Journal of Behavioral Decision Making
J. Behav. Dec. Making, **12**: 149–163 (1999)

Fifty–Fifty = 50%?

BARUCH FISCHHOFF* and WÄNDI BRUINE DE BRUIN
Carnegie Mellon University, USA

ABSTRACT

Several recent surveys have asked respondents to estimate the probabilities of relatively unlikely events, such as dying from breast cancer and smoking. Examination of their response distributions reveals a seemingly inappropriate 'blip' at 50. The two studies reported here indicate that it reflects a response artifact associated with open-ended probability scales. The blip vanishes when a response scale with explicit response options is offered. Apparently, the open-ended format leads some people to use the 50% option as 'fifty–fifty', an expression of having no idea as to the answer. As a result, the accuracy of people's reported beliefs depends on the response scale used, as well as on how it evokes and channels such feelings of epistemic uncertainty. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS probability elicitation; confidence; epistemic uncertainty; risk perception

People often need to know the probabilities of risk-related events. For example, deciding whether to have a mammogram requires knowing the (interrelated) probabilities of having breast cancer, detecting it, and having treatment succeed. Deciding whether to smoke requires knowing the probability of smoking-related health effects, as well as the probability of being able to quit, once having started. Deciding whether to invest in one's education requires knowing the associated change in probability of unemployment. And so on. Poorly understood probabilities can mean ineffective, even dangerous, decisions.

Unfortunately, reliable estimates of these probabilities are often hard to come by, or hard to interpret even when they are available (Freudenberg, 1988; National Research Council, 1989; Slovic, 1987). As a result, it should not be surprising that studies eliciting lay risk judgments have sometimes found troubling misunderstandings. One large national study reported that women significantly overestimate the probability of breast cancer (Black, Nease and Tosteson, 1995). Another reported that people significantly overestimate the probability of lung cancer for smokers (Viscusi, 1992). Yet another reported that respondents substantially overestimate the probability of being the victim of a home burglary at least once in the next year (Dominitz and Manski, 1997). Each of these biases

* Correspondence to: Baruch Fischhoff, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA. E-mail: baruch@cmu.edu

Contract grant sponsor: National Institute of Alcohol Abuse and Alcoholism
Contract grant sponsor: National Institute for Allergies and Infectious Diseases

CCC 0894–3257/99/020149–15$17.50

Accepted 1 September 1998

could add needlessly to individuals' personal worry and to the public demand for risk-reduction policies.*

In the studies just cited, people were found to overestimate relatively small probabilities. Substantive interpretations for these biases are easy to find. For example, the risks in question have all received extensive, and some might argue exaggerated, media attention (Combs and Slovic, 1979; Singer and Endreny, 1994). It would be only natural for people to infer unduly high probabilities, unless they were unusually alert to the possibility of disproportionate reportage (Lichtenstein *et al*., 1978; Tversky and Kahneman, 1973).

Nonetheless, when a pattern is observed across diverse domains, one must ask whether some common methodological feature is at least partly responsible. How people translate subjective feelings into quantitative terms is one of the oldest topics in psychology. Students of psychophysics have documented a variety of potentially relevant biases. For example, drawing people's attention to a quantitative value, however irrelevant to the task, can pull their responses in that direction (Poulton, 1989, 1994; Tversky and Kahneman, 1974). Thus, a high 'anchor' could artifactually increase probability judgments.

Because they were designed for telephone administration, most of these studies showing the '50 blip' all asked respondents to fill in a blank. As a result, they offered no obvious anchor. Nor is there evidence of other common psychophysical biases, such as respondents restricting themselves to a few stereotypical values or showing reluctance to use the extremes of the response scale.

However, in several cases in which studies reported response distributions, inspection of the data revealed a seemingly strange 'blip' at 50. To some extent, this blip might reflect insensitivity to probabilities in the middle range (Kahneman and Tversky, 1979), leading respondents to use 50 for, say, everything between 40 and 60. Yet, even if the 50 responses are spread over this range, a lump remains. Possibly, it could reflect real differences in beliefs. For example, a portion of the population might see much higher risks than the rest, with their estimates coincidentally falling at 50 in each case.

Another possibility is that, for some respondents, 50 means something different than a probability. One possible alternative meaning has emerged in recent interviews regarding the risk of AIDS, which we have conducted with a diverse sample of teens (Fischhoff and Downs, 1997; Fischhoff, Downs and Bruine de Bruin, 1998; Quadrel, Fischhoff and Davis, 1993). Often, when asked about risks, these teens would answer 'fifty–fifty', in a way that suggested, not a probability, but an expression for 'I really don't know'. Exhibit 1 shows several examples. Such usage might be called *Laplacian*, placing equal probability on all hypotheses — in cases where there are two possible outcomes (i.e. something might or might not happen). Or it might be called *epistemic uncertainty*, reflecting not knowing what probabilities to use to express one's state of belief (Gärdenfors and Sahlin, 1982).

Using 'fifty–fifty' to express something other than 50% would artifactually increase aggregate judgments for events typically assigned smaller values. This would be especially true when judged probabilities are otherwise very small. Then, even a modest proportion of such responses could substantially increase mean judgments. For example, eliminating all responses of 50 in the studies cited above would reduce group means by 3–7 percentage points, from the full-sample means of about 15%. Thus, 50s represent a significant proportion of the apparent overestimation of risk.

---

* Evaluating the accuracy of judgments requires sometimes controversial calculations of 'objective' probabilities. For example, Dominitz and Manski (1997) use (with many explicit qualifications) respondents' current health insurance status as a rough approximation of the chances of not having insurance one year hence. Among their respondents, the mean estimate for being a victim of burglary in the next year was 16%, while only 5% of the sample actually became victims of this crime the next year. For such comparisons to be possible at all, assessments must be interpretable in quantitative terms. Thus, one would not avoid the problems discussed here by asking respondents to choose among verbal quantifiers, such as 'very likely' and 'rarely'. These terms have long been known to have different quantitative referents for different people and even for the same person in different contexts (e.g. 'very likely to rain' may connote a much higher number than 'very likely to kill you') (Budescu and Wallsten, 1995).

Exhibit 1. Some adolescents' uses of fifty–fifty

Interviewer:	OK. And how well do you think that this would really protect a person from getting HIV, if they used a new needle?
Respondent:	I wouldn't say 100% sure, I would say almost definite, but there is always that risk if you are involved, I think with that kind of life style, you know, a drug use has, ah, you are eventually going to run into that. So I say, a fifty–fifty chance.
Interviewer:	OK. Hum, OK, so you said the more people that a person has unprotected sex with [the greater the risk]. What if they have unprotected sex with the same person? Does it matter how many times?
Respondent:	It only takes once if the person is infected. It could; it could not. So I guess it's up in the air again, a fifty–fifty chance.
Interviewer:	How well do you think that would really protect a person from getting the AIDS virus, if they use a condom when they are having sex?
Respondent:	I guess it's a fifty–fifty chance, 'cause it might like break or something and then it might not.
Interviewer:	How well do you think [that using condoms] would really protect a person from getting AIDS?
Respondent:	I'd say about . . . I don't know, fifty–fifty, 'cause there's always that possibility that something like the condom would break or something.

One possible contributor to such usage is the open-ended format of these studies, which requires respondents to generate their own probability responses. This procedure may increase the chances of expressing epistemic uncertainty by means of 50, relative to a procedure providing explicit response options. The present studies examine the extent to which an open-ended response mode encourages using 50 as a verbal phrase (fifty–fifty), rather than as a number (50). The studies use different response modes that vary the availability of response options, under the assumption that 'fifty–fifty' becomes less likely when explicit numerical response alternatives are made available. In Study 1, one form had respondents fill in a blank, akin to previous studies. The second form offered a 0–100 probability scale, with 101 tick marks. We expected a smaller proportion of 50 responses with the scale. Half of the respondents receiving each response mode were also given the option of answering 'absolutely no idea', instead of giving a numerical probability. If the 50 response is used to express uncertainty about which number to use, then providing an 'absolutely no idea' option might reduce the blip.

STUDY 1

Method

Procedure

In all versions of the survey, respondents were asked four questions concerning low-probability events, namely (a) 'What is the probability of the university being closed at least once this year because of a bomb threat?' (b) 'What is the probability of someone getting AIDS if they have sex without protection?' (c) 'What is your personal probability of developing cancer by age 80?' (d) 'What is the probability that someone will break into your room or home and steal something this year?' These events were chosen to parallel those in earlier studies, and to provide a variety of events with small statistical probabilities.

Respondents in the *blank* condition were instructed to fill in a space with 'a number between 0% (no chance) and 100% (certainty)'. In the *scale* condition, the probability scale had 101 tick marks, with endpoints labeled '0 = no chance' and '100 = certainty', and additional numbers (as well as slightly larger ticks) at 10, 20, . . . , 80, and 90. The response modes were presented without instructions, lest our words communicate something about epistemic uncertainty.

Approximately half of the participants in the blank and scale conditions were also offered the opportunity to state that they had 'absolutely no idea' regarding each probability. Thus, there were four conditions, differing in whether a scale was provided and whether there was an 'absolutely no idea' option.

After completing the four probability judgments, respondents were asked: (a) how confident they were in each judgment, on a scale anchored at 1 = very confident and 7 = not confident at all, and (b) how clear each probability question had been, with 1 = very clear and 7 = not at all clear. We were interested in respondents' satisfaction with the different response modes. These questions were placed after the four probability judgments, in order to avoid suggesting epistemic uncertainty.

Participants

Participants were asked to fill out a small packet of unrelated questionnaires, including the probability surveys reported here. Seventy students at the University of Pittsburgh and 48 students at Carnegie Mellon University were given the scale version of the questionnaire, half of whom received the 'absolutely no idea' option. An additional 28 aquarium owners at a meeting of Pitt Fish and 93 visitors to the spring carnival at Carnegie Mellon University received the blank condition. Again, half received the 'absolutely no idea' option.* Subjects recruited at the University of Pittsburgh and CMU received a candy bar for their participation, while those visiting the spring carnival were given two ride tickets. The aquarium club received $3 for each participating member. A total of 132 men and 100 women participated in this study; an additional seven respondents did not specify their gender. The average age was 23.4 (sd = 6.0). Mean years of education was 15.0 (sd = 2.2).

Results

Absolutely no idea

Thirty-three of the 120 respondents (=27.5%) in the 'absolutely no idea' condition checked that answer for at least one of the four probability questions, amounting to 53 occasions (or 11% of all responses). Most of the respondents who employed the 'absolutely no idea' option used it for one or two answers only, suggesting that they were not simply minimizing their effort. Respondents used 'absolutely no idea' almost twice as frequently when no scale was presented (13.9% versus 7.9%). However, the related chi-square tests of homogeneity did not reach significance. Cochran's test showed a question effect for the use of 'absolutely no idea' ($\chi(3) = 11.02$, $p < 0.05$), with the bomb-at-university question having as many as all the others combined. For that question, responses of 50 were less common with the 'absolutely no idea' option than without it (6.8% and 1.8%, respectively). For each question, a Kolmogorov–Smirnov test found no difference in the distributions of probability

* Initially, we ran just the scale response mode, thinking that adding the 'absolutely no idea' option would eliminate the 50 blip. When we observed few 50 responses, we realized that the earlier studies had not used a scale, but open-ended questions. As a result, we added a blank response mode. For the sake of brevity, we have combined results with the two response modes here in Study 1, rather than lead the reader through our path of discovery (and having two studies with two versions of one response mode each). This sequential process meant that questionnaires were randomly assigned with respect to the 'absolutely no idea' option, but not with respect to the response mode. We sought similar populations. Subjects receiving the two response modes were similar in age and gender. Those receiving the scale reported two additional years of education (15.9 versus 14.0; $p < 0.001$). We can see no obvious reason why these groups would differ in their use of 50; moreover, similar patterns were observed in the Study 2, with different populations (and random assignment). Nonetheless, this is an imperfect part of our procedure, which we addressed in Study 2.

Exhibit 2. Probability responses in Study 1

Question	Percentage of 50 responses (excluding 'absolutely no idea') Blank conditions	Scale conditions	Chi-square test	Median response	Statistical estimate
AIDS from sex	14.9	13.0	0.17	20	5[a]
Bomb at the university	5.6	2.7	1.13	5	0–9[b]
Break-in	13.0	3.5	6.94**	10	1.07[c]
Cancer by 80	15.9	5.3	6.83**	45	31–37[d]
Average	12.5	6.2	5.92*	NM	NM

*$p < 0.05$.
**$p < 0.01$.
[a]Hearst and Hulley (1988).
[b]In the past 23 years, two buildings at CMU were evacuated because of a bomb threat. The entire school was never closed for this reason (Sgt Simon, personal communication, 29 May 1998).
[c]*Commonwealth of Pennsylvania Uniform Crime Report* (1996); *Pennsylvania Abstract. A Statistical Fact Book* (1996).
[d]Seidman *et al.* (1985).
NM = Not meaningful.

responses with and without the 'absolutely no idea' option ($\alpha = 0.05$). Neither did median tests ($\alpha = 0.05$). As a result, whether the 'absolutely no idea' option was offered is ignored in subsequent analyses.

Use of 50

Exhibit 2 shows the percentage of 50s among the numerical judgments (i.e. excluding 'absolutely no idea' responses). Overall, twice as many 50 responses were provided with the blank as with the scale condition, with significant differences for the break-in and cancer-by-80 questions. A Kruskal–Wallis test, in which individuals were characterized by their total number of 50s, showed significantly more 50 responses when no scale was presented ($\chi(1) = 5.92$, $p < 0.05$). Exhibit 3 depicts the response distributions for the break-in question. (Kendall's test of concordance showed that the number of 50 responses differed by question ($\chi(3) = 16.85$, $p < 0.001$), which could reflect differences in both epistemic uncertainty and overall judged likelihood.)

Overall responses

A Friedman test showed that probability answers differed by question ($\chi(3) = 185.4$, $p < 0.001$). Probabilities were significantly higher in the blank than in the scale condition for the AIDS-from-sex question ($\chi(2) = 5.97$, $p < 0.05$).

Confidence and clarity

Perceptions of confidence and clarity showed positive correlations for each question, but reached significance only for the bomb-at-the-university question ($r = 0.18$, $p < 0.01$). A MANOVA found that confidence ratings were related to question ($F(3, 657) = 9.87$; $p < 0.001$), but not to response mode or to the presence of the 'absolutely no idea' option. Thus, respondents reported being equally comfortable with the blank and the scale conditions. Across questions, mean confidence ranged from 3.25 (for the bomb-at-the-university question) to 4.08 (for the AIDS-from-sex question) (on the 1–7 point scale). Clarity judgments similarly showed an effect for question ($F(3, 663) = 13.31$; $p < 0.001$), but not for response mode. However, an interaction between response mode and question showed that

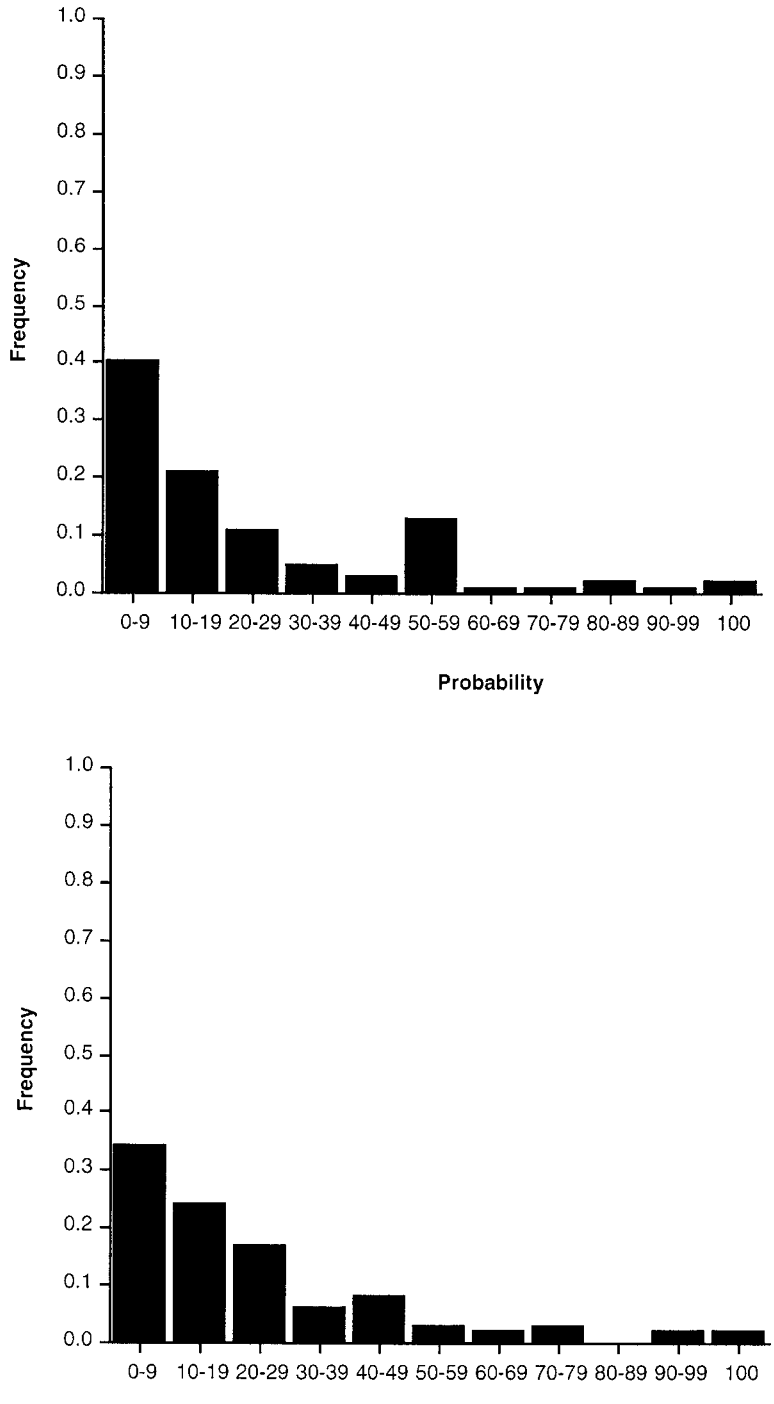

Exhibit 3. Distribution of probabilities given in response to the Study 1 question: 'What is the probability that somebody will break into your room or house and steal something sometime this year?' in the (a) blank and (b) scale conditions.

two of the questions were rated as less clear in the scale condition (AIDS from sex and break-in; $F(3, 663) = 2.90$, $p < 0.05$). For three of the four questions, clarity and confidence were unrelated to the use of 50. Kendall's tau was weakly significant for the bomb-at-the-university question, showing more 50 responses with reports of higher question clarity ($r = 0.15$, $p < 0.05$) and higher confidence ($r = 0.15$, $p < 0.05$).

Group differences

Kendall's tau showed no correlation between the use of 50 and age or gender. Higher grade levels were associated with less frequent use of 50 for one question, break-in ($r = -0.17$, $p < 0.01$). The number of 50s was unrelated to whether respondents reported having taken a class treating probabilities.

Discussion

Offering the 'absolutely no idea' option slightly decreased the use of 50 responses overall, with the difference reaching significance for one individual question. Thus, forcing uncertain respondents to give numerical probabilities, when they would rather not, may contribute to the 50 blip. The frequency with which respondents chose 'absolutely no idea' may understate the frequency with which they felt that subjective state. That response involves an admission of ignorance. Respondents who find that socially undesirable may prefer to say 'fifty–fifty', because that is at least a numerical response.

The blank condition generated twice as many 50 responses as did the linear probability scale. Thus, offering respondents an explicit set of numerical responses may have focused their attention on the numerical interpretation of probabilities, and diminished the availability of the verbal expression 'fifty–fifty'. Respondents reported similar confidence in their answers when using the probability scale and when producing their own probabilities. Two questions were judged less clear in the scale condition, perhaps because the precision implied by the 101-tick scale encouraged respondents to think harder. However, these perceptions of increased ambiguity in the questions did not contribute to any epistemic uncertainty, as expressed by using 50 as an answer.

Overall, respondents overestimated the risks addressed in Study 1, compared with our statistical estimates (Exhibit 2). Estimates for the AIDS-from-sex question were higher with the blank condition, where the increased use of 50 also contributed to higher median probabilities.

Statistical power is naturally limited when observing changes in the rate of a relatively infrequent event (like the use of 50). As a result, a second study was conducted in order to examine the robustness of these results, using different subjects and items. A second probability scale was also added, in order to see the effect of offering a larger set of response options. This scale expands the 0–1% section of the linear scale form, providing additional low-probability response options (in a log format). We expected an even lower proportion of 50 responses with this log-linear scale. Because the 'absolutely no idea' manipulation proved less powerful in Study 1, we dropped it in order to concentrate on the effect of different response modes.

STUDY 2

Method

Procedure

In all three versions of the questionnaire, respondents assessed the probabilities of six events, in the following order: (a) [your] getting struck by lightning some time this year, (b) [your] developing cancer by age 40, (c) someone who smokes a pack or more of cigarettes a day developing lung cancer, (d) someone getting AIDS if they have sex without protection once with someone who is infected, (e) an average woman being diagnosed with breast cancer in her lifetime, and (f) [your] being alive at age 50. These events were chosen to parallel those in earlier studies, to provide a variety of events with small statistical probabilities, and to include at least one high-probability event (f). The events were specified with enough detail to allow answers to be compared to statistical estimates (Fischhoff, 1994; Fischhoff, Bostrom and Quadrel, 1996).

Respondents in the *blank* condition were instructed to fill in a space with 'a number between 0% (no chance) and 100% (certainty)'. The *linear* scale condition was the same as that in Study 1. The *log-linear* scale condition used the linear scale, but added a six-order log scale expanding upon the probability options between 0 and 1%. The response modes were, again, presented without additional instruction.*

After completing the six probability judgments, respondents evaluated each item on the confidence and clarity scales of Study 1.

Participants

An experimenter approached people walking on the UC Berkeley campus and asked them to complete two unrelated questionnaires, in return for $3. The present study was the first of these tasks. A total of 102 men and 58 women participated, with three more not specifying their gender. Respondents reported a mean age of 23.4 (sd $=$ 5.8) and mean years of eduction of 14.3 (sd $=$ 1.9), both very similar to Study 1. Respondents were randomly assigned to one of the three response modes.

Results

Use of 50

Exhibit 4 shows the percentages of 50 responses. As in Study 1, they were much more common in the blank condition than with a numerical scale. These differences were statistically significant for three

Exhibit 4. Percentages of 50 responses in Study 2

Question	Blank	Linear	Log-linear	Total
Lightning	1.9	0.0	0.0	0.7
Cancer by 40	20.4	7.4	10.3	12.9
Lung cancer	15.1	7.4	15.8	12.4
AIDS from sex	23.1	7.4	12.1	14.4
Breast cancer	24.1	7.4	2.6	12.3
Alive at 50	17.3	9.3	12.2	12.9
Average	17.0	6.5	8.8	10.9

* In the data analyses, probabilities marked between the six major tick marks on the log portion of the log-linear scale were rounded up to the nearest tick mark. Thirteen of 55 respondents in the log-linear scale condition used both parts of the scale for every single question, while six additional respondents did so at least once. These respondents were excluded from the analyses. Given our previous success with this response scale, we assume that these difficulties reflect the lack of instructions, omitted in order to equate the conditions more fully and to avoid any wording that might suggest any attitude toward epistemic uncertainty. In the past, brief instructions with one or two test questions have been sufficient to allow even high-risk (low-literacy) teens to use the log-linear scale without apparent trouble (Linville, Fischer and Fischhoff, 1993; Quadrel *et al.*, 1993). Schwartz *et al.* (1998) report a 71% survey response rate and a 98% item response rate with a random national sample of women on a form making extensive use of a log-linear scale. Log-odds scales were used in Bayesian inference 'bookbag and poker chip' studies, producing some of the most accurate responses (Slovic and Lichtenstein, 1971).

individual questions: AIDS from sex ($\chi(1) = 5.09$, $p < 0.05$), cancer by 40 ($\chi(1) = 4.20$, $p < 0.05$), and breast cancer ($\chi(1) = 10.94$, $p < 0.001$). The non-significant differences for the other three questions were in the same direction. After characterizing subjects by their number of 50s, Kruskal–Wallis tests showed that use of 50 depended on the response mode ($\chi(2) = 17.14$, $p < 0.001$), with respondents giving more 50 responses in the blank condition than with the linear scale ($\chi(1) = 15.65$, $p < 0.001$) or the log-linear scale ($\chi(1) = 6.75$, $p < 0.01$). There was no appreciable difference between the two probability scales in the use of 50, even though the log-linear condition offered many more explicit response options than did the linear one. Kendall's test of concordance showed that the number of 50 responses differed by question ($\chi(5) = 19.74$, $p < 0.01$). The AIDS-from-sex question received the highest number of 50s, while the struck-by-lightning question elicited almost no 50s at all. Exhibit 5 depicts distributions for one low-probability event, cancer by 40. Visual inspection of response distributions for the blank condition shows marked blips at 50 for every event, except lightning.

Overall responses

Exhibit 6 shows mean and median responses for each question and response mode. The disparity between these two measures of central tendency shows the skew in the underlying distributions. It is relatively small, except for several of the lowest probabilities. A Friedman test showed, not surprisingly, that responses differed across question ($\chi(5) = 430.67$, $p < 0.001$). Kruskal–Wallis tests comparing judgments in the three conditions reached significance for three items (cancer by 40, lung cancer from smoking, and breast cancer). A median test was also significant for these three items, as well as for a fourth ($\chi(2) = 7.9$, $p < 0.05$ for cancer by 40; $\chi(2) = 17.65$, $p < 0.001$ for lung cancer from smoking; $\chi(2) = 18.67$, $p < 0.001$ for breast cancer; and $\chi(2) = 22.99$, $p < 0.001$ for struck by lightning). The

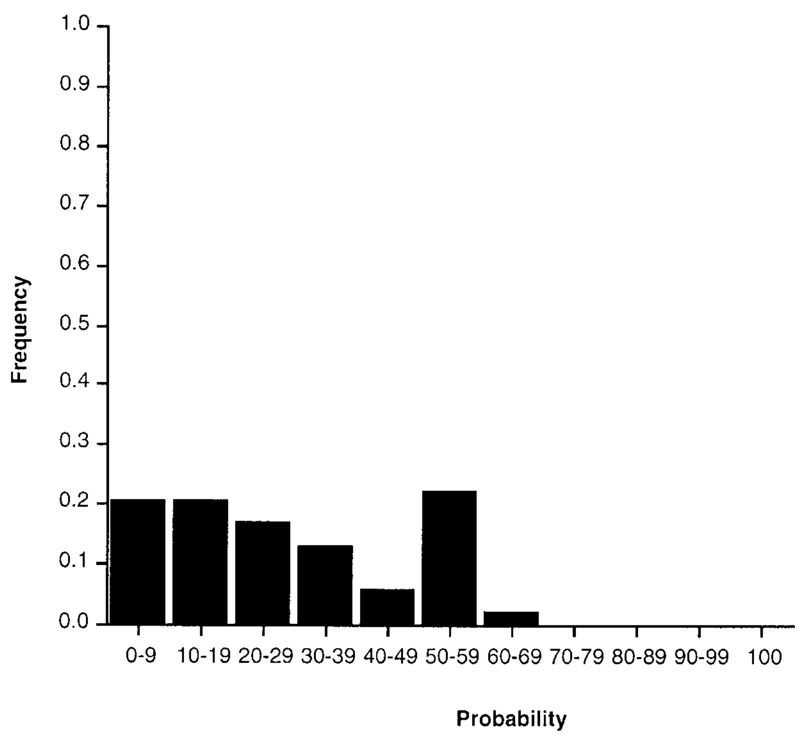

Exhibit 5(a)

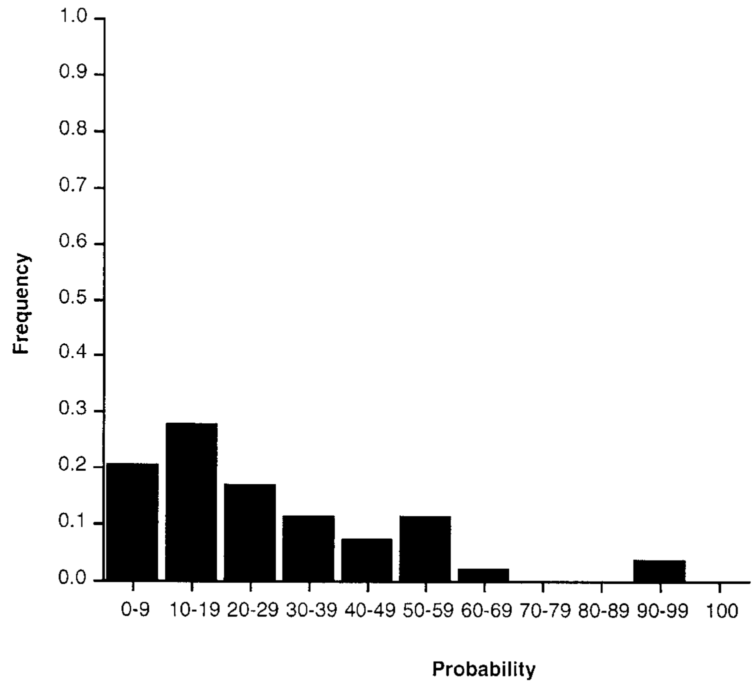

Exhibit 5(b)

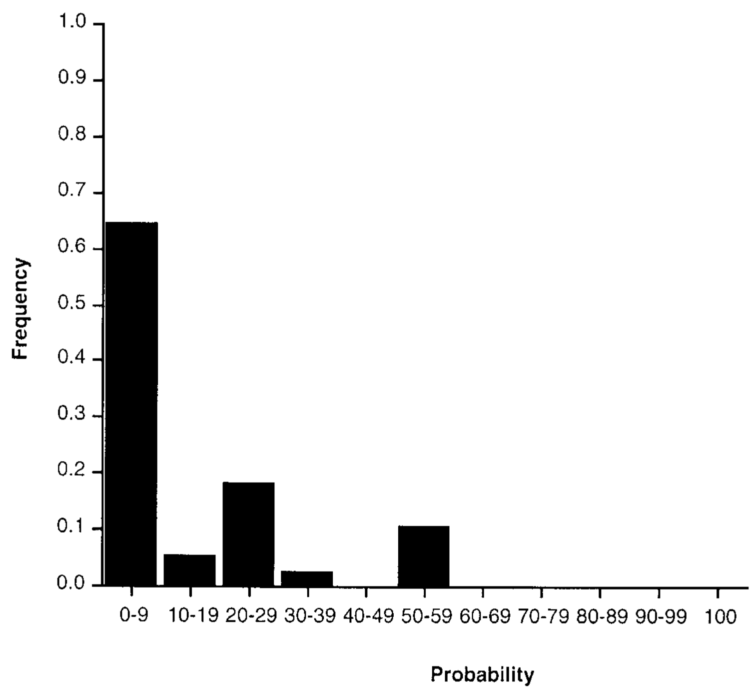

Exhibit 5(c)

Exhibit 5. Distribution of probabilities given in response to the question: 'What is the probability that you will develop cancer by 40?' in the (a) blank, (b) linear, and (c) log-linear conditions.

Exhibit 6. Probability responses in Study 2

Question	Blank Mean	Blank Median	Linear Mean	Linear Median	Log-linear Mean	Log-linear Median	Test K–W $\chi(2)$	Statistical estimates
Lightning	3.1	1.0	3.3	1.0	0.8	0.0[a]	4.43	0.000001[a]
Cancer by 40	25.2	20.0	24.5	20.0	11.4	2.0	18.52*	1–2[b]
Lung cancer	58.7	60.0	54.9	60.0	29.1	16.0	23.61*	10[c]
AIDS from sex	57.2	50.0	50.2	50.0	42.6	50.0	3.99	5[d]
Breast cancer	37.2	40.0	33.3	30.0	18.9	15.0	19.60	10[e]
Alive at 50	74.7	80.0	73.2	79.5	74.0	80.0	0.34	94.4[f]

*$p < 0.01$.
[a]10^{-6}; *Natural Hazard Fatalities for 1994 in the United States* (1995).
[b]Seidman *et al.* (1985).
[c]Viscusi (1992).
[d]Hearst and Hulley (1988).
[e]Black, Nease and Tosteson (1995).
[f]*Vital Statistics of the United States, 1992* (1993).

two questions without significant response-mode effects were the ones with the largest probability responses: AIDS from sex and being alive at 50. Generally speaking, the blank condition elicited the highest probabilities, followed by the linear and log-linear conditions.*

Confidence and clarity

All questions showed a positive correlation between perceived confidence and clarity. The correlations were significant for the lung-cancer-from-smoking ($r = 0.18$, $p < 0.05$) and AIDS-from-sex ($r = 0.27$, $p < 0.01$) questions. As in Study 1, a MANOVA found that confidence ratings were related to question ($F(5, 600) = 13.16$; $p < 0.001$), but not to response mode. Thus, respondents were equally comfortable with the (possibly more demanding) linear and log-linear scales. Across questions, mean confidence ranged from 2.60 (for lightning) to 3.92 (for breast cancer). The question-by-response-mode interaction ($F(10, 600) = 3.01$; $p < 0.01$) had no obvious interpretation. Clarity judgments similarly showed a question effect ($F(5, 575) = 5.39$; $p < 0.001$), with no effect of response mode or interaction. There was no correlation (Kendall's tau) between clarity or confidence judgments and the use of 50.

Group differences

For each question, use of 50 was unrelated to gender or to whether respondents had taken at least one class covering probabilities. Kendall's tau showed that the number of 50s was modestly correlated with overall level of education for one question (AIDS from sex; $r = -0.21$, $p < 0.05$), and with age for one question (cancer by 40; $r = -0.21$; $p < 0.05$).

Discussion

Study 2 replicated the findings of Study 1. Respondents who had to produce their own probability answers, in the blank condition, were twice as likely to use the 50 response as were the respondents offered a probability scale. Adding the log response options to the linear scale did not further reduce the use of 50. These results are consistent with the hypothesis that presenting a probability scale reduces

* The respective means with and without the 50s are, following the order of the events in the exhibits (3.1, 2.2; 25.2, 18.9; 58.7, 60.3; 57.2, 59.4; 37.2, 33.1; 74.7, 79.9).

either respondents' epistemic uncertainty or their ability to express it. Also consistent is the finding, in Study 1, that the probability scale halved respondents' use of the 'absolutely no idea' option. Within the range of variation afforded by our samples, none of these tendencies were related to reported gender, age, years of education, and instruction in probability. Epistemic uncertainty, if that is indeed the causal factor here, appears to be a general experience that occurs when individuals encounter particular items, especially when they need to produce their own numbers.

In both studies, use of 50 depended on the question's content in some, as yet to be determined, way. Examination of item differences might yield some further insight into the nature of these processes. Obviously, there were more 50s with distributions centered at 50, making detection of a blip more difficult. Less obvious is the absence of any blip with the struck-by-lightning question. Arguably, this is an event that epitomizes low probability, without evoking a feeling of 'maybe yes, maybe no'. The other events vary in so many ways that it will take further research to discipline the possible speculations regarding the feelings that they evoke. One potentially relevant feature is that all our questions involved risks to life and limb. With such fateful events, people may be particularly reluctant to commit themselves to a number when they don't really know what to say.

The frequency of 50s was unrelated to the questions' rated clarity. Thus, being less sure about what was being asked did not seem to encourage using 50 as an expression of not knowing what to say. Perhaps people are so accustomed to responding to the gist of questions that ambiguity tends not to generate epistemic uncertainty (Schwarz, 1996). If so, then epistemic uncertainty would seem to emerge when respondents are deciding what to answer, rather than when trying to figure out what they are being asked.

Wallsten, Budescu and Zwick (1993) also observed a 50 blip among respondents using an open-ended numerical response mode, in a study comparing the calibration of those judgments with ones elicited with a list of verbal probability expressions having 'toss up' at its midpoint. They suggest that uncertain respondents might answer 50 when they feel that it will be an equally defensible answer whatever actually happens. That response strategy could, it seems, be either an expression of epistemic uncertainty or a deliberate misrepresentation of another belief. If the latter, then it need not be related to response mode. Wallsten *et al*. also argued that the inherent vagueness of verbal probability terms means that they naturally incorporate an expression of epistemic uncertainty. In this light, it would be interesting to repeat their study with a list of verbal probability questions, whose midpoint is labeled 'fifty–fifty'.

It appears, thus, that the availability of the common phrase 'fifty–fifty' affects the use of the 50 response as an expression of epistemic uncertainty. The presentation of a probability scale draws respondents' attention away from such verbal answers and focuses them on probabilities as numbers. The open-ended response mode may also have increased epistemic uncertainty by requiring respondents to produce their probability answers, a task which is more difficult than simply placing a mark on a probability scale.

Whatever its source(s), reduced reliance on 50 leads to smaller estimates and, hence, to lower overestimates of the small probabilities.* As seen in Exhibit 6, responses with the log-linear scale were actually fairly accurate. However, this improvement could itself also have an element of artifact. The explicit options of the probability scales may have suggested response options that would not have occurred to respondents naturally (Fischhoff, Slovic and Lichtenstein, 1980; Poulton, 1989, 1994; Schwartz and Hippler, 1987). One reasonable expectation is that expanding the lower end of the log-

* Of course, not all feelings of epistemic uncertainty need emerge as 50. People may use 'fifty–fifty' as an anchor for their estimation process, adjusting that starting point, on the basis of whatever beliefs they have, beyond their initial feelings of not knowing what to say. If such adjustment is insufficient (Tversky and Kahneman, 1974), then such an anchor would bias judgments of small probabilities upward, without appearing as a 50 blip.

linear scale encouraged responses between 0 and 1% — although it was not a powerful enough influence to affect the responses to the AIDS-from-sex and alive-by-50 questions, with their high subjective probabilities. Conversely, the linear scale may have been biased against the expression of very small values. For its part, the open-ended response mode may have biased respondents toward fifty–fifty, or the middle, relative to the linear scale. As elsewhere, how accurate people seem to be depends, in part, on how they are asked to answer (Fischhoff and MacGregor, 1983). One clear exception was with the AIDS-from-sex question. The median response of 50 suggests either a robust overestimate of this risk or resolute epistemic uncertainty, which even numerical scales cannot accommodate.*

Understanding individuals' risk perceptions is essential to predicting their behavior and to providing them with needed information. Doing so requires studies with well-chosen response modes and well-formulated questions. The studies reported here explore one obstacle to accurate assessment, as well as a possible remedy. By and large, judgments with a numerical scale showed a smaller 50 blip, without the reduction in confidence that more demanding scales might be suspected of inducing. One part of the acceptability of a response mode is how it evokes and deals with respondents' feelings of epistemic uncertainty. Eliminating anomalous 50 blips seems good, as long as we have confidence that those feelings are not channeled to values suggested by other features of a structured response mode. Increasing that confidence will require further study of both the conditions inducing epistemic uncertainty and the roles played by the response '50'.

ACKNOWLEDGMENTS

Support for this research was provided by the National Institute of Alcohol Abuse and Alcoholism and the National Institute for Allergies and Infectious Diseases, and is gratefully acknowledged. The opinions expressed are those of the authors. We also appreciate the support of Susy Bielak, Stephanie Byram, Robyn Dawes, Julie Downs, Shane Frederick, Melanie Gold, Jenny Hounshell, Liz Kanabe, Chuck Manski, Claire Palmgren, Andy Parker, David Rubovits, and Andrew White in various stages of this project, as well as the comments of three anonymous reviewers.

REFERENCES

Black, W. C., Nease, R. F. and Tosteson, A. N. A. 'Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age', *Journal of the National Cancer Institute*, **8** (1995) 720–731.

Budescu, D. F. and Wallsten, T. S. 'Processing linguistic probabilities: General principles and empirical evidence', in Busemeyer, J. R., Hastie, R. and Medin, D. L. (eds), *Decision Making from a Cognitive Perspective* (pp. 275–318), New York: Academic Press, 1995.

Combs, B. and Slovic, P. 'Newspaper coverage of causes of death', *Journalism Quarterly*, **56** (1979), 832–849.

Commonwealth of Pennsylvania Uniform Crime Report, Pennsylvania: Pennsylvania State Police, 1996.

Dominitz, J. and Manski, C. F. 'Perceptions of economic insecurity. Evidence from the survey of economic expectations', *Public Opinion Quarterly*, **61** (1997) 261–287.

Fischhoff, B. 'What forecasts (seem to) mean', *International Journal of Forecasting*, **10** (1994), 387–403.

Fischhoff, B., Bostrom, A. and Quadrel, M. J. 'Risk perception and communication', in Detels, R., McEwen, J. and Omenn, G. (eds), *Oxford Textbook of Public Health* (pp. 987–1002), London: Oxford University Press, 1996.

Fischhoff, B. and Downs, J. 'Accentuate the relevant', *Psychological Science*, **8**(3) (1997), 1–5.

Fischhoff, B., Downs, J. and Bruine de Bruin, W. 'Adolescent vulnerability: A framework for behavioral interventions', *Applied and Preventive Psychology*, **7** (1998), 77–94.

* Linville *et al.* (1993) found similar estimates with an alternative log-linear scale, and discuss possible sources of this bias.

Fischhoff, B. and MacGregor, D. 'Judged lethality: How much people seem to know depends upon how they are asked', *Risk Analysis*, **3** (1983), 229–236.
Fischhoff, B., Slovic, P. and Lichtenstein, S. 'Knowing what you want: Measuring labile values', in Wallsten, T. (ed.), *Cognitive Processes in Choice and Decision Behavior* (pp. 117–141), Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.
Freudenberg, W. 'Perceived risk, real risk', *Science*, **241** (1988), 44–49.
Gärdenfors, P. and Sahlin, N. E. 'Unreliable probabilities, risk taking and decision making', *Synthèse*, **53** (1982), 361–386.
Hearst, N. and Hulley, S. B. 'Preventing the heterosexual spread of AIDS: Are we giving our patients the best advice?' *Journal of the American Medical Association*, **259** (1988), 2428–2432.
Kahneman, D. and Tversky, A. 'Prospect theory: An analysis of decision under risk', *Econometrica*, **47** (1979), 263–281.
Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M. and Combs, B. 'Judged frequency of lethal events', *Journal of Experimental Psychology: Human Learning and Memory*, **4** (1978), 551–578.
Linville, P. W., Fischer, G. W. and Fischhoff, B. 'AIDS risk perceptions and decision biases', in Pryor, J. B. and Reeder, G. D. (eds), *The Social Psychology of HIV Infection* (pp. 5–38), Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.
National Research Council. *Improving Risk Communication*, Washington, DC: National Academy Press, 1989.
Natural Hazard Fatalities for 1994 in the United States, Washington, DC: National Oceanic and Atmospheric Administration, 1995.
Quadrel, M. J., Fischhoff, B. and Davis, W. 'Adolescent (in)vulnerability', *American Psychologist*, **48** (1993), 102–116.
Pennsylvania Abstract. A Statistical Fact Book, Harrisburg, PA: Pennsylvania State Data Center, 1996.
Poulton, E. C. *Bias in Quantifying Judgment*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1989.
Poulton, E. C. *Behavioral Decision Theory: A new approach*, New York: Cambridge University Press, 1994.
Schwartz, L. M., Woloshin, S., Sox, H. C., Fischhoff, B. and Welch, H. G. *Perceptions of American Women about Screening Mammography: A national survey*, under editorial review, 1998.
Schwarz, N. *Cognition and Communication. Judgmental biases, research methods, and the logic of conversation*, Mahwah, NJ: Lawrence Erlbaum Associates, 1996.
Schwarz, N. and Hippler, H. J. 'What response scales may tell your respondents: Informative functions of response alternatives', in Hippler, H. J., Schwarz, N. and Sudman, S. (eds), *Social Information Processing and Survey Methodology*, New York: Springer-Verlag, 1987.
Seidman, H., Mushinski, M. H., Gelb, S. K. and Silverberg, E. 'Probabilities of eventually dying of cancer — United States, 1985', *CA, A Cancer Journal for Clinicians*, **35**(1) (1985), 36–56.
Singer, E. and Endreny, P. M. *Reporting on Risk*, New York: Russell Sage Foundation, 1994.
Slovic, P. 'Perceptions of risk', *Science*, **236** (1987), 280–285.
Slovic, P. and Lichtenstein, S. 'Comparison of Bayesian and regression approaches to the study of information processing in judgments', *Organizational Behavior and Human Performance*, **6** (1971), 649–744.
Tversky, A. and Kahneman, D. 'Availability: A heuristic for judging frequency and probability', *Cognitive Psychology*, **4** (1973), 207–232.
Tversky, A. and Kahneman, D. 'Judgment under uncertainty: Heuristics and biases', *Science*, **185** (1974), 1124–1131.
Viscusi, W. K. *The Risks of Smoking*, Cambridge, MA: Harvard University Press, 1992.
Vital Statistics of the United States 1992, Centers for Disease Control and Prevention, National Center for Health Statistics, Vol 2, Section 6, 1993.
Wallsten, T. S., Budescu, D. V. and Zwick, R. 'Comparing the calibration and coherence of numerical and verbal probability judgments', *Management Science*, **39** (1993), 176–190.

Authors' biographies:
Baruch Fischhoff is University Professor, in the Department of Social and Decision Sciences and of Engineering and Public Policy, at Carnegie Mellon University. His current research interests include value elicitation, risk prioritization, integrated assessment, adolescent decision making, and risk communication.

Wändi Bruine de Bruin is a PhD candidate in the Department of Social and Decision Sciences, Carnegie Mellon University. She has a Masters degree in cognitive psychology from the Free University of Amsterdam. Her research interests include knowledge elicitation, subjective probabilities, risk communication, and adolescent decision making.

Authors' address:
Baruch Fischhoff and **Wändi Bruine de Bruin**, Department of Social and Decision Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213-3890, USA.