Worth a Second Look

A paper that SPE has published in the past that deserves renewed attention.

Ed Capen and his Arco colleagues in the 1960s and 1970s were pioneers in publicizing and adopting improved techniques for decision making under uncertainty that were emerging from the Department of Engineering Economics at Stanford University and from Harvard Business School. Arguably, they were more advanced than some companies are today. With colleagues Bill Campbell and Bob Clapp, he was also responsible for coining the term "The Winner's Curse," to describe the observation that, in a sealed auction, the winner often significantly overbids.

Although their main focus was on quantitative techniques such as decision tree analysis and Monte Carlo simulation, Capen and his coworkers also recognized a vital human element upon which these techniques relied—the ability of people to accurately assess uncertainty, which is the subject of this issue's "Worth a Second Look." In this paper, Capen showed that most people are grossly overconfident. That is, they specify uncertainty ranges that are too narrow with respect to their actual knowledge of the variable they are assessing.

As you read the paper, it is worth bearing in mind the following. Uncertainty is personal, and probability expresses the relationship between the objective world and the strength of one's subjective knowledge of that world. Thus, uncertainty is in our heads, not in the world. One should, therefore, not speak of "the" probability of an event, but of "my" probability. Individuals may agree upon the probability that should be assigned to very simple, well-defined events such as the toss of a coin. Most events of interest in the oil and gas world are more complex, and there is no reason that individuals' probabilities should agree because their knowledge differs.

Thus, there is no unique, "right" uncertainty range (other than "right" insofar as it is consistent with one's knowledge) for the questions that Capen used. His experiments were designed to measure the extent of his subjects' awareness of their lack of knowledge. Someone who knows a lot would be expected to place narrower ranges than someone who knows a little. Therefore, his methodology, and the validity of his results, are not dependent upon how familiar the subjects were with the topics of his questions. Indeed, similar tests are routinely performed with today's oil and gas employees to make them aware of their overconfidence.

Sadly, since Capen's paper, nothing much would appear to have changed. Recent studies of engineers and geoscientists working in the oil and gas industry show that they are still grossly overconfident even when asked questions related to their work, as opposed to the general knowledge questions that Capen used. Even those who have been trained in uncertainty assessment or whose daily job it is, are overconfident, though not quite so badly.

Does overconfidence matter? Yes, it is if one does not want to be surprised about the outcomes of events more often than is warranted by one's uncertainty assessment. More importantly, if one's decisions are based on one's assessment of uncertainty, as they should and usually are, it does. Greatly.

Steve Begg
Associate Editor, *SPE Economics & Management*
University of Adelaide

What follows is a reprint of an SPE paper from 1976. It has been peer-reviewed.

The correct citation for the paper would be:
Capen, E.C. 1976. The Difficulty of Assessing Uncertainty. *J Pet Technol* **28** (8): 843–850. doi: 10.2118/5579-PA.

The Difficulty of Assessing Uncertainty

E.C. Capen, SPE-AIME, Atlantic Richfield Co.

What do you do when uncertainty crosses your path? Though it seems that we have been taught how to deal with a determinate world, recent testing indicates that many have not learned to handle uncertainty successfully. This paper describes the results of that testing and suggests a better way to treat the unknown.

Introduction

The good old days were a long time ago. Now, though we must harness new technology and harsh climates to help provide needed energy supplies, we are also faced with the complex problem of satisfying not altogether consistent governments, the consumer, our banker, and someone's time schedule. Judging from the delays, massive capital overruns, and relatively low return this industry has experienced lately, it would seem that we have been missing something. At least one explanation is that we have not learned to deal with uncertainty successfully.

Some recent testing of SPE-AIME members and others gives rise to some possible conclusions:

1. A large number of technical people have little idea of what to do when uncertainty crosses their path. They are attempting to solve 1976 problems with 1956 methods.

2. Having no good quantitative idea of uncertainty, there is an almost universal tendency for people to understate it. Thus, they overestimate the precision of their own knowledge and contribute to decisions that later become subject to unwelcome surprises.

A solution to this problem involves some better understanding of how to treat uncertainties and a realization that our desire for preciseness in such an unpredictable world may be leading us astray.

Handling Uncertainty

Our schooling trained us well to handle the certainties of the world. The principles of mathematics and physics work. In Newton's day, force equaled mass times acceleration, and it still does. The physicists, when they found somewhat erratic behavior on the atomic and molecular level, were able to solve many problems using statistical mechanics. The extremely large number of items they dealt with allowed these probabilistic methods to predict behavior accurately.

So we have a dilemma. Our training teaches us to handle situations in which we can accurately predict the variables. If we cannot, then we know methods that will save us in the presence of large numbers. Many of our problems, however, have a one-time-only characteristic, and the variables almost defy prediction.

You may embark on a new project whose technology differs from that used on other projects. Or perhaps your task is to perform a familiar project in a harsh environment. Try to estimate the total cost and completion time. Hard! You cannot foresee everything. And, for some reason, that which you cannot foretell seems to bring forth more ill than good. Hence, the predictions we make are often very optimistic. Even though we see the whole process unfolding and see estimate after estimate turn out optimistic, our next estimate more than likely will be optimistic also.

What happens? Is there some deep psychological phenomenon that prevents our doing better? Because we are paid to know, do we find it difficult to admit we do not know? Or can we obtain salvation through knowledge? As we were trained to handle certainty, can we also find a better way to estimate our uncertainty?

I think so, but it will take some special effort - just as it did when we first learned whatever specialty that got us into the business. As one of the Society's Distinguished Lecturers for 1974–75, I had a unique opportunity to collect information on the way our membership treats uncertainty. I do not claim that what you are about to read will set the scientific or business communities to quaking (others have noticed similar phenomena before[1]). But there are lessons that should help to improve our perceptions of uncertainty and, we hope, increase our economic efficiency by giving us better information on which to base decisions.

SPE-AIME Experiment

The experiment went like this. Each person put ranges around the answers to 10 questions, ranges that described his personal uncertainty. The questions were the following:

1. In what year was St. Augustine (now in Florida) established as a European settlement?
2. How many autos were registered in California in 1972?
3. What is the air distance from San Francisco to Hong Kong in miles?
4. How far is it from Los Angeles to New Orleans via major highways in miles?
5. What was the census estimate of U. S. population in 1900?
6. What is the span length of the Golden Gate Bridge in feet?
7. What is the area of Canada in square miles?
8. How long is the Amazon River in miles?
9. How many earth years does it take the planet Pluto to revolve around the sun?
10. The English epic poem "Beowulf" was composed in what year?

For some, the task was to put a 90-percent range around each answer. The person would think up a range such that he was 90-percent sure the range would encompass the true value. For example, in one section a gentleman put a range of 1500 to 1550 on Question 1. He was 90-percent sure that St. Augustine was established after 1500, but before 1550. In his view, there was only a 5-percent chance that the settlement came into being after 1550. If he were to apply such ranges for many questions, we would expect to find about 10 percent of the true answers outside of his intervals.

Other groups were asked to use 98-percent ranges virtual certainty that their range would encompass the true value. I also asked for ranges of 80, 50, and 30 percent. The 30-percent interval would supposedly allow 70 percent of the true answers to fall outside the range.

Most sections used a single probability range. However, a few groups were divided in two, with each half using different intervals, usually 30 and 90 percent. I shall refer to these ranges as probability intervals.

You may want to test your skill on the test, too. The answers are in the Appendix. Use a 90-percent interval so you can compare with results given later.

Results and Conclusions. My testing turned up traits that should be of interest. [From this point on, the people referred to are the 1,200+ people at the local section meetings who answered the questions sufficiently to be counted. There were a significant number (350 or so) at the meetings who either had no idea of how to describe uncertainty or thought it chic not to play the game.]

1. People who are uncertain about answers to a question have almost no idea of the degree of their uncertainty. They cannot differentiate between a 30- and a 98-percent probability interval.

2. The more people know about a subject, the more likely they are to construct a large probability interval (that is, one that has a high chance of catching the truth), regardless of what kind of interval they have been asked to use. The converse seems to hold also; the less known, the smaller the chance that the interval will surround the truth.

3. People tend to be a lot prouder of their answers than they should be.

4. Even when people have been previously told that probability ranges tend to be too small, they cannot bring themselves to get their ranges wide enough, though they do somewhat better.

5. Simultaneously putting two ranges on the answers greatly improved performance, but still fell short of the goal.

Such conclusions come from the following observations. Looking at the data collected on each of the sections, we find that the average number of "missed" questions was close to 68 percent. We could adopt the following hypothesis:

SPE-AIME sections will miss an average 68 percent of the questions, no matter what probability ranges they are asked for.

Mathematical statisticians have invented a way to test such hypotheses with what they call confidence intervals. They recognize, for instance, that the Hobbs Petroleum Section average of 6.26 misses out of 10 questions is subject to error. Slightly different questions, a different night, a longer or shorter bar - all kinds of things could conspire to change that number. By accounting for the variability of responses within the Hobbs chapter and the number of data points that make up the average, these statistical experts can put a range around the 6.26 much like the ranges the members were asked to use. Except that (unlike the members) when the statistician says he is using a 95-percent range, he really is!

For Hobbs, that range comes out to be 5.45 to 7.07. Since that range includes 6.8, or 68-percent misses, the statistician will agree that, based on his data, he would not quarrel with the hypothesis as it applies to Hobbs.

Table 1 shows all the 95-percent ranges and **Fig. 1** illustrates how these ranges compare with the 68-percent hypothesis. You will see a portion of the Los Angeles Basin Section whose confidence interval (5.24 to 6.68) does not include 6.8. There are three possible explanations:

1. The group has a bit more skill at handling such a problem than most.

2. Being part of an audience that was asked to use two different ranges, there was a more conscious effort on their part to use a wider range.

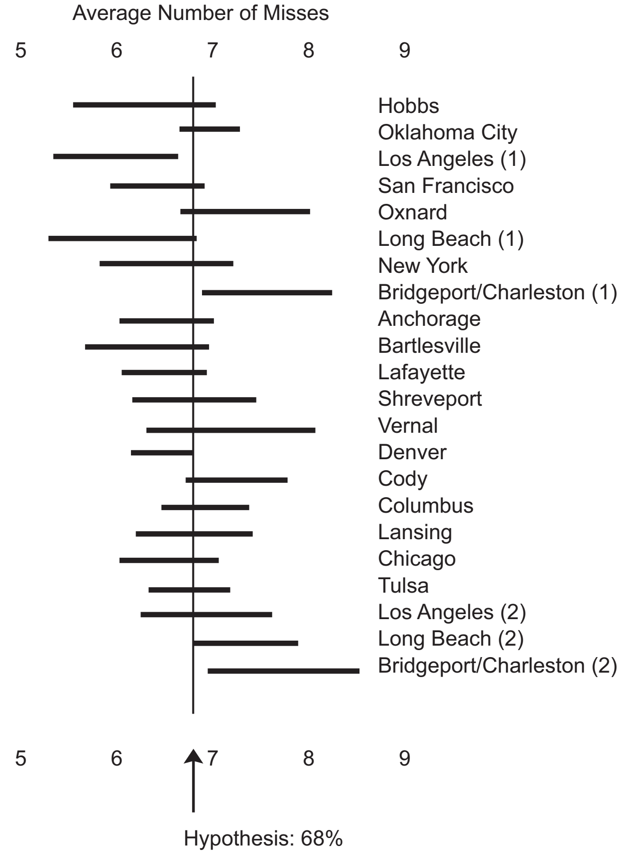

Fig. 1—The 95-percent confidence intervals of SPE-AIME sections. Average number of misses on 10-question quiz.

TABLE 1—SUMMARY OF 95-PERCENT RANGES

SPE-AIME Section	Number of Usable Responses	Requested Range (percent)	Expected Number of Misses	Actual Number of Average Misses	95-Percent Confidence Interval
Hobbs Petroleum	34	98	0.2	6.26	5.45 to 7.07
Oklahoma City	11	98	0.2	7.00	6.64 to 7.36
Los Angeles Basin (1)	28	90	1	5.96	5.24 to 6.68
San Francisco	61	90	1	6.41	5.89 to 6.93
Oxnard	26	90	1	7.38	6.64 to 8.12
Long Beach (1)	28	90	1	6.04	5.20 to 6.88
New York	29	90	1	6.52	5.76 to 7.28
Bridgeport/Charleston (1)	16	90	1	7.63	6.89 to 8.37
Anchorage	63	90	1	6.54	6.00 to 7.08
Bartlesville	44	90	1	6.30	5.61 to 6.99
Lafayette	79	90	1	6.51	6.03 to 6.99
Shreveport	41	90	1	6.83	6.18 to 7.48
Vernal	13	80	2	7.23	6.30 to 8.16
Denver	129	80	2	6.46	6.12 to 6.80
Cody	42	80	2	7.31	6.74 to 7.88
Columbus	27	50	5	6.96	6.47 to 7.45
Lansing	30	50	5	6.83	6.16 to 7.50
Chicago	41	50	5	6.54	5.97 to 7.11
Tulsa	53	50	5	6.79	6.33 to 7.25
Los Angeles Basin (2)	27	30	7	7.00	6.26 to 7.74
Long Beach (2)	28	30	7	7.39	6.80 to 7.98
Bridgeport/Charleston (2)	15	30	7	7.82	6.97 to 8.67

TABLE 2—COMPARISON OF RESULTS					
Section	Number of Useable Response	Requested Range (percent)	Expected Number of Misses	Actual Number Average Misses	95-Percent Confidence Interval
Atlantic Richfield R&D	52	98	0.2	4.52	3.84 to 5.20
SPE-AIME Section (Hobbs and Oklahoma City)	145	98	0.2	6.83	6.50 to 7.16

3. The statistics are misleading, and the group is not different from the others. We expect this to happen about 5 percent of the time. (Our testing mechanism was a 95-percent confidence interval.)

Likewise, the Bridgeport/Charleston (W. Va.) sections had ranges that did not encompass 6.8. In their defense, the meal service had been poor, the public address system had disappeared, and there were more than the normal misunderstandings. Even so, their lower limits of 6.87 and 6.97 just barely missed the 6.8 target.

One group of highly quantitative people also took the test. I mention this group because of the large number of members it includes and because it provides evidence that the more quantitative people may do a little better in estimating uncertainty - but still not as well as they would like. (See **Table 2.**)

The 68 percent would not be expected to hold on all kinds of questions or all kinds of people. In fact, it is clear that the number would have been higher had it not been for relatively easy questions such as Questions 1 and 4. Most people know St. Augustine was a Spanish community and, therefore, had to be established between 1492 and 1776. By making the range a bit more narrow than that, they could be reasonably sure of bracketing the true answer. Even so, more than one-third of the members missed that one - regardless of their instructions on range.

Based on a sample of the 1,200+ quizzes, here are the average misses for each question:

Question	Average Misses (percent)
1	39
2	67
3	60
4	50
5	69
6	68
7	76
8	69
9	74
10	85

Questions such as Questions 9 and 10 were difficult, and we found 80 percent or so misses - again regardless of the requested probability of a miss.

People who have no idea of the answer to a question will apparently try to fake it rather than use a range that truly reflects their lack of knowledge. This trait may be as universal a part of human nature as laughter; certainly it is not peculiar to SPE-AIME members.

Is the Problem Costly?

Why should anyone get excited about such results? Because, I think, similar behavior on the job can cost industry a bundle. Our membership at various levels of management is responsible for all sorts of daily estimates that ultimately work their way into investment decisions. To the extent that the success of the investments relies on those estimates, business can be in trouble. If one's range so seldom encompasses the truth on tough questions, then the more common single-point estimates have little chance of being very close. Even those beloved "what-ifs" cannot be of much help since such questions would only be expected to test "reasonable" ranges. This research seems to indicate that most of us have little idea of what is a reasonable range.

Other Experiments

Earlier, I mentioned that we might be able to practice this business of estimating uncertainty and improve our track record. Experience with the SPE-AIME sections says that the practice may have to be substantial. Having established the 68-percent norm during the early part of my tour, I was able to do some other experimenting later.

One section had the benefit of knowing ahead of time what all the other sections had done. They knew before they started that no matter what range I had asked for, the membership always responded with about 68-percent misses, or a 32-percent probability interval. This group of 143 knew, then, that the tendency was to give much too tight a range and that they should be very careful not to fall into the same trap. (See **Table 3.**) It would seem that my warning had some effect. The mere telling of the experience of others is not, however, enough to shock most people into an acceptable performance.

Menke, Skov, and others from Stanford Research Institute's (SRI) Decision Analysis Group have experimented along similar lines (and, in fact, their work gave me the idea for these tests). They say that if groups repeatedly take quizzes such as those described here, they are able to improve. Initially, people gave 50-percent ranges even though 98-percent ranges had been asked for. After several such tests (different each time, of course), the participants were able to reach a 70-percent range, but could never quite break that barrier. Their results show, apparently, that many intelligent men and women (they dealt largely with business executives) can never admit all their uncertainty. SRI made sure that some of their tests were built from subject matter familiar to the executives, such as questions extracted from their own company's annual report. Therefore, the phenomenon we are describing must have very little to do with the type of question.

Value of Feedback

For several years now we have asked our exploration people for 80-percent ranges on reserves before drilling an exploratory well. But we recognized that the act of putting down a 10-percent point and a 90-percent point would not in itself be sufficient. We also asked them to see what their 80-percent range told them about other points on the distribution curve. If one is willing to assume a certain form of probability distribution, then the 80-percent range also specifies every other point. Hence, the explorationist can essentially put himself into a feedback loop. He puts two points into a simple time-share computer program, and out pop all the others. He now may check the 90-percent point, the 50-percent

TABLE 3—KNOWLEDGE OF PREVIOUS RESULTS					
Section	Number of Useable Response	Requested Range (percent)	Expected Number of Misses	Actual Number Average Misses	95-Percent Confidence Interval
New Orleans	143	90	1	5.46	5.08 to 5.84

TABLE 4—RESULTS USING FEEDBACK PROCESS					
Section	Number of Useable Response	Requested Range (percent)	Expected Number of Misses	Actual Number Average Misses	95-Percent Confidence Interval
Bay City	26	90	1	5.04	3.99 to 6.09
Bay City	26	50	5	8.31	7.67 to 8.95
Houston	98	90	1	4.05	3.63 to 4.47
Houston	98	50	5	7.32	6.94 to 7.70

point, or any other. He well may find some that do not fit his notions - for example, his 80-percent range does not yield a 40-percent range that suits him. So he compromises one or the other until he gets the fit he likes.

All that is design and theory. In practice, most people throw in the 80-percent range and just accept whatever comes out. Based on the recent testing with SPE-AIME groups, I would have to guess that the 80-percent range constructed without feedback is actually much more narrow - perhaps 50 percent. It would take a lot of data, which we do not have, to measure the range. Almanacs and encyclopedias cannot help much here.

My estimate of 50 percent comes from the following judgment. It must be more difficult to put ranges on exploration variables than to put them on questions such as when St. Augustine was founded. On the other hand, it should be easier for a geologist to conceive of his vocational uncertainties than for him to handle Beowulf-type questions. Since the audiences' average ranges on those two questions were about 40 and 85 percent, respectively, I chose 50 percent.

The feedback process, if used, can be of benefit. The following experiment was performed with some sections. I asked the members to write down two ranges simultaneously. That forced some sort of feedback. And since both ranges could not have 68-percent misses, it seemed logical to expect that such a ploy would yield better results - which, in fact, was what happened. (See **Table 4.**)

By having to use two ranges, the members were able to greatly improve their 90-percent range compared with those who worked with only one interval. The 50-percent range, however, was shoved in the other direction. I would guess that the best strategy for one faced with an uncertainty problem would be to consider whole distributions (that is, many ranges), continually playing one against the others. That scheme should result in even better definition of one's uncertainty.

Even then, studies suggest that people may come up short. I once saw the results of a full-scale risk analysis, including a probability distribution of project cost. A few months later the same people did another risk analysis on the very same project. Amazingly, the cost distributions did not even overlap. Changes had taken place on that project in the space of a few months that moved the results far beyond those contemplated when the experts were laying out their original ranges. People tend to build into their ranges those events that they can see as possibilities. But since much of our uncertainty comes from events we do not foresee, we end up with ranges that tend to be much too narrow.

Are the Tests Valid?

There may be those who still feel that the kinds of questions I used cannot be used as indicators of what one does in his own specialty. I know of several arguments to counter that view, but no proof.

The less one knows about a subject, the wider should be his range. An English scholar might have a 90 percent range of A.D. 700 to 730 for the "Beowulf" question. The typical engineer might recognize his limitations in the area and put A.D. 500 to 1500. Both ranges can be 90-percent ranges because the degree of uncertainty is a very personal thing. One's knowledge, or lack of it, should not affect his ability to use 90-percent ranges. So the type of question should not matter.

I mentioned earlier that SRI's use of material from a company's own annual report did not change the results. Regardless of whether one is an expert, the ranges generally come in too narrow.

Another criticism of these questions has been that they test one's memory of events already past rather than the ability to predict the future. Conceptually, is there any difference regarding the uncertainty? There may be more uncertainty associated with, for instance, the timing of an event yet to take place. But it seems that the difference is only one of degree when compared with recalling a date in history from an obscure and seldom-used brain cell. In either case, one does not know for sure and must resort to probability (likely a nontechnical variety) to express himself.

Bean Counting

You may find a third argument even more compelling. We asked groups of people to estimate the number of beans in a jar. Not only were they asked for their best-guess single number but also for a 90-percent range. The players were mostly professional people with technical training, and most had or were working part time on advanced degrees. Since we built in a reward system (money), the estimators were trying to do a good job, at least with their best guess. The following table gives their results. The jar contained 951 beans.

Best Guess	90-Percent Range
217	180 to 250
218	200 to 246
250	225 to 275
375	200 to 500
385	280 to 475
390	370 to 410
450	400 to 500
500	150 to 780
626	500 to 700
735	468 to 1,152
750	500 to 1,500
795	730 to 840
800	750 to 850
960	710 to 1,210
1,000	900 to 1,100
1,026	700 to 1,800
1,053	953 to 1,170
1,070	700 to 1,300
1,080	700 to 1,400
1,152	952 to 1,352
1,200	500 to 3,600
1,200	1,000 to 1,500
1,201	1,000 to 1,400
1,300	500 to 2,000
1,300	600 to 2,000
1,400	1,200 to 1,600
1,500	400 to 1,800
1,500	800 to 2,000
1,600	1,350 to 1,950
1,681	1,440 to 2,000
1,850	1,400 to 2,200
4,655	4,000 to 5,000
5,000	2,000 to 15,000

The experiment provides added insight because everyone could see the beans. No one had to test his memory of geography or history or his company's performance reports. The jar was somewhat square in cross-section so as not to introduce any tricks in estimating volume, though no one was allowed to use a ruler. Still, the requested 90-percent ranges turned out to be more like 36-percent

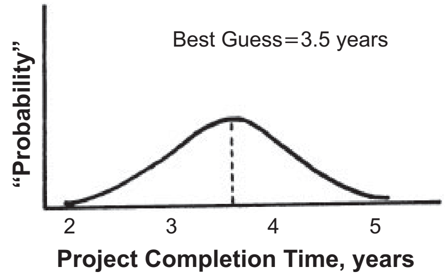

Fig. 2—Estimating with symmetrical uncertainty.

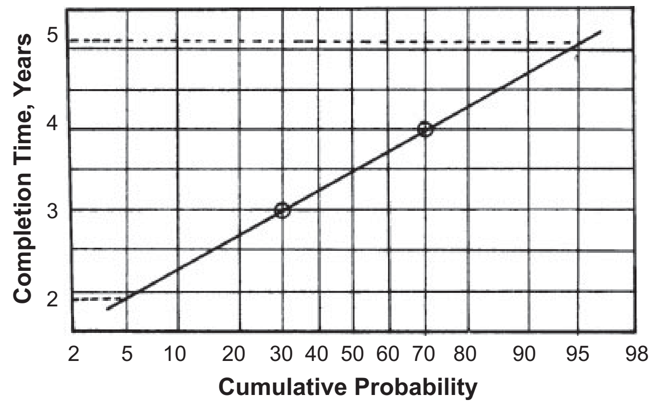

Fig. 3—Determining range, symmetrical uncertainty.

ranges because only 12 of 33 included the true value. After our testing, Elmer Dougherty of the U. of Southern California tried the same experiment and privately reported very similar results. We then asked some of our exploration people to go through the exercise, and they too repeated the earlier performances of others.

Interestingly, we have three more bean estimates made by people using a computer model (Monte Carlo simulation) to get ranges. They estimated their uncertainty on the components (length, width, heighth, and packing density) to get an over-all range. All included the true value of 951. Equally competent people not using the simulation approach could not do as well.

Best Guess	90-Percent Range
1,120	650 to 1,900
1,125	425 to 3,000
1,200	680 to 2,300

This experiment provides evidence that even a simple approach to probability modeling usually will be a lot better than what one dreams up in his head when it comes to assessing uncertainty.

Still More Experiments

Few people give in easily when confronted with this kind of material. They complain that I am testing groups and it was the "other guys" who caused the problems we see reflected in the data. Or they did not know my game was a serious one. Or they had no real incentive to do well, as they normally have on the job. Or that while they admit to having missed cost estimates, project completion times, producing rates, inflation rates, crude oil prices, etc., now and then, those were caused by external circumstances and certainly nothing they could have been responsible for. (Who ever said that we should only estimate that part of uncertainty for which we have responsibility?)

To counter such talk, I have engaged in other testing. One group had money riding on their ability to properly assess probability ranges. I asked them for 80-percent ranges and even agreed to pay them if, individually, they got between 60 and 90 percent. If they did not, they had to pay me. The group was so convinced the game was in their favor that they agreed to pay for the privilege of playing! And it was not sight unseen, either. They had already taken the test before the wager (same 10 questions given to SPE-AIME sections). They lost. But the point is that before getting their results, they did not feel that the questions were in some way beyond their capabilities.

At the SPE-AIME Fall Technical Conference and Exhibition in Dallas, I needed to save time while presenting this paper but I still needed to illustrate the point. I used a color slide of some beans spread about in an elliptical shape. It was the easiest test yet; the audience could clearly see every single bean. We used a 12- ft screen so the images would be large even for those in the rear. Still, only about one-third of the several hundred present came up with a 90-percent range that encompassed the true value.

As early as 1906, Cooke[2] did some testing of meteorological questions to see how well he could assess uncertainty. Since then, others[3,4] have examined the problem and noticed similar results. Lichtenstein *et al.*[5] have an extensive bibliography.

Don Wood of Atlantic Richfield Co. has been using a true/false test to study the phenomenon. The subject answers a question with true or false and then states the probability he thinks he is correct. Most people are far too sure of themselves. On those questions they say they have a 90-percent chance of answering correctly, the average score is about 65 percent.

To illustrate his findings, Wood describes the results on one of his test questions: "The deepest exploratory well in the United States is deeper than 31,000 ft." Several knowledgable oil men have said the statement is false and that they are 100-percent sure of their answer. Other oil men have said true, also believing they are 100-percent sure of being correct. Two petroleum engineers argued about another of Wood's questions: "John Wayne never won an academy award." Each was 100-percent sure of his answer, but one said true and the other said false. By the way, an Oklahoma wildcat has gone deeper than 31,000 ft and "True Grit" won an Oscar for the actor.

Where this paper reports results on how SPE-AIME groups act, Wood gives a test that has enough questions so that an individual can calibrate himself apart from any group. The grade one receives after taking the test may be loosely defined as the probability he knows what he is doing. It comes from a chi-square goodness-of-fit test on binomial data. Typical scores have been smaller than 1×10^{-5}, or less than 1 chance in 100,000.

Every test we have performed points in the same direction, as have most of the tests performed by others. The average smart, competent engineer is going to have a tough time coming up with reasonable probabilities for his analyses.

What Can We Do?

First, think of a range of uncertainty without putting any probability on that range. Since our sample showed that people tend to use the same range no matter what kind of range they were asked for, it seems plausible that a range such as we obtained during the tour would be forthcoming.

Having written it down, we arbitrarily assign some relatively small probability to the range encompassing the truth, say 40 percent. Decide on the form of the error. For example, in estimating project completion time, one may feel his uncertainty is symmetrical (± 6 months). (See **Fig. 2.**)

If the uncertainty is best expressed as symmetrical, then get some normal probability paper like that illustrated in **Fig. 3.** Plot the low end of your range at the 30-percent point and the high end at the 70-percent point. Note that 70 – 30 = 40. Your range has a 40-percent chance of encompassing the truth. Connect the points with a straight line and extend the line all the way across the paper. By reading the ordinates at the 5-percent and 95-percent points, you have your 90-percent range (95 – 5 = 90). Our ± 6 months has been converted to ± 1½ years. If that range seems uncomfortably large, good! Remember that if you are like most people, your natural tendency is to make such ranges too narrow. To repeat an earlier idea, uncertainty comes about because of what we do not know. Ranges constructed using what we do know are likely to be too small. (Bias, either pessimism or optimism, may be a problem too, but we have not addressed it here.)

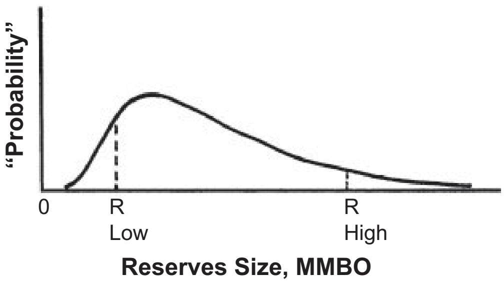

Fig. 4—Estimating with asymmetrical uncertainty.

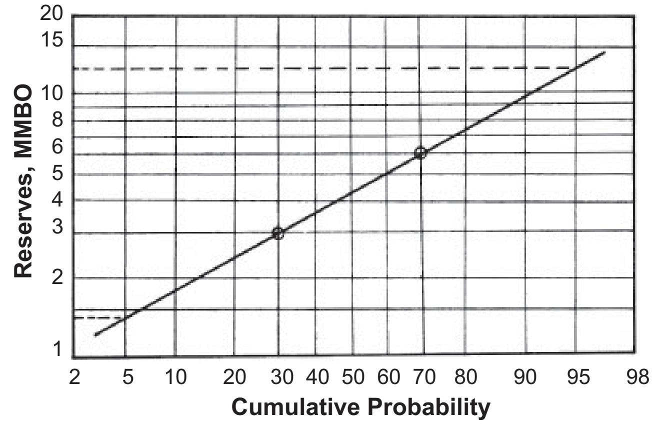

Fig. 5—Determining range, asymmetrical uncertainty.

You may feel the uncertainties are asymmetrical with a long tail region to the right, such as in estimating reserves (see **Fig. 4**). One cannot have less than 0 bbl, though with small probability he can have very large numbers.

In such cases, use log-probability paper as in **Fig. 5.** Say the range is 3 to 6 million bbl. Again, go through the ritual of plotting the low and high, drawing the line, and checking to see how comfortable you are with a 90-percent range. This time our range has been converted from 3 to 6 to something like 1.4 to 12. Discomfort is a good sign.

Because they fit so much of the world so well, the normal and lognormal distributions are logical choices for describing uncertainty. Do not worry a great deal about this apparent straight jacket. A realistic range (that is, wide) is often more important than the form of the distribution anyway.

Nor is there anything particularly holy about defining your original range as 40 percent. I could have used 50 or 30 percent. I am just proposing a simple way to get started in this business of defining the degree of your uncertainty and at the same time paying homage to the finding that people tend to overestimate the extent of their knowledge.

If each bean counter had plotted his range on log-probability paper as a 40-percent range and graphically determined his 90-percent range, 25 of the ranges (or 76 percent) would have included the true value of 951. Using such a technique, the group would have achieved a significant improvement in their ability to set ranges. After all, 76 percent is not that far from their target of 90 percent.

As you begin to keep records of your probability statements and compare them with actual outcomes, you will begin to build your own rules for making estimates. And, ultimately, your own tested rule is going to work better for you than anything others design.

The Value of Training

Winkler and Murphy[6] reported on some meteorologists who showed little or no bias in assessing probability. Training through years of almost immediate feedback on their predictions very likely accounts for this rare but enviable behavior. The oil business seldom allows such feedback. We may not find the answers to our predictions for several years, and by then we have been retired, promoted, banished, or worse.

But since training in this area appears to be vital, I urge you to set up a program for yourself. Every month make some predictions about the future, predictions whose outcome will become known during the following few weeks. Assign probabilities to your predictions, and religiously check your results. Find out what happens when you are 90-percent sure, 70-percent sure, etc. Example:

1. The next holiday weekend will see more highway deaths recorded than the similar period last year.
True 60 percent
2. The Cincinnati Reds will lead their division on July 4.
True 70 percent
3. XXX Corp. common stock will close above $Z before Sept. 1.
False 50 percent

To find out how well you are doing, consult some binomial probability tables (or a friendly expert). Say you had 20 statements to which you assigned a 70-percent chance of being right. You would have expected to get 14 of them right. What if you only got 10 right? Is that good? The tables show a probability of 4.8 percent of getting 10 or less right under conditions when you expect to get 14 right out of 20. It would be long odds (1 in 20) to claim, therefore, that you had learned to set the probabilities correctly. Better practice some more. Ask your stockbroker to do likewise.

Does a Better Range Lead to a Better Mean?

One might be tempted to argue that improving our understanding of uncertainty would not in itself improve the estimate of the mean, best guess, or whatever people tend to use for making their decision. But look, for example, at the Alyeska Pipeline and the 1969 cost estimate of $900 million. Most everyone associated with the project knew that it could not cost much less than $900 million. If everything had gone off without a hitch (roughly equivalent in probability of occurrence to all the molecules congregating on one side of a room), it might have come in for around $800 million.

What kind of things could happen to drive the cost in the other direction?

1. Labor problems such as jurisdictional disputes and the lack of an adequate supply of necessary skills in such a harsh environment.
2. Weather.
3. Shortages of equipment and supplies resulting from the unique nature of the project and remoteness of the site.
4. Design problems. An axiom of engineering: All doth not work that man designeth.
5. Economy of scale in full retreat. Some are so large that they are most difficult to effectively.
6. Bureaucratic delays brought about by masses of government regulations.

(Note that the list does not include the large cost increase brought about by government inflationary policies and the oil embargo, nor does it include the problems caused by so-called environmentalists. Reasonably intelligent forecasters might have missed those events back in early 1969.)

An analysis of these six items would have led one to imagine some chance for a pipeline costing as much as $3 billion giving the following range: rock-bottom cost = $0.8 billion; best estimate = $0.9 billion; high-side cost = $3.0 billion.

How long could such a "best guess" survive in such a range? Merely writing down the numbers exposes the best guess to sharp criticism and doubtless would force it to a higher and more realistic level. Though the new best guess would still have been far below present cost estimates of almost $8 billion, it nevertheless would have been very useful. Crude prices, we remember, were much lower then.

It seems logical, then, to expect that quite a number of projects would benefit similarly from a better range analysis. Consider the bean counters mentioned earlier. What if all those whose best guesses were less than 500 had known that there was a chance the truth might be up around 1,000? Is it not likely that they would have moved those best guesses up somewhat?

The Payoff

The payoff for having a better grasp on uncertainty should be quite a sum. In recent years both industry and government could have been more cautious in their estimates and perhaps achieved a better return for their investments.

The *Oil and Gas Journal* of Oct. 9, 1967, quoted management at the Great Canadian Oil Sands plant dedication: "Operating in the northland offers no unusual problems - in fact, it has some advantages." *Business Week,* Jan. 5, 1974, quoted the GCOS President: "We're the proud owners of a $90 million loss. This is the cost of being a pioneer."

Most tax payers remember the many government programs that ended up costing much more than original estimates (TFX, C5A, Interstate Highway Program, BART, and the Dallas-Fort Worth Regional Airport, for example). There has been a long history of cost underestimates for all kinds of projects because of not adequately accounting for future unknowns.

The whole planning and budget process stands at the mercy of supposedly expert estimates. It may be that we have gotten ourselves into trouble by looking for "the answer" (never attainable) when we should have concentrated on realistically setting our uncertainties. If the ranges are adequate, then at least the plan can cope with possible events of the future.

A better view of our uncertainties should have a significant effect on our success as risk takers and ultimately on profits.

Acknowledgments

I would like to thank the many SPE-AIME members who took the test and made this project so enjoyable. Also, I extend thanks to friends in Atlantic Richfield Co. who gave many helping hands.

References

1. Tversky, A. and Kahneman. D.: "Judgment Under Uncertainty: Heuristics and Biases," *Science* (1974) **185,** 1124–1131.
2. Cooke, W. E.: "Forecasts and Verifications in Western Australia." *Monthly Weather Review* (1906) **34,** 23–24.
3. Fischhoff, B. and Beyth, R.: " 'I Knew It Would Happen' - Remembered Probabilities of Once-Future Things," *Organizational Behavior and Human Performance* (1975) **13,** 1–16.
4. Lichtenstein, S. and Fischoff, B.: "Do Those Who Know More Also Know More About How Much They Know?," *Bull.,* Oregon Research Institute, Eugene (1976) **16,** 1.
5. Lichtenstein, S., Fischoff, B., and Phillips, L. D.: "Calibration of Probabilities: The State of the Art," *Bull.,* Oregon Research Institute, Eugene.
6. Winkler, R. L. and Murphy, A. H.: "Evaluation of Subjective Precipitation Probability Forecasts." *Proc., First National Conference on Statistical Meteorology,* American Meteorological Society (May 27–29, 1968) 148–157.

Appendix

Answers to the ten questions used in the quiz.

1. A.D. 1565.
2. 12.8 million (10.3 million autos).
3. 6,904 miles.
4. 1,901 miles.
5. 76.2 million people.
6. 4,200 ft.
7. 3.85 million sq miles.
8. 3,900 miles.
9. 248.4 years.
10. A.D. 700 to 730.

The answers to the questions came from the *Official Associated Press Almanac,* 1974 edition. Any source can be in error, and thus I discovered after the testing that I had been led astray on Question 2. The source said automobiles, but in checking other sources, I am now sure they meant motor vehicles. Strangely, the "new" answer does not affect our results very much. Most of those who missed that one were so far off that they were beyond help.

Original manuscript received in Society of Petroleum Engineers office July 27, 1975. Paper accepted for publication Jan. 8, 1976. Revised manuscript received April 19, 1976. Paper (SPE 5579) was first presented at the SPE-AIME 50th Annual Fall Technical Conference and Exhibition, held in Dallas, Sept. 28–Oct. 1. 1975.