

grow fonder vs. out of sight, out of mind). Research is needed to determine which version is correct or what their respective ranges of validity are. A third strategy, adopted immediately below, is to present empirical results that contradict conventional wisdom.

Informing people about risks

It is often claimed that people don't want to know very much about the health risks they face, since such information makes them anxious. Moreover, they cannot use that information very productively, even if it is given. If true, these claims would legitimate having someone else (e.g., physicians, manufacturers, government) do the deciding about what health (and therapeutic) risks are acceptable, and not invest too much effort on information programs. Recently, however, a number of investigators have replaced anecdotal evidence with systematic observation and found that, by and large, people want to be told about potential risks (Alfidi 1971; Weinstein 1979). In clinical settings, this desire has been observed with such risky treatments as psychotropic medication (Schwarz 1978), endoscopy (Roling et al. 1977), and oral contraceptives (*Applied Management Sciences* 1978; Joubert and Lasagna 1975). For example, when casual laborers at a state employment office were shown a pamphlet explaining the risks faced by temporary workers in a nuclear power plant, 90 percent gave the most affirmative answer possible to the question "If you had taken such a job without being shown this pamphlet, would you feel that you had been deprived of necessary information?" (Fischhoff, in press).

Risk-Taking Propensity

We all know that some people are risk takers and others are risk avoiders; some are cautious, whereas others are rash. Indeed, attitude toward risk might be one of the first attributes that comes to mind when one is asked to describe someone's personality. In 1962, Slovic compared the scores of 82 individuals on 9 different measures of risk taking. He found no consistency at all in people's propensity for taking risks in the settings created by the various tests. Correlations ranged from $-.35$ to $.34$, with a mean of $+.006$. That is, people who are daring in one context may be timid in another, a result that has been replicated in numerous other studies (e.g., Davidshofer 1976).

The surprisingness of these results may tell us something about ourselves as well as about the people we observe. One of the most robust psychological discoveries of the last 10 years has been identification of the "fundamental attribution error," the tendency to view ourselves as highly sensitive to the demands of varying situations, but to see others as driven to consistent behaviour by dominating personality traits (Nisbett and Ross 1980). This misperception may be due to the fact that we typically see most others in only one role, as workers or spouses or parents or bowlers or drivers or whatever, in which the situational pressures are quite consistent. Thus, we may observe accurately the evidence available to us, but fail to understand the universe from which these data are drawn.

Protective Behavior

For years, the United States has been building flood control projects. Despite these great expenditures, flood losses today (in constant dollars) are greater than they were before this enterprise began. Apparently, the behavioral models of the dam and levee builders failed to account for the extent to which eliminating the recurrence of small-to-moderate floods reduced residents' (and particularly newcomers') sensitivity to flood dangers, which in turn led to overbuilding of the flood plain. As a result, when the big (100-year) floods come, exceeding the containment capacity of the protective structures, much more is in their path (White 1974).

The official response to this situation was the National Flood Insurance Program (Kunreuther et al. 1978), designed according to economic models of human behaviour, which assumed that flood plain residents are all-knowing, all-caring, and entirely "rational" (as defined by economics). Initially, premiums were greatly subsidized by the federal government in order to make the insurance highly attractive; these subsidies were to be gradually withdrawn once the insurance-buying habit was established. Unfortunately for the program, few people bought the insurance. The typical explanation for this failure was that residents expected the government to bail them out in the event of flood. However, a field survey found this speculation, too, to be in error. Flood plain residents reported expecting no help, feeling that they were willingly bearing an acceptable risk. When residents thought about insurance at all, they seemed to rely on a melange of ad hoc principles like, "I can't worry about everything" and "The chances of getting a return (reimbursement) on my investment (premium) are too small," rather than on the concepts and procedures of economics (Kunreuther et al. 1978; Slovic et al. 1977.).

Setting Acceptable Risk (Double) Standards

A prominent approach to determining socially acceptable levels of risk are the "revealed preference" analyses advanced by Starr (1972). Such analyses begin by calculating the current risks and benefits to society of selected technologies. According to Starr's interpretation, the illustrative data in Figure 1 reflect society's success in achieving a (nearly) optimal balance between the risks and benefits of different technologies. In addition to allowing more beneficial activities to be more risky, society has also imposed less stringent standards upon voluntarily-incurred risks.

Since people (and, through their combined efforts, society) respond to the risks they perceive, which are not necessarily those calculated by a particular scientist,

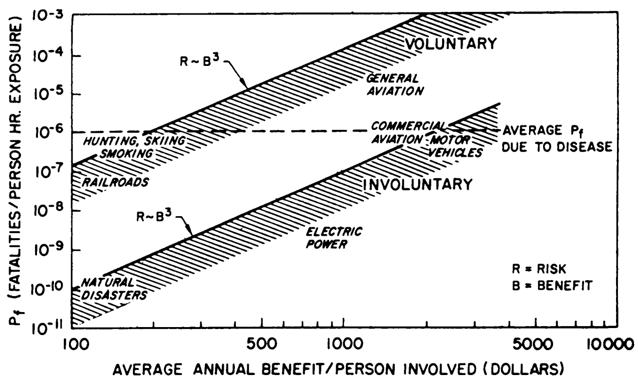

Figure 1. Current risk and benefit levels for selected technologies, as computed by Starr (1972, p. 33). Roughed-in lines were interpreted as reflecting society's revealed preference for tolerating more risk for more beneficial activities and for lower risk levels for involuntarily encountered risks (holding benefit constant).

it is worth asking whether people agree with Starr about the fine-tuning of our world. Figure 2 shows a set of subjective estimates of the risks and benefits of Starr's technologies. In this judgmental space, the relationship found in Starr's computational space (Figure 1) disappears; equally beneficial items may be high or low in risk. Similar results have been obtained using a variety of technologies, respondents, and response modes (Fischhoff et al. 1978; Slovic, Fischhoff, and Lichtenstein 1980). The typical correlation between perceived risk and perceived benefit was about −.20. Consideration of voluntariness made little difference.

Although respondents did not believe that society had effected Starr's hypothesized risk-benefit tradeoff, they indicated that they would like such a relationship to exist, as shown by Figure 3, which contrasts judgments of *acceptable* risk with those for perceived benefit. Moreover, respondents felt that voluntary and involuntary risks should be treated differently. Complicating the picture was the fact that they also wanted double standards for known and unknown risks, for risks with immediate and delayed consequences, and for those differing on other qualitative features. More recent evidence suggests that since involuntary risks affect large numbers of people and often impose risks on individuals other than those who gain the benefits, voluntariness may, in fact, be a surrogate for people's concern about catastrophic potential or equity (Slovic et al. 1980).

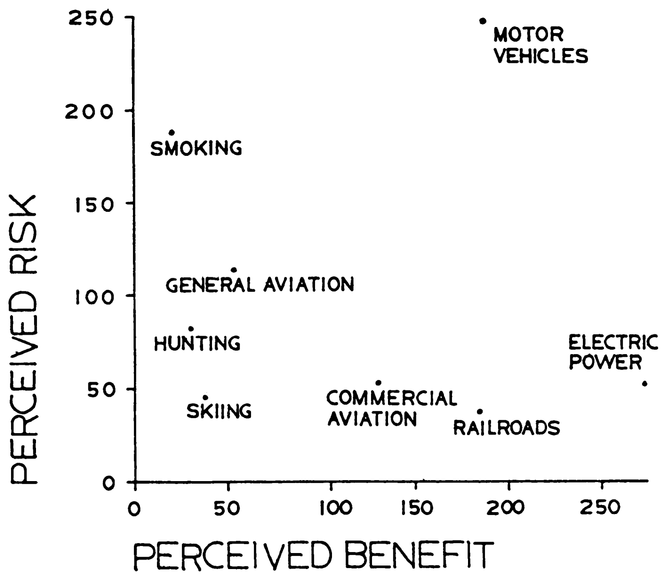

Figure 2. Judgments of risk and benefit for activities and technologies used by Starr (in Fig. 1). Respondents were 41 members of the League of Women Voters. Similar patterns were revealed with other items, response modes and respondent groups. Source: *Fischhoff et al. (1978, p. 136).*

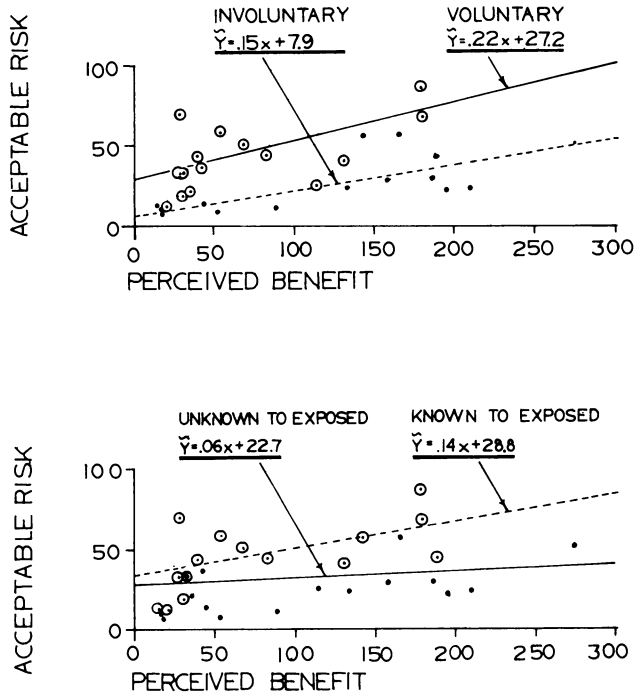

Figure 3. Judgment of acceptable risk levels and perceived current benefit. In the top figure, technologies were sorted into those judged most voluntary (circled) and those judged least voluntary. Regression lines for the two subsets of items were interpreted as reflecting a double standard, with more safety being required of involuntary risks. The flower figure presents an analogous analysis for the activities whose risks are most and least well understood by those exposed to them. Source: *Fischhoff et al. (1978, p. 138).*

Conclusion

The common theme of these examples is that nothing goes without saying regarding human behaviour. Casual observations cannot be presumed to be valid. Indeed, a growing body of research on intuitive theories of behavior (Fischhoff 1980a; Nisbett and Ross 1980) suggests that such theories are often poorly formulated, based on scanty or poorly sampled evidence and insensitive to inconsistent data. Having decided to look at empirical evidence, one must then resist the tendency to discount its conclusions as "obvious." Perhaps the most general antidote to pooh-poohing reported results is to ask oneself "Had the opposite been reported, could I have explained it just as readily?" (Slovic and Fischhoff 1977).

ASSERTION 2: STUDYING BEHAVIOR IS DIFFICULT

Judgments of Risk

At first blush, assessing the public's risk perceptions would seem to be very straightforward. Just ask questions like, "What is the probability of a nuclear core meltdown?" or "How many people die annually from asbestos-related diseases?" or "How does wearing a seat belt affect your probability of living through the year?" Once the results are in, they can be compared with the best available technical estimates, with deviations interpreted as evidence of the respondents' ignorance.

Unfortunately, how one asks the question may in large part determine the content (and apparent wisdom) of the response. Lichtenstein and her colleagues (1978) asked two groups of people to estimate the frequency of death in the United States from each of 40 different causes. The groups differed only in the information given them about one cause in order to help scale their responses; one being told that about 50,000 people die annually in motor vehicle accidents, the other being told about the 1,000 annual deaths from electrocution. Although both reports were accurate, provision of a larger number increased respondents' estimates of most frequencies. Such anchoring on the original number changed the smallest estimates by roughly a factor of 5.

Fischhoff and MacGregor (1980) asked people to judge the lethality of various potential causes of death using one of four formally equivalent formats: (a) For each afflicted person who dies, how many survive? (b) For each 100,000 people afflicted, how many will die? (c) x people were afflicted with this malady, how many people died? (d) y people died from this malady, how many were afflicted but did not die? When responses were translated into a common format, they revealed even more dramatic effects of question phrasing on expressed risk perceptions. For example, when people estimated the lethality rate for influenza directly, their mean response was 393 deaths per 100,000 cases. When told that 80,000,000 people catch influenza in a normal year and asked to estimate the number who die, respondents' mean response was 4,800, representing a death rate of only 6 per 100,000 cases. This slight change in the question changed the estimated rate by a factor of more than 60. Similar discrepancies occurred with other questions and other hazards.

Another study (Fischhoff 1980b) asked respondents to estimate the risks of an unnamed drug (actually, an oral contraceptive) as these were described in two package inserts distributed by the manufacturer, one designed for doctors and one for patients. Readers of the patients' form thought that the risk of death from blood clots (the major risk described) was 5.1 times as large for users as for nonusers; readers of the doctors' form thought that it was "only" 2.5 times as large. On the other hand, readers of the patients' form estimated a much lower overall death rate (1 in 40,000 vs. 1 in 2,000 with the doctors' form). Thus, the risk seemed greater in the doctors' form by one measure, less by another, almost identical measure. The reason for this discrepancy seems to be that the patients' version gave a number of representative death and morbidity rates, revealing that the absolute value of a risk that seemed relatively high was an order of magnitude smaller than that imagined by readers of the doctors' form. Had only one risk question been asked, one would have had a rather different picture of readers' knowledge and the effect of the textual differences between the inserts.

Such effects are hardly new; indeed, some have been recognized for close to 100 years. Early psychologists discovered that different numerical judgments may be attached to the same physical stimulus (e.g., the loudness of a tone) as a function of whether it is presented in the context of increasingly intense or weak alternatives, whether the set of alternatives is homogeneous or diverse, and whether the respondent makes one or many judgments. Even when the same presentation is used, different judgments might be obtained with a numerical or a comparative (ordinal) response mode, with instructions stressing speed or accuracy, with a bounded or an unbounded response set, and with verbal or numerical response labels.

The range of these effects may suggest that the study of judgment is not just difficult, but actually impossible. Closer inspection, however, reveals considerable orderliness underlying this apparent chaos. Poulton (1968) discovered enough order to offer six "laws" of the "new psychophysics." Presented in Figure 4, these show how the judgmental value (Ψ) assigned to a physical stimulus (ϕ) will vary depending upon how it is elicited.

Judgments are sensitive to these factors because the formulation of a response always involves an inferential process. The respondent must decide what the questioner (or the task) means, how the physical stimulus impinges on the relevant senses, and how to translate that sensation into an acceptable response. When the task is novel, one necessarily turns to its details and its purveyors for hints as to what to say. Presentation and elicitation effects are inevitable because one has to pose a problem some way, and the method used will affect the resultant message. Such effects can, in principle, be turned around and used to manipulate the information that goes into people as well as that which comes out of them. By judicious formatting, one may be able to exaggerate or minimize perceived risks in ways that could never be faulted in a court of law. Indeed, in this work, psychologists may be laboriously discovering effects known well and long to merchandisers.

Judgments of Values

Once the facts of an issue have been estimated and communicated, it is usually held that lay people should (in a democracy) be asked about their values. What do they want—after the experts have told them what they can (conceivably) have? Here, too, the straightforward

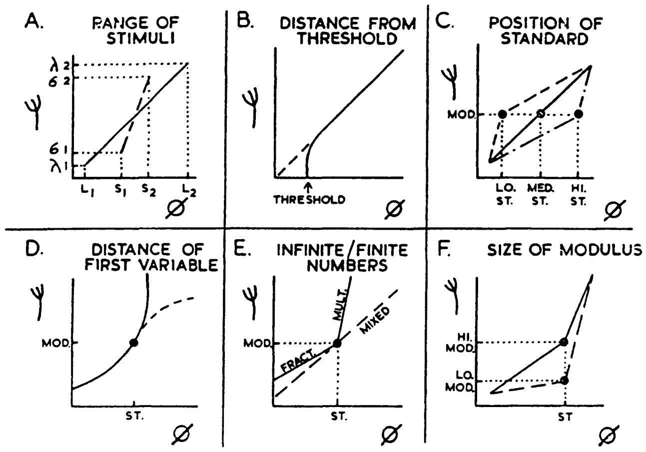

Figure 4. Six laws of the new psychophysics. In each figure ϕ represents the physical magnitude of a stimulus (e.g., the loudness of a tone) and Ψ represents the subjective magnitude assigned to it. The figures depict theoretical models of how the design of an experimental task can influence the subjective magnitude of any individual stimulus and the relationship between the subjective magnitudes of a pair of stimuli. A shows the effects of using a broad (L_1, L_2) or narrow (S_1, S_2) range of stimuli. B shows the effect of using or not using stimuli close to the threshold of perceptibility. C shows the effect of choosing a standard stimulus (ST) from various positions in the range. MOD (or modulus) refers to the number given to that stimulus. D shows the effect of having the first judged stimulus close to or far from the standard (relative to the entire set of variables). E represents the effect of using bounded or unbounded sets of numbers and whether fractions are required or allowed as numerical responses. F reflects the use of a large or small number for the modulus (e.g., 10 vs. 1,000). Further details may be found in Poulton (1968, 1977).

strategy of "just ask them" runs into trouble.

The problem of poorly (or even misleadingly) worded questions in attitude surveys is well known, although not necessarily well resolved (Payne 1952; Zeisel 1980). For example, a major trade publication recently presented the results of a survey of public attitudes toward the chemical industry containing the following question: "Some people say that the prime responsibility for reducing exposure of workers to dangerous substances rests with the workers themselves, and that all substances in the workplace should be clearly labeled as to their levels of danger and workers then encouraged or forced to be careful with these substances. Do you agree or disagree?" It is hard to know what one is endorsing when one says yes, no, or I don't know to such a complex and unclear question.

Although annoying, ambiguous wording is, in principle, a relatively easy problem to deal with because there are accepted ways to "do it right." Other issues in value elicitation are more troublesome. Even though opinions refer to an internal object (one's thoughts and desires), they face many of the same formulation problems as psychophysical judgments of external objects. For example, just as the judged noisiness of a tone depends greatly upon the range of options offered (see Figure 4A), Parducci (1974) has found that judged satisfaction with one's state in life may depend upon the range of states considered. In an attempt to establish a dollar value for aesthetic degradation of the environment, Brookshire, Ives, and Schulze (1976) asked visitors to Lake Powell how much they would be willing to pay in increased users' fees in order not to have an ugly (coal-fired) power plant looming on the opposite shore. They asked "Would you pay \$1, \$2, \$3?" and so on, until the respondent answered "No" and then they retreated in decrements of a quarter (e.g., "Would you pay \$5.75, \$5.50, . . . ?"). Rather different numerical values might have been obtained had the bidding procedure begun at \$100 and decreased by steps of \$10 or with other plausible variants. Any respondents who were not sure what they wanted in dollars and cents might naturally and necessarily look to the range of options presented, the difference between first and second options, and so on, for cues as to what are reasonable and plausible responses.

A psychophysicist can trust that people will have a fully-elaborated sensory response to any clearly-presented stimulus. The student of values has no such assurance. At first glance, it might seem as though questions of value are the last redoubt of unaided intuition. Who knows better than an individual what he or she prefers? When people are considering simple, familiar events with which they have direct experience, it may be reasonable to assume that they have well-articulated opinions. Regarding the novel, global consequences potentially associated with CO_2-induced climatic change, nuclear meltdowns, or genetic engineering, that may not be the case (Fischhoff, Slovic, and Lichtenstein 1980). Our values may be incoherent, not thought through. In thinking about what are acceptable levels of risk, for example, we may be unfamiliar with the terms in which issues are formulated (e.g., social discount rates, minuscule probabilities, or megadeaths). We may have contradictory values (e.g., a strong aversion to catastrophic losses of life and a realization that we're no more moved by a plane crash with 500 fatalities than by one with 300). We may occupy different roles in life (parents, workers, children) that produce clear-cut but inconsistent values. We may vacillate between incompatible, but strongly held, positions (e.g., freedom of speech is inviolate, but should be denied to authoritarian movements). We may not even know how to begin thinking about some issues (e.g., the appropriate tradeoff between the opportunity to dye one's hair and a vague, minute increase in the probability of cancer 20 years from now). Our views may undergo changes over time (say, as we near the hour of decision or of experiencing the consequence) and we may not know which view should form the basis of our decision.

As a result, the particular or peculiar way that issues are posed by nature, scientists, politicians, merchants, or the media may have great influence over which responses emerge as apparent expressions of people's values. In cases where people do not know, or have difficulty appraising, what they want, problem representations may become major forces in shaping the values (apparently) expressed in the responses they require. Representations can induce random error (by confusing the respondent), systematic error (by hinting at what the "correct" response is), or unduly extreme judgments (by suggesting clarity and coherence of opinion that are not warranted). In such cases, the method becomes the message. If elicited values are used to guide policy, they may lead to decisions not in the decision maker's best interest, to action when caution is desirable (or the opposite) or to the obfuscation of poorly formulated views needing careful development and clarification.

An extreme, but not uncommon, situation is having no opinion and not realizing it. In that state, we may respond with the first thing that comes to mind once a question is asked and then commit ourselves to maintaining that first expression and to mustering support for it, while suppressing other views and uncertainties. As a result, we may be stuck with stereotypic or associative responses, generated without serious contemplation. The low rates of "no opinion" responses encountered by surveys addressing diverse and obscure topics suggests that most people are capable of providing some answer to whatever question is put to them. In many instances, however, these responses may reflect a desire to be counted rather than deeply held opinions.

Table 1 summarizes some of the ways in which an elicitor may affect a respondent's judgments of value. These begin with deciding whether there is something to question. By asking about the desirability of premarital sex, interracial dating, daily prayer, freedom of expression, or the fall of capitalism, the elicitor may legitimate possibilities that were previously viewed as unacceptable or cast doubts on events that were previously unquestioned. Opinion polls help set our national agenda by the questions they do and do not ask. Advertising helps set our personal agendas by the questions it induces us to ask ourselves (two-door or four-door?) and the answers it takes for granted (more is better).

Once the issue has been evoked, it must be given a label. In a world with few hard evaluative standards, such interpretations may be very important. For example, we found that the attractiveness of insurance may decline if its one certain consequence is labeled "sure loss" rather than "premium" (Fischhoff, Slovic, and Lichtenstein 1980). When these two versions are presented sequentially people often reverse their preferences for the two options. Table 2 shows a labeling effect that produced a reversal of preference with practicing physicians; most preferred treatment A over treatment B, and treatment D over treatment C, despite the formal equivalence of A and C and of B and D. Saving lives and losing lives afforded very different perspectives on the same problem.

People solve problems, including the determination of their own values, with what comes to mind. The more detailed, exacting, and creative their inferential process is, the more likely they are to think of all they know about a question. The briefer that process becomes, the more they will be controlled by the relative accessibility of various considerations. Accessibility may be related to importance, but it is also related to the associations that are evoked, the order in which questions are posed, imaginability, concreteness, and other factors only loosely related to importance. As one example of how an elicitor may (perhaps inadvertently) control respondents' perspective, Turner (1980) observed a large difference in responses to as simple a question as "Are you happy?" on two simultaneous surveys of the same population. The apparent source of the difference was that one preceded the happiness question with a set of questions about married life. In the U.S., married people generally report being happier than unmarried people. Reminding them of that aspect of their life apparently changed the information they brought to the happiness question.

It would be comforting to be able to say which way of phrasing this question is most appropriate. However, there is no general answer. One needs to know why the question is being asked. If one wants to predict the quality of casual encounters, then a superficial measure of happiness may suffice. However, an appraisal of na-

Table 1. Ways that an Elicitor May Affect A Respondent's Judgments of Value

Defining the issue
- Is there a problem?
- What options and consequences are relevant?
- How should options and consequences be labeled?
- How should values be measured?
- Should the problem be decomposed?

Controlling the respondent's perspective
- Altering the salience of perspectives
- Altering the importance of perspectives
- Choosing the time of inquiry
- Changing confidence in expressed values
- Changing the apparent degree of coherence

Changing the respondent
- Destroying existing perspectives
- Creating perspective
- Deepening perspectives

Source: Fischhoff, Slovic, and Lichtenstein (1980, p. 123).

Table 2. Two Formulations of a Choice Problem

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. The accepted scientific estimate of the consequences of the program are as follows:

Lives Saved

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is ⅓ probability that 600 people will be saved, and ⅔ probability that no people will be saved.

Which of the two programs would you favor?

Lives Lost

If Program C is adopted, 400 people will die.

If Program D is adopted there is ⅓ probability that nobody will die, and ⅔ probability that 600 people will die.

Which of the two programs would you favor?

Source: Tversky and Kahneman (1981).

tional malaise or suicide potential may require a questioning procedure that evokes a fuller appreciation of the components of respondents' lives. It has been known for some time that white interviewers evoke more moderate responses from blacks on race-related questions than do black interviewers. The usual response has been to match the races of interviewer and interviewee (Martin 1980). This solution may be appropriate for predicting voting behavior or conversation in same-race bars, but not for predicting behavior of blacks in white-dominated work places.

Conclusion

The fact that one has a question is no guarantee that others have answers, or even that they have devoted any prior thought to the matter. When one must have an answer (say, because public input is required by statute), there may be no substitute for an elicitation procedure that educates respondents about how they might look at the question. The possibilities for manipulation in such interviews are obvious. However, one cannot claim to be serving respondents' best interests (letting them speak their minds) by asking a question that only touches one facet of a complex and incompletely formulated set of views.

ASSERTION 3: LAY PEOPLE ARE NOT STUPID OR IRRATIONAL

Given the methodological cautions just discussed, no one has a clear picture of what people know about risk. Our best guess, based on such research as is available, is that people's perceptions may sometimes be erroneous but they are seldom stupid or irrational.

How Accurate Are Lay Perceptions of Risk?

Since the answer obviously includes an element of "it depends," and since systematic research on this question has only just begun, it is hard to give a general assessment. For any kind of risk information, one should ask a series of questions:

1. What are its formal properties?
2. What are its observable signs?
3. How are those signs revealed to the individual?
4. Are they contradicted, supported, or hidden by immediate experience?
5. Do people have an intuitive grasp of such information?
6. If their intuitions are faulty, what is the nature of their misunderstanding and how severe are its consequences?
7. Does natural experience provide useful feedback and induce improvement?

These questions ask, in essence, whether people's cognitive skills are adequate for coping with the information they receive. Existing research suggests that their skills are far from perfect. People seem to lack the intuitions and cognitive capacity for dealing with complex, probabilistic problems. As a result, they resort to judgmental heuristics, or rules of thumb that allow them to reduce such problems to simpler and more familiar terms (Tversky and Kahneman 1974; Slovic, Fischhoff, and Lichtenstein 1977). On the bright side, these strategies are quite adaptive, in the sense that they always produce some answer and that answer is often moderately accurate. They are maladaptive in that they can produce erroneous judgments and in that the ease with which they are applied may inhibit the search for superior methods.

Figure 5 shows average subjective estimates of the frequency of death from 41 sources, obtained in the study by Lichtenstein et al. (1978) mentioned earlier. One pattern emerging from these results is that people have a highly consistent subjective scale of frequency. Several different ways of asking people to assess these risks produced highly similar subjective orderings of their magnitude. A second finding of note is that these judgments correlated fairly well with available statistical estimates of frequency.

Nevertheless, there are some blemishes on this seemingly rosy picture. One is that differences between the judged frequencies of the most and least frequent events were much smaller than the corresponding differences in the statistical estimates. The former varied over 3–4 orders of magnitude, the latter over 6. Difficulties in using large numbers do not seem to have been

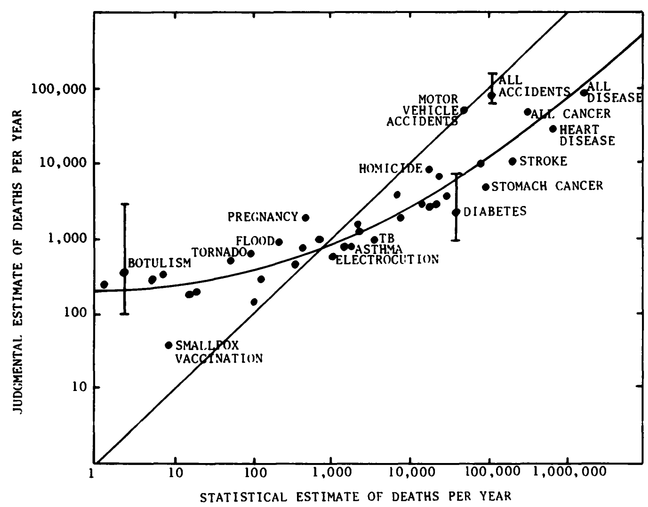

Figure 5. Relationship between judged frequency and the actual number of deaths per year for 41 causes of death. Respondents were told that about 50,000 people per year die from motor vehicle accidents. If judged and actual frequencies were equal, the data would fall on the straight line. The points, and the curved line fitted to them, represent the averaged responses of a large number of lay people. While people were approximately accurate, their judgments were systematically distorted. To give an idea of the degree of agreement among subjects, vertical bars are drawn to depict the 25th and 75th percentile of individual judgment for botulism, diabetes, and all accidents. Fifty percent of all judgments fall between these limits. The dispersion of responses for the other 37 causes of death was similar. Source: *Slovic et al. (1979, p. 565).*

the culprit here, as there was no such flattening in comparable tasks using the frequency of words (in written English) and occupations (in the U.S.) for subject matter.

A final pattern in the results is that relative to the primary bias (the flatness of best-fit line of curve), one can see substantial *secondary* biases, large differences in the estimated frequency of events with similar statistical frequencies. For example, accidents were judged to cause as many deaths as diseases, whereas diseases actually take about 15 times as many lives. Homicides were incorrectly judged to be about as frequent as fatal strokes, although the latter actually claim about 11 times as many lives. Frequencies of death from botulism, tornadoes, and pregnancy (including childbirth and abortion) were also overestimated.

This last pattern of responses seems to illustrate use of one of the most general judgmental heuristics: availability. People who use this heuristic judge an event as likely or frequent if instances of it are easy to imagine or recall. Frequently occurring events generally come to mind more readily than do rare events. Thus, availability is often an appropriate cue. However, availability is also affected by numerous factors unrelated to frequency of occurrence. For example, a recent disaster or a vivid film such as *Jaws* can seriously distort risk judgments (Tversky and Kahneman 1974).

Table 3 lists the lethal events whose frequencies were most poorly judged in our studies. In keeping with availability considerations, overestimated items were dramatic and sensational whereas underestimated items tended to be unspectacular events that claim one victim at a time and are common in nonfatal form. Indeed, Combs and Slovic (1979) found that overestimated hazards also tended to be disproportionately mentioned in the news media. Availability might also be blamed for the flattening in Figure 5, as life is too short for experience to provide large differences in observed frequencies. A study by Slovic, Fischhoff, and Lichtenstein (1979) reveals a similar pattern of results in all these respects for estimates of the fatalities in an average year from various technologies.

Is this good performance or bad? One possible summary is that it may be about as good as can be expected, given that these people were neither specialists in the hazards considered, nor exposed to a representative sample of information.

Table 3. Bias in Perceived Frequency

Most Overestimated	*Most Underestimated*
1. All accidents	1. Smallpox vaccination
2. Motor vehicle accidents	2. Diabetes
3. Pregnancy, childbirth, abortion	3. Stomach cancer
4. Tornado	4. Lightning
5. Flood	5. Stroke
6. Botulism	6. Tuberculosis
7. All cancer	7. Asthma
8. Fire and flames	8. Emphysema
9. Venemous bite or sting	
10. Homicide	

Source: Slovic, Fischhoff, and Lichtenstein (1979, p. 16).

Such accurate perception of misleading samples of information was suggested earlier as an explanation why people perceive consistent attitudes toward risk in others that are not supported by systematic observation. It might also underlie another apparent judgmental bias, people's predilection for exaggerating their personal immunity from many hazards. The vast majority of individuals believe themselves to be better than average drivers, more likely than average to live past 80, less likely than average to be injured by tools they operate and so on (Svenson 1979). Although such perceptions are obviously unrealistic, from the perspective of each individual's experience, the risks do look very small. Consider automobile driving: despite driving too fast, tailgating, and so on, poor drivers make trip after trip without mishap. This personal experience may demonstrate to them their exceptional skill and safety. Moreover, their indirect experience via the media shows them that when accidents happen, they happen to others. One could hope that people would see beyond the limits of their own minds and information, but inability to do so need not render them incompetent to make decisions in their own behalf.

ASSERTION 4: EVEN WHEN LAY PEOPLE SEEM TO BE BADLY CONFUSED, THEIR BEHAVIOR MAY BE STILL BE QUITE REASONABLE

Could They Have Been Better Informed?

Failure to understand sampling bias suggests that had they merely been provided with better information, lay people's performance would have been markedly better. The source of much technical information is, of course, the technical community. There are a number of ways in which the experts fail, either deliberately or inadvertently, to inform the public. One is by not telling the whole story about the hazards they know best, because they fear that the information would make the public anxious, because dissemination is not their job, or because they have a vested interest in keeping things quiet (Hanley 1980).

What happens when the tale that the experts tell is incomplete? Two contradictory hypotheses come to mind. One is that when major items are omitted from an expert's presentations, they are quickly discovered, discrediting the expert and perhaps producing an exaggerated perception of the presentation's incompleteness ("If I caught that omission, how many are there that I didn't catch?"). On the other hand, perhaps what is out of sight is effectively out of mind, in keeping with availability considerations. Fischhoff, Slovic, and Lichtenstein (1978) presented people with various versions of a fault tree describing ways in which a car might fail to start. These versions differed in how much of the full tree were left out. The participants' task was to estimate the degree of completeness. We found great insensitivity to omissions; even omission of major, commonly known systems, like the ignition and fuel systems, led to only minor decreases in perceived completeness.

A second foible that may be exacerbated by the experts is the difficulty people have in resolving the conflicts generated by life's gambles (Lichtenstein and Slovic 1973). People often attempt to deny uncertainty in order to reduce its attendant anxiety. They want statements of fact, not probability. Thus, just before hearing a blue-ribbon panel of scientists report being 95 percent certain that cyclamates do not cause cancer, a former Food and Drug Administration commissioner, Alexander Schmidt said, "I'm looking for a clean bill of health, not a wishy-washy, iffy answer on cyclamates." Likewise, former Senator Edmund Muskie has called for "one-armed" scientists who do not respond "on the one hand, the evidence is so, but on the other hand . . . " when asked about the health effects of pollutants. In this atmosphere, unduly confident, one-fisted debators, ready to make definitive statements beyond the data available, may unjustifiably win the day from more even-handed scholars. The temptation may be very great to give people the simple answers they seem to want.

Social as well as psychological processes help to make balanced presentations an endangered genre. The constraints of legal settings (Piehler et al. 1974), the exigencies of the political arena, and the provocations of the news media all encourage adversarial encounters that are inhospitable to properly qualified scientific evidence. Lay people viewing such shouting matches perhaps should not be faulted for wondering about scientists or feeling that "since they can't agree, my guess may be as good as theirs."

A final way to keep the public from knowing any better is to present results in terms meaningful only to other experts. Summary statistics (in technical jargon) may not be the best or only way to tell about risks. Alternatives and complements might be descriptions of credible accidents (carefully specifying the time period considered) or explanations of how things work and what ameliorative strategies are possible. One positive repercussion of the nuclear power plant accident at Three Mile Island was that for a time the public was educated in plain English about the process of nuclear power generation, not just presented with conflicting assertions about its overall safety.

Were They Solving a Different Problem?

In analyzing people's behavior, perhaps the most reasonable assumption is that there is some method in any apparent madness. For example, Zentner (1979) berates the public because its rate of concern about cancer (as measured by newspaper coverage) is increasing faster than the cancer rate. One reasonable explanation for this pattern is that people may believe that too little concern has been given to cancer in the past (e.g., our concern for acute hazards like traffic safety and infectious disease allowed cancer to creep up on us). A second is that people may realize that some forms of cancer are the only major causes of death whose rate is increasing. Just as it is counterproductive for lay people to view technology promoters as evil on the basis of insufficient or misinterpreted evidence, it is counterproductive for promoters to view lay people as misinformed and irresponsible on similar grounds.

Other apparently irrational behavior can be attributed to the rational pursuit of unreasonable objectives. This can happen when one rejects the problem definition deemed reasonable by the presenting body. Consider, for example, an individual who is opposed to increased energy consumption but is only asked about which energy source to adopt. The answers to these narrow questions provide a de facto answer to the broader question of growth. Such an individual may have little choice but to fight dirty, engaging in unconstructive criticism, poking holes in analyses supporting other positions, or ridiculing opponents who adhere to the more narrow definition.

Indeed, some participants in technology debates are in it for the fight. Many approaches to determining acceptable-risk levels (e.g., cost-benefit analyses) make the political-ideological assumption that our society is sufficiently cohesive and common-goaled that its problems can be resolved by reason and without struggle. Although such a "get on with business" orientation will be pleasing to many, it will not satisfy all. For those who do not believe that society is in a fine-tuning stage, a technique that fails to mobilize public consciousness and involvement has little to recommend it (Fischhoff et al. 1981). Their strategy may involve a calculated attack on what they interpret as narrowly defined rationality.

Even when experts and lay people have the same goals, they may be solving different problems. For example, lay people are often maligned for their failure to wear seat belts. However, their formulation of the problem may be quite different from that of the safety expert. The latter sees the tens of thousands of lives that might be saved from the small statistical reduction in each trip's probability of ending in death. Drivers may see only that minute probability. From their perspective, the effort of buckling up may not be worth a probabilistic reduction in the minute chance of dying on any given trip. When shown that the .00000025 probability of an average trip ending in a fatal accident compounds to a lifetime probability of .01, people may view seat belts more favorably (Slovic, Fischhoff, and Lichtenstein 1978).

Other apparent differences of opinion between lay people and technical people may be traced to differences in semantics. Figure 6 contrasts expert and lay assessments of the risk from 25 activities and technologies. The upper figure suggests that for experts, "risk" and "annual fatalities" may be synonymous (and that their intuitive estimates of the latter agreed with available calculated estimates). The lower figure shows a strikingly lower slope for lay judgments of risk. It would be tempting to explain the contrast between these figures as reflecting lay people's insensitivity to the great differences in risk from these technologies. However, when lay people were asked to assess average year fatal-

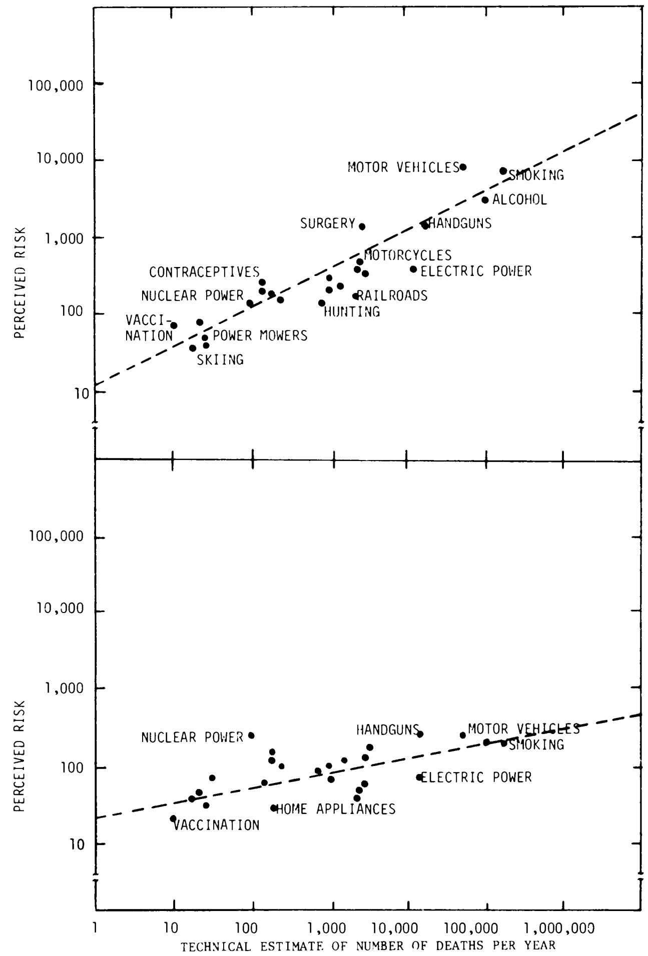

Figure 6. Judgments of perceived risk for experts (top) and lay people (bottom) plotted against technical estimates of annual fatalities for 25 technologies and activities. Each point represents the average responses of the participants. The dashed lines are the straight lines that best fit the points. The experts' risk judgments are seen to be more closely associated with annual fatality rates than are the lay judgments. Source: *Slovic, Fischhoff, and Lichtenstein (1979, p. 19).*

ities, they gave responses (not shown) that were strikingly similar to those of the experts (in Fig. 6). It is only when asked about "risk" that they give something different. Subsequent research (Slovic, Fischhoff, and Lichtenstein 1979, 1980) has shown that lay judgments of risk can be predicted by a combination of fatality estimates for an average year and estimates of the number of deaths that would arise from the most disastrous year one imagines happening in one's lifetime.

The lay participants in these studies have typically been drawn from a strongly antinuclear population. Once these semantic questions are clarified, the locus of their disagreement with pronuclear experts is seen to lie in assessments of the catastrophic potential of plants, the aspect of riskiness whose assessment is most heavily a matter of judgment.

Another example of terminological differences being interpreted to the detriment of lay people's reputation may be found in weather forecasting. Some meteorologists have recently proposed abandoning the use of probabilistic precipitation forecasts because of reports that the public is confused by them. A study by Murphy et al. (1980), however, found that confusion is created not by the use of probability, but by uncertainty about what the forecasted event is: Does a 70 percent chance of rain mean "rain during 70 percent of the day," "rain over 70 percent of the area," "70 percent chance of rain somewhere in the area," or "70 percent chance of rain at some particular spot?" The last mentioned option is the correct one (with the spot being the local weather station); it was picked by a minority of respondents, whether the likelihood was described with a number (70 percent) or verbal label (very likely).

A precondition for understanding other people's beliefs is understanding their conceptual universe. With Figure 1, Starr (1972) argued that voluntariness was an important qualitative feature of risk to lay people and that setting a double standard according to voluntariness was not an unreasonable basis for social policy. Fischhoff et al. (1978) found that people believed that voluntariness should set a double standard (although it was not doing so today). However, other qualitative criteria, like how well a hazard is known to those exposed to it (Figure 3) or whether it evokes a feeling of dread, produced similar response patterns.

Figure 7 describes an attempt to order this universe of concepts as a first step to understanding what people want, to be followed by analysis of how reasonable their desires are as guides to social policy. It reflects judgments of 90 hazards on 18 qualitative aspects of risk that have been cited, by one or another commentator in risk issues, as dominating lay perceptions. It shows that these 18 aspects could be adequately summarized by two underlying dimensions. The vertical dimension seems to reflect some combination of novelty and mystery; the horizontal dimension, the possibility of uncontrollable consequences. Nuclear power is uniquely aversive on both dimensions.[1] Whatever the interpretation given to these dimensions, they do provide structure and parsimony to a universe of discourse for which the relationship between terms was poorly understood. Understanding *how* people think about risks is a precondition for even collecting the data with which we will discipline our speculations about *what* they think.

ASSERTION 5: EXPERTS ARE FALLIBLE, TOO

By definition, experts know more than anybody else about the risks in their domain. Seldom, however, do they know everything. Typically, they have hard facts or a trustworthy model for only part of the problem or a related problem. The final steps toward the information

[1]Alternatively, one could view nuclear power as defining a category of hazard and rotate these axes (around the current origin) so that nuclear power anchors one end of one axis (and bicycles its other end) while warfare and cosmetics roughly mark the extremes on the other axis. From this perspective, the two dimensions might be labeled "insidiousness" or "dread" (for which nuclear power receives extreme marks) and "intentionality (of risk)" (for which crime, warfare, terrorism, nuclear weapons, dynamite, and handguns receive outstanding scores).

Figure 7. Two dimensions underlying 18 qualitative aspects of risk. Source: *Slovic, Fischhoff, and Lichtenstein (1980).*

needed by decision makers must be traversed by informed speculation or judgment. It is natural to ask how well experts fulfill these assignments. Although the evidence is spotty, there is no particular reason to believe that the thought processes of experts are appreciably different from those of lay people. Their fund of substantive knowledge tells experts where to look for information and what solutions have worked in the past (DeGroot 1965; Feigenbaum 1978). When forced to go beyond the limits of the available data or to convert their incomplete knowledge into judgments usable by risk assessors, they may fall back on intuitive processes, just like everyone else. A few examples of the kinds of problems they, like lay people, may encounter follow.

Insensitivity to Sample Size

In an article entitled "Belief in the Law of Small Numbers," Tversky and Kahneman (1971) showed that even statistically sophisticated individuals have poor intuitions about the size of sample needed to test research hypotheses adequately. In particular, they expect small samples to represent the populations from which they were drawn to a degree that can only be assumed with much larger samples. This tendency leads them to gamble their research hypotheses on underpowered small samples, to place undue confidence in early data trends, and to underestimate the role of sampling variability in causing results to deviate from expectations (preferring, instead, to offer causal explanations for discrepancies). In a survey of standard hematology texts, Berkson, Magath, and Hurn (1939–40) found that the maximum allowable difference between two successive blood counts was so small that it would normally be exceeded by chance 66 to 85 percent of the time. They mused about why instructors often reported that their best students had the most trouble attaining the desired standard.

Small samples mean low statistical power; that is, a small chance of detecting phenomena that really exist in the data. Cohen (1962) surveyed published articles in a respected psychological journal and found very low power. Even under the charitable assumption that all underlying effects were large, a quarter of the studies had less than three chances in four of showing statistically significant results. He goes on to speculate that the one way to get a low-power study published is to keep doing it again (perhaps making subtle variations designed to "get it right next time") until a significant result occurs. As a result, published studies may be unrepresentative of the set of conducted studies in a way that inflates the rate of spuriously significant results.

Page (1981) has similarly shown the low power of representative toxicological studies. In designing such studies, one inevitably must make a tradeoff between avoiding false alarms (e.g., erroneously calling a chemical a carcinogen) and misses (e.g., erroneously calling a chemical a non-carcinogen). For any given false alarm rate, the smaller the sample size the larger the miss rate. Because scientists are typically most concerned about false alarms, that rate is usually set at some conventional value (e.g., .05). As a result, variations in sample size affect primarily the miss rate, with decreased samples leading to increased miss rates. In this way, wayward intuitions may lead to underpowered experimental designs that represent, perhaps inadvertently, a social policy that protects chemicals more than people.

Hindsight Bias

Experimental work has shown that in hindsight, people consistently exaggerate what could have been anticipated in foresight. They not only tend to view what has happened as having been inevitable, but also to view it as having appeared relatively inevitable before it happened. People believe that others should have been able to anticipate events much better than was actually the case. They even misremember their own predictions so as to exaggerate in hindsight what they knew in foresight (Fischhoff 1980a).

The revisionist history of strategic surprises (e.g., Lanir 1978; Wohlstetter 1962) argues that such misperceptions have vitiated the efforts of scholars and scalpers attempting to understand questions like, "Who goofed at Pearl Harbor?" These expert scrutinizers were not able to disregard the knowledge they had only as a result of knowing how things turned out. Although it is flattering to believe that we personally would not have been surprised, failing to realize the difficulty of the task that faced the individuals we are second-guessing may leave us very exposed to future surprises.

Methodological treatises for professional historians contain numerous warnings about related tendencies. One such tendency is telescoping the rate of historical processes, exaggerating the speed with which "inevitable" changes are consummated (Fischer 1970). Mass immunization against polio seems like such a natural idea that careful research is needed to show that its adoption met substantial snags, taking almost a decade to complete (Lawless 1977). A second variant of hindsight bias may be seen in Barraclough's (1972) critique of the historiography of the ideological roots of Naziism; looking back from the Third Reich, one can trace its roots to the writings of many authors from whose writings one could not have projected Naziism. A third form of hindsight bias, also called "presentism," is to imagine that the participants in a historical situation were fully aware of its eventual importance ("Dear Diary, The Hundred Years' War started today." Fischer 1970). More directly relevant to the resolution of scientific disputes, Lakatos (1970) has argued that the critical experiment, unequivocally resolving the conflict between two theories or establishing the validity of one, is typically an artifact of inappropriate reconstruction. In fact, "the crucial experiment is seen as crucial only decades later. Theories don't just give up, a few anomalies are always allowed. Indeed, it is very difficult to defeat a research program supported by talented and imaginative scientists" (pp. 157–158). Future generations may be puzzled by the persistence of the antinuclear movement after the 1973 Arab oil embargo guaranteed the future of nuclear power, or the persistence of nuclear advocates after Three Mile Island sealed the industry's fate—depending on how things turn out. Perhaps the best way to protect ourselves from the surprises and reprobations of the future in managing hazards is to "accept the fact of uncertainty and learn to live with it. Since no magic will provide certainty, our plans must work without it" (Wohlstetter 1962, p. 401).

Judging the Quality of Evidence

Since cognitive and evidential limits prevent scientists from providing all the answers, it is important to have an appraisal of how much they do know. It is not enough to claim that "these are the ranking experts in the field," for there are some fields in which the most knowledgeable individuals understand a relatively small portion of all there is to be known.

Weather forecasters offer some reason for encouragement. Figure 8 demonstrates the validity of short-term probabilistic precipitation forecasts. There is at least some measurable precipitation on about 70 percent of the occasions for which they say that there is a 70 percent chance of rain. The conditions under which forecasters work and train suggest prerequisites for good performance in probabilistic judgment: (a) great amounts of practice; (b) the availability of statistical data offering historical precipitation base rates (indeed, forecasters might be fairly well calibrated if they ignored the murmurings of their intuitions and always responded with the base rate); (c) computer-generated predictions for each situation; (d) a readily verifiable criterion event (measurable precipitation), offering clear feedback; (e) explicit admission of the imprecision of the trade and the need for training. In experimental work, we have found that large amounts of clearly characterized, accurate, and personalized feedback can improve the probability assessments of lay people (Lichtenstein and Fischoff 1981).

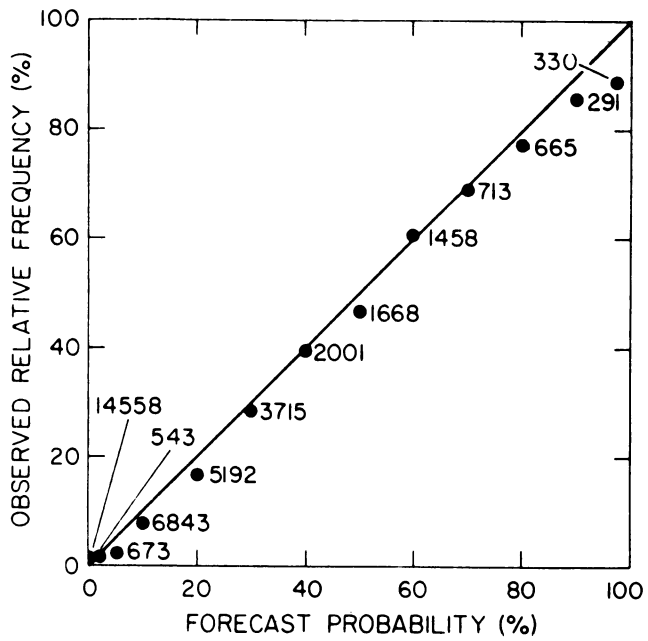

Figure 8. Validity of probabilistic precipitation forecasts. Source: *Murphy & Winkler (1977, p. 6).*

Training professionals to assess and express their uncertainty is, however, a rarity. Indeed, the role of judgment is often acknowledged only obliquely. For example, civil engineers do not routinely assess the probability of failure for completed dams, even though approximately one dam in 300 collapses when first filled (U.S. Government 1976). The "Rasmussen" Reactor Safety Study (U.S. Nuclear Regulatory Commission 1975) was an important step toward formalizing the role of risk in technological systems, although a subsequent review was needed to clarify the extent to which these estimates were but the product of fallible, educated judgment (U.S. Nuclear Regulatory Commission 1978).

Ultimately, the quality of experts' assessments is a matter of judgment. Since expertise is so narrowly distributed, assessors are typically called upon to judge the quality of their own judgments. Unfortunately, an extensive body of research suggests that people are overconfident when making such assessments (Fischhoff, Slovic, and Lichtenstein 1977). A major source of such overconfidence seems to be failure to appreciate the nature and tenuousness of the assumptions upon which judgments are based. To illustrate with a trivial example, when asked "To which country are potatoes native? (a) Ireland (b) Peru?" many people are very confident that answer (a) is true. The Irish potato and potato blight are familiar to most people; however, that is no guarantee of origin. Indeed, the fact that potatoes were not indigenous to Ireland may have increased their susceptibility to blight. Table 4 suggests some forms of

Table 4. Some Common Problems Leading to Underestimation of Risks

- Failure to consider the ways in which human errors can affect technological systems. Example: Due to inadequate training and control room design, operators at Three Mile Island repeatedly misdiagnosed the problems of the reactor and took inappropriate corrective actions. A minor incident thus became a major accident.
- Overconfidence in current scientific knowledge. Example: The failure to recognize the harmful effects of X rays until use had become widespread and largely uncontrolled.
- Failure to appreciate how technological systems function as a whole. Example: The DC-10 failed in several early flights because its designers had not realized that decompression of the cargo compartment would destroy vital parts of the plane's control system running through it.
- Slowness in detecting chronic, cumulative effects. Example: Although accidents to coal miners have long been recognized as one cost of operating fossil-fueled plants, the effects of acid rains on ecosystems were slow to be discovered.
- Failure to anticipate human response to safety measures. Example: The partial protection afforded by dams and levees gives people a false sense of security and promotes development of the flood plain. When a rare flood does exceed the capacity of the dam, the damage may be considerably greater than it would have been had the flood plain not been "protected."
- Failure to anticipate "common-mode failures." Example: Because electrical cables controlling the multiple safety systems of the reactor at Browns Ferry, Alabama, were not spatially separated, all five emergency core cooling systems were crippled by a single fire.

insensitivity to assumptions to which technical experts may be particularly prone. In one quasi-experimental study, Hynes and Vanmarke (1976) asked several internationally known geotechnical experts to predict the height of fill at which an embankment would fail and to give confidence intervals for their estimates. Without exception, the true values fell outside the confidence intervals, a result akin to that observed with other tasks and respondent populations (Lichtenstein, Fischhoff, and Phillips 1982). One of the intellectual challenges facing engineering is to systematize the role of judgment, both to improve its quality and to inform those who must rely upon it in their decision making.

Conclusion

These speculations about expert judgment should be followed by repetition of the caveat that began this paper: In the absence of systematic, reliable data, moderately informed speculation is the best one can hope to offer. A case can be made that experts differ from lay people of comparable social origin and educational level, not in the way they think, but in the substantive knowledge they have at their disposal. It is not clear that this knowledge affords them any special advantage in going beyond the available data or into realms outside their expertise. Indeed, the price for acquiring such depth of field may be a reduction in the breadth of their view. Laboratory toxicologists may, for example, be insensitive to behavioral effects experienced in the field; designers may not see flaws that are apparent to operators; theoreticians may tend to forget the simplifying assumptions underlying their models. Even when the experts have a (near) monopoly on the best available facts of a matter, they need not have a monopoly on the set of possibly valid perspectives, particularly with problems having complex social ramifications or involving the interactions of diverse systems. In such situations "the more the merrier" may be more appropriate than "too many cooks spoil the broth." There is no way to get the right answer to many risk problems; all that we can hope to do is avoid the mistakes to which each of us is attuned; the more perspectives involved, the more local wisdom is brought to bear on a problem.

IMPLICATIONS

More respectful and balanced relations between the expert and lay communities (to which any expert belongs except for a narrow range of problems) are good not only for science, but also for society. In many, if not most, cases, effective hazard management requires the cooperation of a large body of lay people. These people must agree to do without some things and accept substitutes for others; they must vote sensibly on ballot measures and choose legislators who will serve as surrogate hazard managers; they must obey safety rules and use the legal system responsibly. Even if the experts were much better judges of risk than lay people, giving experts an exclusive franchise for hazard management would mean substituting short-term efficiency for the long-term effort needed to create an informed citizenry.

For nonexperts, these findings pose an important series of challenges: to be better informed, to rely less on unexamined or unsupported judgments, to be aware of the qualitative dimensions that strongly condition our risk judgments, and to be more open to new evidence; in short, to realize the potential of being just as educable as the experts.

For experts, these findings pose what may be a more difficult challenge: to recognize and admit one's own cognitive limitations, to temper risk assessments with the important qualitative aspects of risk that influence the responses of lay people, and to create continuing and respectful relations with the public.

What do we do if disagreements persist between the experts and the public (treating them for the moment as corporate bodies)? In a democratic system, "we" don't do anything. The political process resolves the issue, for better or for worse. Elected representatives, through their votes, appointees and bureaucracy, do what needs to be done to balance the public will, the public weal, and their own needs for popularity, fulfillment, and so on.

Assume, however, that there is a dispassionate institution entrusted with resolving such disagreements (or that our courts or legislatures or civil service constitute such institutions). Could it responsibly act in accordance with the public's "fears" rather than upon the experts' "facts?" The answer could be "yes" if one (or more) of three conditions holds:

1. The lay public knows something that the experts do not. In that case, the dispassionate institution should change its best estimate of what the facts are.
2. The lay public does not know anything special, but has good reason not to be all that convinced by the evidence supporting the experts' testimony. In such situations, it may be appropriate to leave the best estimate unchanged, but to increase substantially the confidence intervals around it. The result might be delay, hedging of bets, or even switching to a more certain course of action.
3. The public is truly unreasonable, but has a deep emotional investment in its beliefs. There are costs to a society for overriding the strong wishes of its members; these include anomie, alienation, resentment, distrust, sabotage, stress, and even psychosomatic effects (whose impact is physical even when their source is illusory). Such costs could tip the balance against the action indicated by the experts' best guess.

REFERENCES

ALFIDI, J. (1971), "Informed Consent: A Study of Patient Reaction," *Journal of the American Medical Association,* 1971, 216, 1325–1329.

Applied Management Sciences (1978), "Survey of Consumer Perceptions of Patient Package Inserts for Oral Contraceptives," NTIS No. PB–248–740.

BARRACLOUGH, G. (1972), "Mandarins and Nazis," *New York Review of Books,* 19(6), 37–42.

BERKSON, J. MAGATH, T.B., and HURN, M. (1939–40), "The Error of Estimate of the Blood Cell Count as Made With the Hermocytometer," *American Journal of Physiology,* 128, 309–323.

BROOKSHIRE, D.S., IVES, B.C., and SCHULZE, W.D. (1976), "The Valuation of Aesthetic Preferences," *Journal of Environmental Economics and Management, 3,* 325–346.

COHEN, J. (1962), "The Statistical Power of Abnormal-Social Psychological Research: A Review," *Journal of Abnormal and Social Psychology,* 65(3), 145–153.

COMBS, B., and SLOVIC, P. (1979), "Causes of Death: Biased Newspaper Coverage and Biased Judgments," *Journalism Quarterly,* 56, 837–843, 849.

DAVIDSHOFER, I.O. (1976), "Risk-Taking and Vocational Choice: A Reevaluation," *Journal of Counseling Psychology,* 23, 151–154.

DE GROOT, A. (1965), *Thought and Choice in Chess,* The Hague: Mouton, 1965.

FEIGENBAUM, E.A. (1978), "The Art of Artificial Intelligence: I. Themes and Case Studies in Knowledge Engineering," in *Proceedings of the International Artificial Intelligence Conference,* MIT, 1014–1029.

FISCHER, D.H. (1970), *Historians' Fallacies,* New York: Harper and Row.

FISCHHOFF, B. (1980a), "For Those Condemned to Study the Past: Reflections on Historical Judgment," in *New Directions for Methodology of Behavior Science: Fallible Judgment in Behavioral Research,* eds. R.A. Shweder and D.W. Fiske, San Francisco: Jossey-Bass.

——— (1980b), *Informing People About the Risks of Oral Contraceptives,* Decision Research Report 80–3.

——— (in press), "Informed Consent for Transient Nuclear Workers," in *Equity Issues in Radioactive Waste Management,* eds. R. Kasperson and R.W. Kates, Cambridge, Ma.: Oelgeschlager, Gunn and Hain.

FISCHHOFF, B.; LICHTENSTEIN, S.; SLOVIC, P.; DERBY, S.; and KEENEY, R. (1981), *Acceptable Risk,* New York: Cambridge University Press.

FISCHHOFF, B., and MACGREGOR, D. (1980), *Judged Lethality,* Decision Research Report 80–4.

FISCHHOFF, B.; SLOVIC, P.; and LICHTENSTEIN, S. (1977), "Knowing With Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology: Human Perception and Performance,* 3, 552–564.

———(1978), "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation," *Journal of Experimental Psychology: Human Perception and Performance,* 4, 330–334.

——— (1980), "Knowing What You Want: Measuring Labile Values," in *Cognitive Processes in Choice and Decision Behavior,* ed. T. Wallsten, Hillsdale, N.J.: Erlbaum.

FISCHHOFF, B.; SLOVIC, P.; LICHTENSTEIN, S.; READ, S.; and COMBS, B. (1978), "How Safe Is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," *Policy Sciences,* 8, 127–152.

HANLEY, J. (1980), "The Silence of Scientists," *Chemical & Engineering News,* March 24, 5.

HYNES, M., and VANMARCKE, E. (1976), "Reliability of Embankment Performance Prediction," in *Proceedings of the ASCE Engineering Mechanics Division Specialty Conference,* Waterloo, Ontario: University of Waterloo Press.

JOUBERT, P. and LASAGNA, L. (1975), "Commentary: Patient Package Inserts," *Clinical Pharmacology and Therapeutics,* 18(5), 507–513.

KUNREUTHER, H.; GINSBERG, R.; MILLER, L.; SAGI, P.; SLOVIC, P.; BORKAN, B.; and KATZ, N. (1978), *Disaster Insurance Protection,* New York: John Wiley.

LAKATOS, I. (197), "Falsification and Scientific Research Programmes," in *Criticism and the Growth of Scientific Knowledge,* eds. I. Lakatos and A. Musgrave, Cambridge: Cambridge University Press.

LANIR, Z. (1978), *Critical Reevaluation of the Strategic Intelligence Methodology,*" Tel Aviv: Center for Strategic Studies, Tel Aviv University.

LAWLESS, E.W. (1977), *Technology and Social Shock,* New Brunswick: Rutgers.

LICHTENSTEIN, S., and FISCHHOFF, B. (1980), "Training for Calibration," *Organizational Behavior and Human Performance,* 26, 149–171.

LICHTENSTEIN, S.; FISCHHOFF, B.; and PHILLIPS, L.D. (1982), "Calibration of Probabilities: The State of the Art to 1980," in *Judgment Under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky, New York: Cambridge University Press.

LICHTENSTEIN, S., and SLOVIC, P. (1973), "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas," *Journal of Experimental Psychology,* 101, 16–20.

LICHTENSTEIN, S.; SLOVIC, P.; FISCHHOFF, B.; LAYMAN, M.; and COMBS, B. (1978), "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory,* 4, 551–578.

MARTIN, E. (1980), *Surveys as Social Indicators: Problems in Monitoring Trends,* Chapel Hill: Institute for Research in Social Science, University of North Carolina.

MURPHY, A.H.; LICHTENSTEIN, S.; FISCHHOFF, B.; and WINKLER, R.L. (1980), "Misinterpretations of Precipitation Probability Forecasts," *Bulletin of the American Meteorological Society,* 61, 695–701.

MURPHY, A.H., and WINKLER, R.L. (1977), "Can Weather Forecasters Formulate Reliable Probability Forecasts of Precipitation and Temperature?," *National Weather Digest,* 2(2), 2–9.

NISBETT, R., and ROSS, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment,* Englewood Cliffs, N.J.: Prentice-Hall.

PAGE, T. (1981), "A Framework for Unreasonable Risk in the Toxic Substances Control Act," *Carcinogenic Risk Assessment,* ed. R. Nicholson, New York: New York Academy of Sciences.

PARDUCCI, A. (1974), "Contextual Effects: A Range-Frequency Analysis, in *Handbook of Perception,* (vol. II), eds. E.C. Carterette and M.P. Friedman, New York: Academic Press.

PAYNE, S.L. (1952), *The Art of Asking Questions,* Princeton: Princeton University Press.

PIEHLER, H.R.; TWERSKI, A.D.; WEINSTEIN, A.; and DONAHER, W.A. (1974), "Product Liability and the Technical Expert, *Science,* 186, 1089–1093.

POULTON, E.C. (1968), "The New Psychophysics: Six Models for Magnitude Estimation," *Psychological Bulletin,* 69, 1–19.

——— (1977), "Quantitative Subjective Assessments are Almost Always Biased, Sometimes Completely Misleading," *British Journal of Psychology,* 68, 409–425.

ROLING, G.T.; PRESSGROVE, L.W.; KEEFFE, E.B.; and RAFFIN, S.B. (1977), "An Appraisal of Patients' Reactions to 'Informed Consent' for Peroral Endoscopy," *Gastrointestinal Endoscopy,* 24(2), 69–70.

SCHWARZ, E.D. (1978), "The Use of a Checklist in Obtaining Informed Consent for Treatment With Medicare," *Hospital and Community Psychology,* 29, 97–100.

SLOVIC, P. (1962), "Convergent Validation of Risk-Taking Measures," *Journal of Abnormal and Social Psychology,* 65, 68–71.

SLOVIC, P., and FISCHHOFF, B. (1977), "On the Psychology of Experimental Surprises," *Journal of Experimental Psychology: Human Perception and Performance,* 3, 544–551.

SLOVIC, P.; FISCHHOFF, B.; and LICHTENSTEIN, S. (1977), "Behavioral Decision Theory," *Annual Review of Psychology,* 28, 1–39.

——— (1978), "Accident Probabilities and Seat Belt Usage: A Psychological Perspective," *Accident Analysis and Prevention,* 10, 281–285.

——— (1979), "Rating the Risks," *Environment,* 21(3), 14–20, 36–39.

——— (1980), "Perceived Risk," in *Societal Risk Assessment: How Safe Is Safe Enough?* eds. R. Schwing and W.A. Albers, New York: Plenum Press, 181–214.

SLOVIC, P.; FISCHHOFF, B.; LICHTENSTEIN, S.; CORRI-

GAN, B.; and COMBS, B. (1977), "Preference for Insuring Against Probable Small Losses: Implications for the Theory and Practice of Insurance," *Journal of Risk and Insurance*, 44, 237–258.

STARR, C. (1972), "Benefit-Cost Studies in Sociotechnical Systems," in *Perspective on Benefit-Risk Decision Making,"* Washington D.C.: National Academy of Engineering.

SVENSON, O. (1979), "Are We All Less Risky and More Skillful Than Our Fellow Drivers?" *Acta Psychologica*, 47, 143–148.

TURNER, C.F. (1980), "Surveys of Subjective Phenomena," in *The Measurement of Subjective Phenomena*, ed. D. Johnston, Washington, D.C.: GPO.

TVERSKY, A., and KAHNEMAN, D. (1971), "The Belief in the Law of Small Numbers," *Psychological Bulletin*, 76, 105–110.

——— (1974), "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 211, 453–458.

U.S. GOVERNMENT (1976), *Teton Dam Disaster*, Washington, D.C.: Committee on Government Operations.

U.S. NUCLEAR REGULATORY COMMISSION (1975), *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*, Washington, D.C.

U.S. NUCLEAR REGULATORY COMMISSION (1978), *Risk Assessment Review Group to the U.S. Nuclear Regulatory Commission*, Washington, D.C.

WEINSTEIN, N.D. (1979), "Seeking Reassuring or Threatening Information About Environmental Cancer," *Journal of Behavioral Medicine*, 16, 220–224.

WHITE, G. (ed.) (1974), *Natural Hazards: Local, National and Global*, New York: Oxford University Press.

WOHLSTETTER, R. (1962), *Pearl Harbor: Warnings and Decision*, Stanford, Calif.: Stanford University Press.

ZEISEL, J. (1980), "Lawmaking and Public Opinion Research: The President and Patrick Caddell," *American Bar Foundation Research Journal*, 1, 133–139.

ZENTNER, R.D. (1979), "Hazards in the Chemical Industry," *Chemical and Engineering News*, Nov. 5, 25—32.