Risk Analysis, Vol. 22, No. 4, 2002

What Number is "Fifty-Fifty"?: Redistributing Excessive 50% Responses in Elicited Probabilities

Wändi Bruine de Bruin,[1]* Paul S. Fischbeck,[2,3] Neil A. Stiber,[4] and Baruch Fischhoff[2,3]

Studies using open-ended response modes to elicit probabilistic beliefs have sometimes found an elevated frequency (or blip) at 50 in their response distributions. Our previous research[1–3] suggests that this is caused by intrusion of the phrase "fifty-fifty," which represents epistemic uncertainty, rather than a true numeric probability of 50%. Such inappropriate responses pose a problem for decision analysts and others relying on probabilistic judgments. Using an explicit numeric probability scale (ranging from 0–100%) reduces thinking about uncertain events in verbal terms like "fifty-fifty," and, with it, exaggerated use of the 50 response.[1,2] Here, we present two procedures for adjusting response distributions for data already collected with open-ended response modes and hence vulnerable to an exaggerated presence of 50%. Each procedure infers the prevalence of 50s had a numeric probability scale been used, then redistributes the excess. The two procedures are validated on some of our own existing data and then applied to judgments elicited from experts in groundwater pollution and bioremediation.

KEY WORDS: Risk perception; subjective probability; fifty-fifty; epistemic uncertainty; de-biasing

> The 50-50-90 rule: Any time you have a 50-50 chance of getting something right, there's a 90% probability you'll get it wrong.
>
> —Joke of the Day, February 16, 2001[4]

Probabilities are a standard way to describe situations of uncertainty and risk. Using probabilities appropriately is essential to formulating and communicating beliefs. Risk and decision analysts depend on explicit probability assessments to build models, predict the results of complex interactions among chance events, as well as creating decision trees, and identify the appropriate courses of action in uncertain situations.[5] These efforts are threatened when responses to probability questions reflect something other than probabilistic beliefs. In previous work, we have identified one potential anomaly: using 50 as shorthand for the verbal phrase "fifty-fifty," reflecting a feeling of epistemic uncertainty—or not knowing what number to use[6]—rather than the quantity 50%.[1–3] We initially observed such *nonnumeric* use of 50 in open-ended interviews.[1,3] When asked to explain their probability responses, respondents who say "50" are more likely to add "I don't know" or "either it happens or it doesn't" than those who use other numbers.[7] Failing to distinguish nonnumeric from numeric 50 responses would result in a "blip" of 50s relative to the rest of the response distribution. Indeed, an apparent excess of 50s has been observed in open-ended studies eliciting judgments about a wide variety of topics, including the risks of breast cancer, lung cancer (from smoking), and mortality.[8–10]

In addition to respondents' own reports about the meaning of their 50 responses, evidence of

1 Department of Technology Management, Eindhoven University of Technology.
2 Department of Social and Decision Sciences, Carnegie Mellon University.
3 Department of Engineering and Public Policy, Carnegie Mellon University.
4 URS Corporation.
* Address correspondence to Wändi Bruine de Bruin, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.

0272-4332/02/0800-0713$22.00/1

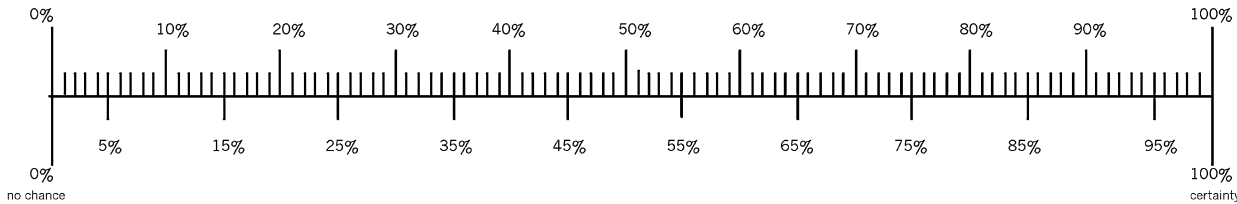

Fig. 1. The probability scale.

nonnumeric use comes from several covariates of the prevalence of 50s.(1–3) The use of 50 is related to response mode, question content, and respondent characteristics. For example, more 50s are observed with an open-ended response mode than with a probability scale, while leaving the rest of the response distribution relatively unchanged. An open-ended probability question asks respondents to mention a number between 0% and 100%, encouraging them to think about uncertainty in verbal terms, making "fifty-fifty" more likely to come to mind as an expression of epistemic uncertainty. A response scale that explicitly presents numeric options (such as shown in Fig. 1) diminishes the availability of verbal probability expressions, including "fifty-fifty," encouraging respondents to resolve their epistemic uncertainty. Consistent with this interpretation, an "absolutely no idea" response option, if offered, is used less with a probability scale than with an open-ended one.[1,3]

Furthermore, singular questions (e.g., "What is the probability that a Californian resident will die in the next earthquake?") elicit more 50s than equivalent distributional ones (e.g., "What is the percentage of Californian residents that will die in the next earthquake?"). The latter appears to evoke more analytical, and more numeric, thinking,[11–16] thereby reducing epistemic uncertainty and the use of nonnumeric 50s to express it.

Use of 50 increases with questions that could generate epistemic uncertainty, such as those addressing threatening topics. We found more 50s in response to probability questions about negative events with lower perceived control.[2,3] A national sample of teens produced particularly large 50 blips when assessing their risk of dying in the next year or before age 20.[17] A nonnumeric 50 provides an escape from contemplating such events, leaving any epistemic uncertainty unresolved, while still providing a number. Indeed, studies in survey design have found that questions covering threatening topics generally elicit more nonresponses, such as "don't know" answers.[18]

Finally, 50s are given more frequently by teens and lower-education adults—individuals who might understand probabilities less well.[1,3] Use of 50 is negatively correlated with numeracy, even after partialing out age and education.[1,3] Many of these 50s may not be intended as numeric.

These results suggest that surveys may be designed to discourage the nonnumeric use of 50 by using probability response scales (e.g., Fig. 1), as well as by providing an "absolutely no idea" option. Phrasing questions in distributional terms may also stimulate more numeric thinking. However, even with deliberate design, it may be impossible to eliminate nonnumeric 50s. Some unique events (e.g., about one's personal risk) may have no distributional equivalent. In addition, epistemic uncertainty may be unavoidable with studies of risk that deal with threatening and often uncontrollable events. To design education programs, the public's perception of these risks needs to be elicited—including those individuals with a poor understanding of probability and a tendency to say "50." Finally, previous surveys, unaware of this issue, have used open-ended response modes, which has clouded interpretation of their results. Future studies may be constrained to use an open-ended response mode (e.g., when administered by telephone) or choose to do so, in order to assess the extent of epistemic uncertainty.

When the nonnumeric use of 50 cannot be prevented, investigators need some way to distinguish between the two meanings of 50 lest they misinterpret respondents' intent. If some 50 responses do not represent the numeric value of 0.5, then it is misleading to treat them as such in statistical summaries of respondents' beliefs or in subsequent risk and decision analyses. For example, studies using open-ended questions have shown an excessive and striking use of 50s for low-probability events. Rather than overestimation, these 50 responses could reflect epistemic uncertainty. Such nonnumeric use of 50 may have contributed to the public's seeming overestimation of the probability of a smoker getting lung cancer,[8] cited by tobacco companies as evidence that smokers had not been led to underestimate these risks.

On the other hand, deleting all 50 responses from the analyses ignores those respondents who actually intended to say "50%" as an honest reflection of their numeric beliefs. This article offers two methods to reanalyze open-ended data so that it is as if they were collected with a probability response scale, correcting for the prevalence of nonnumeric 50s. Each estimates the reduction in 50 responses that would have been achieved with such a scale. The techniques are calibrated on our own data[2] and then used to correct probability judgments of experts regarding uncertain processes of groundwater pollution and bioremediation,[19] which were elicited with an open-ended response mode.

1. ANOTHER LOOK AT OUR DATA:[2] WHAT NUMBER IS "FIFTY-FIFTY"?

The differences in response distributions with open-ended and probability scale response modes suggest that the two elicit different cognitive processes, affecting respondents' treatment of epistemic uncertainty. When asked to generate their own probability, respondents may think in more verbal terms, raising vague beliefs captured by the verbal phrase "fifty-fifty." A nonnumeric 50 response reflects this epistemic uncertainty, while at the same time using a seemingly numeric expression that follows conversational norms set by the researcher.[20–23] However, it is not intended, and should not be interpreted, as a precise numeric probability.

One might ask, then, what happens with these vague beliefs when probability response scales are used (e.g., Fig. 1). It is possible that probability scales fail to resolve epistemic uncertainty, leading respondents to select tickmarks that might not have occurred to them naturally.[24–27] If so, then nonnumeric 50s would be a better indication of respondents' actual thinking, providing a window into qualitative aspects of their beliefs. Exploiting this possibility requires identifying the rate of nonnumeric 50s, as does restricting analyses to numeric 50s.

Alternatively, seeing a full set of numeric response options on a scale could help respondents (who would have said "50" to an open-ended question) to reduce their epistemic uncertainty. Those who resolve it completely will arrive at a specific number. Others may narrow it down to an interval (e.g., 10–20%, 45–50%).[5] Whereas a nonnumeric 50 would have been given as an open-ended response, a probability scale could lead to marking a number in that range. In that case, the scale would more accurately capture respondents' numeric beliefs, which may not be close to 50% at all.

Thus, respondents who use a nonnumeric 50 with an open-ended response mode may select a different response on a probability scale. There are also reasons for numeric responses to cluster in the middle range (e.g., between 40–60%). With any quantitative scale, respondents gravitate toward the middle,[28] which would be 50 with a 0–100% probability response range. With open-ended probability questions, 50 would be accentuated by the availability of the verbal phrase "fifty-fifty." Whatever the reason for the salience of 50, it may be used as an anchor from which insufficient (or even no) adjustment is made.[25,26,29] In addition, people's insensitivity to middle-range probabilities, when weighing them for decision-making purposes,[30] may encourage saying "50" for any answer that is "somewhere in the middle." Finer distinctions may not seem worth making. Such "midpoint effects" would cluster imprecise middle-range numeric probabilities around the 50 response.

Having to generate one's own probabilities (as with an open-ended response mode) should draw more mid-range responses toward 50, relative to a probability scale that explicitly offers 101 tickmarks, for different probabilities. While 50 is salient in the open-ended format, a scale offers alternatives for respondents looking for a response "somewhere in the middle." They should find it just as easy to mark a number near 50 as 50 itself. As a result, the probability scale would smooth the 50 blip by distributing responses more equally in the 40–60% range.

We believe, however, that nonnumeric 50s reflect more than just a mechanical midpoint effect. More likely, they often reflect epistemic uncertainty—not knowing exactly what number to use. If so, a probability scale that helped resolve this uncertainty should encourage numeric responses using the entire range (e.g., 0–100%).

Two of our studies[2] posed the same questions with open-ended and scale formats, allowing us to estimate the rate of nonnumeric 50s with the former. Comparing these two response distributions also allows estimating the responses that would have been given by respondents who provided a nonnumeric 50 with the open-ended format had a scale been presented instead. Specifically, this comparison could indicate whether the probability scale leads to a

5 More research is needed to understand whether the degree of epistemic uncertainty, as reflected in such a range,[6] affects the use of 50.

Table I. Percentage of Observations in the Lower, Middle, and Upper Range(2)

	Open-Ended			Probability Scale		
Question	$p \leq 0.35$	Middle	$p \geq 0.65$	$p \leq 0.35$	Middle	$p \geq 0.65$
Study 1						
Break-in	77	17	6	81	13	6
Cancer by 80	42	33	26	42	21	37
AIDS from sex	61	22	17	70	25	5
Bomb at university	84	9	7	88	5	6
Study 2						
Lightning	98	2	0	100	0	0
AIDS from sex	29	27	44	43	19	39
Breast cancer	46	41	13	63	24	13
Alive at 50	0	33	67	7	17	76
Cancer by 40	70	30	0	74	22	4
Lung cancer from smoking	21	36	43	22	39	39

redistribution of the nonnumeric 50s over the middle or over the entire response range. Table I presents the proportion of responses in the middle range for each question in the open-ended and scale conditions of two studies.(2) Study 1 asked (1) "What is the probability of the university being closed at least once this year because of a bomb threat?" (2) "What is the probability of someone getting AIDS if they have sex without protection?" (3) "What is your personal probability of developing cancer by age 80?" and (4) "What is the probability of someone breaking into your room or home and stealing something some time this year?" Probability questions in Study 2 considered "What is the probability that ..." (1) "You will be struck by lightning some time this year?" (2) "You will develop cancer by age 40?" (3) "Someone who smokes a pack or more of cigarettes a day will develop lung cancer?" (4) "Someone will get AIDS if they have sex without protection once with someone who is infected?" (5) "An average woman will be diagnosed with breast cancer in her lifetime?" and (6) "You will be alive at age 50?" Both studies had an open-ended and a scale condition. Open-ended questions instructed respondents to "write down a number between 0% (no chance) and 100% (certainty)" in a blank space. The scale questions asked respondents to place a mark on a probability scale (Fig. 1).

To increase the chances of finding a midpoint effect, we defined "somewhere in the middle" broadly as between 35–65%. If the open-ended response mode merely encourages saying "50," rather than other values in the middle range, there should be similar percentages of 35–65% responses with the two response modes. Table I shows that this was not the case. For eight of the nine probability questions, the open-ended condition evoked a higher percentage of mid-range probabilities, suggesting that the additional 50s were drawn from responses that would have been outside the midrange with a probability scale. (Although both studies had the same trend, the open-ended response mode elicited significantly more mid-range probabilities only in Study 2($\chi(1) = 6.05$, $p < 0.05$.) Thus, the 50 blip with open-ended questions seems to reflect more than just a mechanical midpoint effect.

The remainder of this article offers two methods to deal with the potential excess of 50s obtained with open-ended response modes. Each first assesses the percentage of nonnumeric 50 responses (i.e., those being drawn from outside the middle range), then redistributes them as if a numeric probability scale had been used. The first method uses the beta function, the second one an averaging heuristic.

2. BETA METHOD

The first method fits a curve to the histogram of observed open-ended responses,[6] assuming that it represents the form of the underlying distribution. To this end, we chose the beta family of functions. The beta distribution is used extensively to model uncertainty in binomial probability assessments because it is restricted to values over a fixed interval (i.e., 0–1), can approximate varied distributional shapes (e.g.,

6 The choice of response categories (e.g., 0–9, 10–19, ..., 90–99, 100) is somewhat arbitrary. Different divisions are possible, but should not influence the results reported here. In this division, the 50 responses are included in the category 50–59%, almost all of which are 50s.

skew in either direction, symmetry, kurtosis, and limited bimodality), is a general form of several common distributions (e.g., uniform, triangular), and has conjugate properties with the binomial distribution.[31] The beta distribution has been used to merge subjective judgments with formal probabilistic models in large Bayesian belief networks.[32,33] It also fits with our response distributions well.

A beta function is continuous from 0 to 1, with shape parameters α_1 and α_2 that can be set to fit observed CDF response distributions.[7] The fitted distribution shows the expected percentage of observations in each response category, including the proportion of 50s that would be expected in the data. We assume that these reflect the numeric 50 responses that would have been revealed had a probability scale been used, with any excess 50s being the result of using 50 as a nonnumeric proxy for "fifty-fifty." The procedure redistributes responses to match the best-fit curve. Thus, it estimates (and reduces) the number of observations in overused response categories (such as 50) and distributes them to underused ones, as though that is where they would have been had the scale response mode been used.[8]

When applied to the entire observed distribution, this procedure produces a beta function that is still biased toward the middle of the range (because it includes some nonnumeric 50s). The extent of this bias can be estimated by fitting a beta function to the observed distribution, omitting all 50s.

3. AVERAGING METHOD

The second method estimates the expected percentage of numeric 50s (included in the 50–59% category), as the mean of the number of responses in the two neighboring categories, 40–49% and 60–69%.[2] The number of 50s greater than this mean is treated as nonnumeric—reflecting the verbal phrase "fifty-fifty." Assuming that these excess 50s could come from anywhere in the 0–100% range, we redistributed them over all other categories, in proportion to their existing size.

4. TESTING THE TECHNIQUES

These two redistribution techniques were applied to the response distributions obtained with open-ended questions (Figs. 2A and 3A).[2] If the redistribution methods are effective, there should be better correspondence between the distributions with the open-ended and probability scale response modes after the redistribution than before. Correspondence was measured, for each question, in terms of (1) the *absolute difference* in the percentage of 50s and (2) the *sum of the squared differences* (SSD) between these two percentages.

Table II presents these measures, comparing the scale distribution with that from (1) the original, uncorrected open-ended data; (2) the best-fit beta function, using the entire open-ended distribution; (3) the best-fit beta function, using the open-ended distribution after excluding all 50s; and (4) the averaging rule. With both the beta and the averaging method, an excess of 50s was observed with each open-ended question. As a result, the correction procedures reduced the proportion of 50s. They also produced a percentage of 50s closer to that found with the probability scale for each question except AIDS-from-sex in Study 1 (where both methods overcorrected). Overall, the beta function produced a closer correspondence with the scale distribution, with nearly identical proportions of numeric 50s. The fit was slightly better when the beta function was fitted without the 50s than with those responses included. On average, a slight blip remained after the averaging correction, compared to scale responses.

The corrected open-ended distributions (Figs. 2A, 3A) also looked more like those with the probability scale (Figs. 2B, 3B) than the uncorrected open-ended distributions (Figs. 2A, 3A). The beta function fitted with all responses reduced SSD for 8 of 10 events, the beta fitted without 50s for 6 of 10, and the averaging method for 7 of 10. On average, the sum of the squared differences was smallest for the beta function fitted on the complete distribution (including the 50s). Overall, by both criteria, each redistribution method reduced the differences between the (corrected) open-ended and scale distributions. The

7 The beta distribution is continuous from 0 to 1, with the general shape

$$\begin{array}{ll} \text{if} \quad 0 < x < 1 & f(x) = \frac{(x^{\alpha 1-1}(1-x)^{\alpha 2-1})}{B(\alpha 1, \alpha 2)} \\ \text{otherwise} & f(x) = 0 \end{array}$$

where $B(\alpha 1, \alpha 2)$ is the beta function defined by

$$B(\alpha 1, \alpha 2) = \int_0^1 t^{\alpha 1 - 1}(1-t)^{\alpha 2 - 1}$$

for any real numbers $\alpha 1 > 0$ and $\alpha 2 > 0$.

8 We are not claiming that any blip is anomalous. Legitimate reasons for an excess may be that it reflects the modal response, or because subgroups of respondents have different beliefs. A peak at 50, however, seems anomalous because it occurs systematically across questions and can be manipulated by changes in sampling, questions, and response modes.

Table II. Differences Between Open-Ended Distributions and Scale Response Modes[2]

	Original Open-Ended Data		Open-Ended After Beta Correction		Open-Ended After Beta Correction (Without 50s)		Open-Ended After Averaging Correction	
	50 diff.[a]	SSD[b]	50 diff.[a]	SSD[b]	50 diff.[a]	SSD[b]	50 diff.[a]	SSD[b]
Study 1								
Break-in	+10%	0.022	+2%	0.014	+2%	0.016	−1%	0.017
Cancer by 80	+5%	0.030	+2%	0.018	+2%	0.015	+2%	0.027
AIDS from sex	+1%	0.009	−7%	0.015	−7%	0.016	−9%	0.017
Bomb at university	+3%	0.092	0%	0.107	0%	0.107	−1%	0.102
Study 2								
Lightning	+2%	0.003	0%	0.001	0%	0.004	0%	0.004
AIDS from sex	+14%	0.040	+1%	0.038	0%	0.049	−6%	0.022
Breast cancer	+15%	0.036	+4%	0.011	+3%	0.010	−1%	0.012
Alive at 50	+10%	0.044	0%	0.017	−1%	0.015	−2%	0.039
Cancer by 40	+14%	0.020	−4%	0.010	−3%	0.010	−7%	0.013
Lung cancer from smoking	+14%	0.040	+6%	0.021	+6%	0.021	+4%	0.021
Average	+8.5%	0.034	+0.4%	0.025	+0.2%	0.026	−2.1%	0.027

[a]Difference in percentages of 50s, with a positive sign indicating more 50s in the open-ended condition.
[b]Differences in distributions, as measured by the sum of squared differences in histograms.

next section considers strengths and weaknesses of each method.

5. FURTHER EVALUATION OF THE METHODS

The beta correction, whether fitted with or without the 50s, shows slightly better results than the averaging technique. Its other advantages include being based on well-understood statistical procedures and its consideration of the cardinal values of all entered observations. Thus, the beta function is not disproportionately affected by observations of any category and hence will smooth any unexpected blip or dip in the distribution; it takes fuller advantage of the metric properties of the responses.

Unfortunately, the beta distribution shows larger sums of squared differences when the data are bi- or multimodal, as with the AIDS-from-sex question from Study 1.[2] Bimodality can occur with distributions that include responses from individuals representing different populations. For example, the risk of developing breast cancer is higher for older women than for younger ones. A sample of knowledgeable younger and older women (e.g., mothers and daughters) would produce a bimodal distribution of personal probabilities of breast cancer. If the identity of such subgroups is known, separate distributions could be created. Each could be corrected separately—merging the redistributed responses—if an overall distribution were desired.

The averaging rule is less sensitive to bimodality insofar as it considers only two response categories. It faces a problem when one category is overused, for whatever reason. Whereas the beta function would take this into account (and level the category), the averaging rule maintains the overuse by assigning more 50 responses to the category. The averaging rule will also produce a poor fit with scale responses when the 40s or 60s category is extremely large or small, affording undue weight to chance variations.

A final problem with the averaging method arises with distributions whose modal response is 50. In this case, the average number of observations in the neighboring categories would underestimate the expected number of 50s. Although such a distribution may include nonnumeric 50 responses, applying the averaging rule may "overcorrect" for them. This problem does not affect the beta function, which fits the entire distribution. As it happens, none of the questions in Tables I and II had such a distribution.

Although the averaging rule was not as effective as the beta procedure in "de-biasing" the data in Table II, it is much easier to use than the beta function. Indeed, simply eyeballing a histogram of the open-ended data gives a good indication of the proportion of numeric 50s expected by this method—and the excess compared to what is observed.

Thus, both techniques "de-biased" the open-ended responses in the sense of making them more similar to scale responses using the same questions. These distributions were taken from a study with a

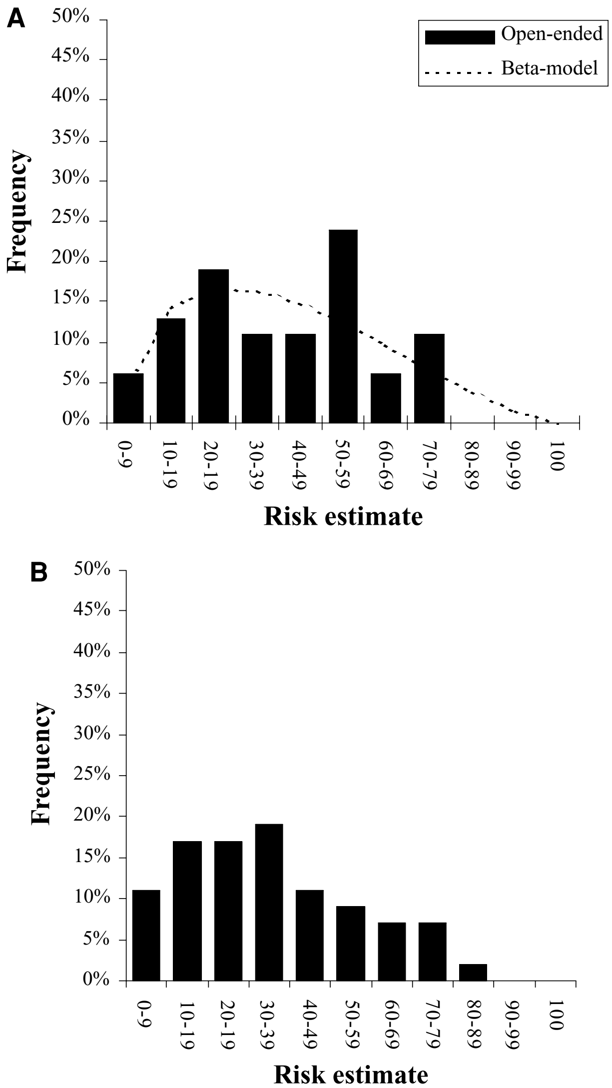

Fig. 2. Open-ended probability distribution and its beta correction (A) as well as the scale distribution (B), for the breast-cancer question.[3]

large number of observations, allowing relatively stable estimation, with a better chance for unimodal distributions to emerge. Results could be less satisfactory with smaller samples. Furthermore, even these results apply only to the pooled data, indicating that some—but not which—respondents did not mean a numeric value when they said "50." The beta and averaging methods correct population statistics, rather than individual responses. The corrected data might provide a sounder basis for general decisions (e.g., the design of public risk communications).

The next section shows one kind of policy application—refining expert judgments—in the con-

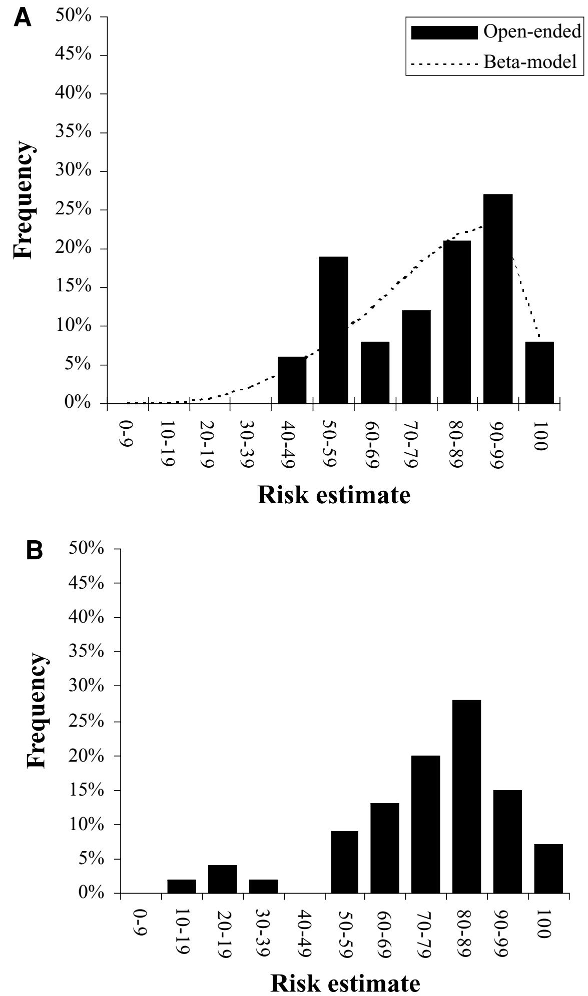

Fig. 3. Open-ended probability distribution and its beta correction (A) as well as the scale distribution (B), for the alive-at-50 question.[3]

text of a model of groundwater pollution. It also asks the empirical question of whether excess 50s will be observed in the judgments of experts working within their field of expertise.

6. AN EXPERT MODEL OF GROUNDWATER POLLUTION

Stiber *et al.*[19] developed a causal model of groundwater pollution to support an expert system for evaluating the adequacy of reductive de-chlorination as a remedial option for sites with groundwater contaminated by trichloroethylene (TCE). It follows the formalism of the influence diagram.[34,35] Each node represents a chance variable that plays a role in the reductive de-chlorination process. An arrow means that the value of a node depends on the value of the preceding node. For example, *environmental conditions* affect *anaerobic degradation by reductive dechlorination*. These environmental conditions include the temperature, pH, and dissolved oxygen content of the groundwater. The bacteria that perform the biodegradation have a range of conditions within which their performance is optimal, and a range within which their activity is slowed or impossible. Changes in the value of each link cascade through the other nodes, affecting the predicted probability of reductive de-chlorination.

Because many of the conditional probabilities in this influence diagram could not be obtained empirically, expert elicitation was used to complete it. Specifically, 22 experts in biological degradation of chlorinated solvents were each asked to assess 98 different conditional probabilities for three variations of the model. For example, one question asked, "Given that anaerobic degradation is occurring, what is the probability that vinyl chloride is detected?" These judgments were elicited over the telephone[9] using an open-ended format similar to those often used with lay subjects. The experts' mean responses were used to compute the contingencies and model predictions.

7. TESTING—AND CORRECTING—EXPERT JUDGMENTS

7.1. Testing

The 22 experts used an excess of 50s. On average, the percentages of 50s observed with the 98 probability questions exceeded the expected percentage by 4% according to the beta method (SD = 0.07), and by 5% according to the averaging method (SD = 0.10). The difference ranged from −12% to 32% for the beta and −19% to 45% for the averaging method.

According to the beta method, 58 out of the 98 questions had excess 50s; according to the averaging method, 59 of 98. Though some questions showed fewer 50s than expected by the correction procedures, across all 98 response distributions an overuse, possibly due to nonnumeric 50s, was observed ($t(97) = 7.35$, $p < 0.001$ for the beta function; $t(97) = 7.85$, $p < 0.001$ for the averaging rule).

7.2. Correcting

We applied the beta and averaging methods to redistribute excess, nonnumeric 50s for those questions where there were more than expected (58 questions with beta, 47 with averaging). The averaging rule was not used on 12 questions that showed 50 as the modal response.

In the previous section (Table II), the extent of discrepancy was evaluated by comparing the debiased response distribution with that obtained using the probability scale. Because that response mode was not used with the experts, the corrected and uncorrected distributions were compared, using the SSD criterion. Because the size of the 50 blip affects this measure, the 50s were omitted from the calculation. Generally, the corrected distributions showed a good fit with the remaining categories of the original open-ended data. The mean SSD was 0.003 (SD = 0.004) for the beta method, and 0.002 (SD = 0.003) with the averaging rule.

The beta method produced a mean difference of 1.42 (SD = 4.17), with the largest being a reduction of 31.9. The averaging rule produced a mean difference of 1.86 (SD = 1.34), with the largest being 6.87. Both corrections significantly changed the mean expert judgment on these questions: $t(57) = 2.61$, $p < 0.05$) for the beta method and ($t(46) = 9.55$, $p < 0.001$) for averaging.

8. DEVELOPING A PREDICTIVE MODEL

This section presents a simple model for predicting the rate of excess 50s obtained with open-ended response modes using the already examined lay[(2)] and expert data[(19)] to estimate model parameters. The model assumes that the likelihood of a 50 response depends on respondents' degree of epistemic uncertainty and the closeness of their corresponding numeric probability to 0.5.[10] Respondents are more likely to say "50" if they (1) experience epistemic uncertainty and (2) have numeric probabilities closer to 0.5. To represent the first factor, we define an *epistemic*

9 "Each probability can be expressed as a fraction (from 0 to 1.0); or, as a percentage (from 0% to 100%)—whichever is more comfortable with you. You may use as many (or as few) decimal places and significant figures as you feel appropriate. Please try to provide probabilities that are as precise as possible. A probability of 1.0 (100%) means that you are completely certain about an event's occurrence. Similarly, a probability of 0.0 (0%) means that you believe the event is a total impossibility. During the elicitation, try to imagine all possibilities and avoid using 1.0 (100%) and 0.0 (0%) unless they are truly fitting."

10 This can be considered the probability that the respondent would provide, if given enough time for careful reflection as well as instruction on likely cognitive biases that can affect probability judgments.

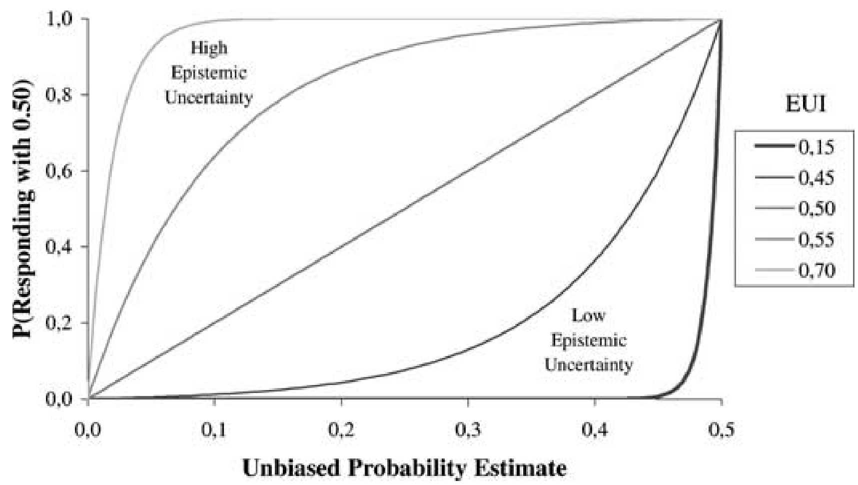

Fig. 4. The effects of epistemic uncertainty on the likelihood of a nonnumeric 50 response.

uncertainty index (EUI) equal to 0 for well-understood events and 1.0 for completely uninformed priors. The second factor is an unbiased estimate of the probability, p_u, putting aside any epistemic uncertainty surrounding it. The probability of a 50 response becomes: P(responding 0.5) = f(p_u, EUI).

Fig. 4 shows a family of exponential functions expressing such behavior. In domains of low epistemic uncertainty, the probability of a 50 response is essentially 0—unless the unbiased probability is close to 0.5. However, with high epistemic uncertainty, a 50 response is possible even when respondents believe the numeric probability to be small.

If this model holds, then the observed mean probability (p_o) for a group of individuals with similar levels of epistemic uncertainty and "true" numeric probabilities would be:

$$p_o = f(p_u, EUI)\,0.5 + (1 - f(p_u, EUI))\,p_u.$$

This is simply the weighted average of the "true" numeric probability estimate and 0.5 (the 50 response). Fig. 5 shows the mean observed probabilities for hypothetical groups having different levels of EUI and unbiased probability estimates.

Using the lay and expert data sets presented earlier, the reasonableness of this model can be explored. Fig. 6 is an X-Y scatter plot of the average assessed probabilities and their corrected "true" values for the lay and expert data for which the beta method was significant. Overlaid on these plots are the predictive curves for increasing levels of epistemic uncertainty (EUI). Note that the expert responses fall much closer to the diagonal, indicating a lower EUI, as might be expected for individuals making judgments in their

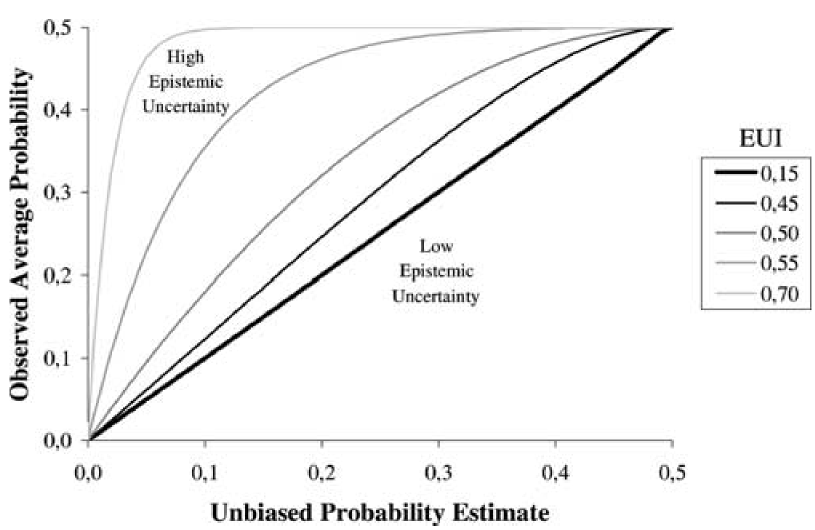

Fig. 5. The effects of epistemic uncertainty on the average probability response.

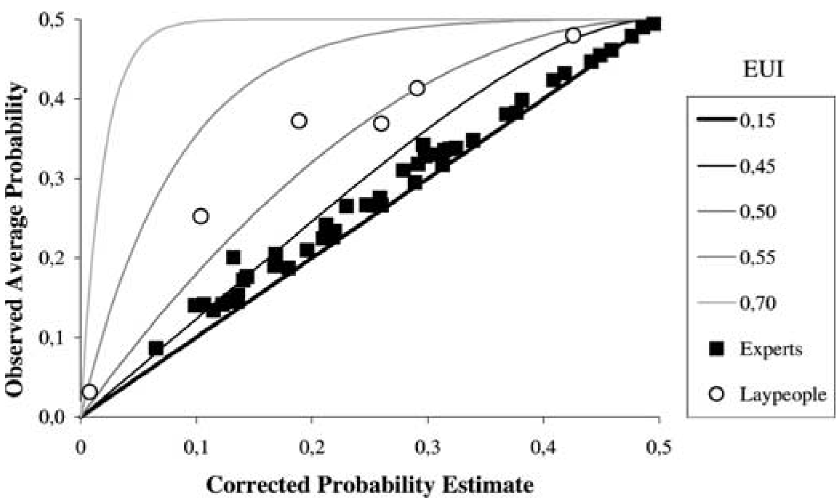

Fig. 6. The effects of epistemic uncertainty on the lay[2] and expert[19] elicitations. The corrected probability estimates were derived using the beta correction technique.

domains of expertise. Moreover, the experts also display relatively consistent levels of EUI across the various questions. Mean lay responses show more variance in EUI across questions, which spanned many different topics. The consistent expert bias allows a standard postassessment correction factor for any questions thought to come from a domain for which experts experience similar EUI.

The "fifty-fifty" bias leads, on average, to probability judgments closer to 0.5 than is warranted. If the size of the bias is known, one can estimate its effects on the expected value (or utility) of options and of information in a given decision problem. Depending on the decision, the bias can lead to an overestimation or underestimation of the true values. Applying formal techniques similar to those used in determining the expected value of information in classic decision analysis,[36] the cost of the bias can be estimated.

9. DISCUSSION

We presented two methods for estimating the magnitude of the nonnumeric 50s and eliminating its effects in existing open-ended data. These methods were applied to responses from our previous study,[2] comparing responses with open-ended and scale response modes. The two methods showed similar results, with the beta function performing slightly better than the averaging rule in the sense of the distributions of corrected open-ended responses distributions looking more like the scale distributions. Clearly, these are not the only possible correction methods. However, both performed fairly similarly in estimating the excess of 50s potentially due to epistemic uncertainty and in redistributing them as if a probability scale had been used.

The beta and averaging methods were also applied to the probability judgments given by groundwater pollution experts. They showed an excess of 50s for 60% of the 98 probability questions. Thus, even with experts, the open-ended elicitation procedure may have inadvertently induced nonnumeric "fifty-fifty" responses.

In this application, the beta procedure fitted the original response distributions less well than did the averaging rule. Possibly, the small number of responses (from just 22 experts) resulted in less smooth distributions, creating problems for the unimodal beta function. If so, then the averaging rule might be recommended for redistributing nonnumeric 50s from studies eliciting a relatively small numbers of judgments with an open-ended response mode.

The predictive model (Fig. 6) indicated that the experts[19] experienced less epistemic uncertainty than did the lay respondents.[2] Nonetheless, the experts, too, showed an excess of 50s for most questions, with significant effects on mean probabilities. Although we have no direct evidence that the experts' 50 blips result from epistemic uncertainty, our previous research suggests this relationship.[1–3] Depending on the sensitivity of the specific decision, such mis-estimation may have little effect or could seriously affect the calculations and choices. Awareness of these possibilities can facilitate the choice of elicitation method or adoption of correction procedure.

REFERENCES

1. Bruine de Bruin, W. (1998). People's understanding of probability: "It's a 'fifty-fifty' chance." Doctoral dissertation. Pittsburgh, PA: Carnegie Mellon University, Department of Social and Decision Sciences.

2. Fischhoff, B., & Bruine de Bruin, W. (1999). "Fifty-fifty" = 50%? *Journal of Behavioral Decision Making*, *12*, 149–163.
3. Bruine de Bruin, W., Fischhoff, B., Millstein, S. G., & Halpern-Felsher, B. L. (2000). Verbal and numeric expressions of probability: "It's a 'fifty-fifty' chance." *Organizational Behavior and Human Processes*, *81*, 115–131.
4. Joke of the Day. (2000). 50-50-90 rule. *Joke of the Day Archive*. http://www.joker.org/archive/query.cgi?ID = 1114.
5. von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.
6. Gärdenfors, P., & Sahlin, N. E. (1982). Unreliable probabilities, risk taking and decision making. *Synthese*, *53*, 361–386.
7. Janssens, A. C. J. W., de Boer, J. B., Hintzen, R. Q., van Doorn, P. A., Passchier, J., & van der Meché, F. G. A. (2001). "It might happen or not": Patients' perceptions of prognostic risk in Multiple Sclerosis. Paper presented at the meeting of the Biannual Conference on Subjective Probability, Utility, and Decision Making. Amsterdam, the Netherlands.
8. Viscusi, W. K. (1993). *The risks of smoking*. Cambridge, MA: Harvard University Press.
9. Black, W. C., Nease, R. F., & Tosteson, A. N. A. (1995). Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *Journal of the National Cancer Institute*, *8*, 720–731.
10. Dominitz, J., & Manski, C. F. (1997). Perceptions of economic insecurity. Evidence from the survey of economic expectations. *Public Opinion Quarterly*, *61*, 261–287.
11. Fiedler, K. (1988). The dependence of conjunction fallacy on subtle linguistic factors. *Psychological Research*, *50*, 123–129.
12. Kahneman, D., & Tversky, A. (1982a). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). New York: Cambridge University Press.
13. Kahneman, D., & Tverksy, A. (1982b). Variants of uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 509–520). New York: Cambridge University Press.
14. Reeves, T., & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General*, *122*, 207–226.
15. Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
16. Tversky, A., & Koehler, D. J. (1994). Support theory: A nonexistensional representation of subjective probability. *Psychological Review*, *4*, 547–567.
17. Fischhoff, B., Parker, A. M., Bruine de Bruin, W., Downs, J., Palmgren, C., Dawes, R., & Manski, C. F. (2000). Teen expectations for significant life events. *Public Opinion Quarterly*, *64*, 189–205.
18. Bradburn, N. M., Sudman, S., Blair, E., Locander, W., Miles, C., Singer, E., & Stocking, C. (1979). *Improving interview method and questionnaire design. Response effects to threatening questions in survey research*. San Francisco, CA: Jossey-Bass.
19. Stiber, N. A., Pantazidou, M., & Small, M. J. (1999). Expert system methodology for evaluating reductive dechlorination at TCE sites. *Environmental Science and Technology*, *33*, 3012–3020.
20. Fox, C. R., & Irwin, J. R. (1998). The role of context in the communication of uncertain beliefs. *Basic and Applied Social Psychology*, *20*, 57–70.
21. Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3 Speech acts* (pp. 95–113). New York: Academic Press.
22. Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Erlbaum.
23. Wallsten, T. S., & Budescu, D. V. (1995). A review of human linguistic probability processing: General principles and empirical evidence. *Knowledge Engineering Review*, *10*, 43–62.
24. Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117–141). Hillsdale, NJ: Lawrence Erlbaum Associates.
25. Poulton, E. C. (1989). *Bias in quantifying judgment*. Hillsdale, NJ: Lawrence Erlbaum Associates.
26. Poulton, E. C. (1994). *Behavioral decision theory: A new approach*. New York: Cambridge University Press.
27. Schwarz, N., & Hippler, H. J. (1987). What response scales may tell your respondents: Informative functions of response alternatives. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology*. New York: Springer Verlag.
28. Poulton, E. C. (1982). Biases in quantitative judgments. *Applied Ergonomics*, *13*, 31–42.
29. Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
30. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–281.
31. Law, A. M., & Kelton, W. D. (1982). *Simulation modeling and analysis*. New York: McGraw-Hill.
32. Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, *20*, 197–243.
33. Winkler, R. (1967). The assessment of prior distributions in Bayesian analysis. *American Statistical Association Journal*, *62*, 776–800.
34. Howard, R. A. (1989). Knowledge maps. *Management Science*, *35*, 903–922.
35. Burns, W. J., & Clemen, R. T. (1993). Covariance structure models and influence diagrams. *Management Science*, *39*, 816–834.
36. Clemens, R. T. (1996). *Making hard decisions: An introduction to decision analysis* (2nd ed.). Belmont, CA: Duxbury.