Experience and Cognitive Biases in Oil and Gas Industry Personnel

Matthew Brian Welsh (matthew.welsh@adelaide.edu.au)
Australian School of Petroleum, University of Adelaide
Adelaide, SA 5005 AUSTRALIA

Abstract

Demonstrations of cognitive bias in expert decision makers are often criticised on the basis of their content being irrelevant to the participants. Herein two cognitive biases, anchoring and overconfidence, are shown to occur in batteries designed with questions relevant to the participants' domains of expertise. Results suggest that neither industry experience nor risk training removes the biases' effects. Instead, a relationship is found between the amount of time lapsed since such training occurred and calibration/overconfidence. This is taken as partial support for the proposition that regular training in risk, uncertainty and the mode of action of biases aids in remediation.

Introduction

The study of bias in oil and gas (O&G) industry decision making dates back to Capen's (1976) work in introducing the concept of overconfidence to O&G professionals. Despite this early start, however, the industry continues to report problems that seem to result from known cognitive biases (Hawkins, Coopersmith, & Cunningham, 2002; Shuenemeyer, 2002).

Biases, such as anchoring (Tversky & Kahneman, 1982) and overconfidence (Morgan & Henrion, 1990), are observed in the decisions of most people working under conditions of uncertainty and risk – defining features of O&G. Therefore, it is in the industry's best interests to follow research on risk, biases and their remediation (see, for example: Begg, Bratvold, & Campbell, 2002).

Despite this, and the best efforts of a segment of the industry, there remains resistance to the acceptance and application of findings. Many O&G personnel question the applicability of laboratory-based findings to their work in the field. Questions used in Capen's (1976) assessment of overconfidence among O&G personnel are described as "arcane" and thus, despite repeated demonstrations of overconfidence, some personnel feel that their actual work is beyond the reach of this type of bias. Similar objections have been raised about the generalisability of other biases.

To counter these objections, a discussion of two known biases and the evidence that they are impacting the O&G industry will be presented. The questions commonly used to assess these biases will be discussed, in terms of their function and how this relates to their form. Then, the creation of a new battery of bias questions designed using O&G examples will be discussed, along with expectations as to the effect that this will have on responses.

Following the presentation of the biases, the role of experience in improving decision making will be discussed, particularly in light of the sorts of training currently employed within the industry.

Bias in the Oil and Gas Industry

Anchoring Bias resulting from the use of the anchoring heuristic – people's tendency to base estimates on any value they have at hand, regardless of its relevance (Chapman & Johnson, 2002) - is also familiar to many O&G personnel.

A number of O&G companies, in fact, ask employees to estimate the end points of ranges before their best guess, to prevent anchoring on that guess and ending up with too narrow a range as suggested by Tversky & Kahneman (1982). Laboratory work, however, shows the opposite, with respondents who provide their best guess first ending up with wider ranges (Block & Harper, 1991; Clemen, 2001).

This, Rolle (2001) suggests, indicates a difference between experts and non-experts in their reaction to anchors. It is also reasonable to expect that susceptibility to anchoring is inversely related to the respondent's degree of knowledge of the true answer. Thus, experts, within their bailiwicks, may be less susceptible to anchoring than laypeople asked the same question.

The questions commonly used to assess anchoring come in two parts: the first is a greater/less than question about the value of interest; and the second directly asks the person for their best guess as to the true value. Half of a group are shown a high value as the anchor in the greater/less than question, and half a low (Tversky & Kahneman, 1982).

This format was chosen for the anchoring question used herein, with the only change being that the value being requested be something about which O&G personnel could make an educated guess, rather than having to rely on the anchor as their only cue – world oil reserves.

Overconfidence As noted above, overconfidence was the first cognitive bias openly discussed in O&G (Capen, 1976). It is, therefore, the best known bias, particularly as O&G exploration personnel are commonly required to give 80% ranges to describe parameters of a potential oil field and, historically, these ranges have been too narrow, precisely as the overconfidence bias predicts. Hawkins et al (2002), for example, report that 'actuals' fall outside the predicted 80% ranges more than half the time.

The relationship between this effect and the "arcane" questions asked by Capen (1976), however, remains difficult to convince people of and, thus, while they accept that overconfidence occurs in the real world and that they demonstrate this in the laboratory, they do not always make the connection between the two. This is despite the fact that, as these questions test the respondent's ability to state their own degree of uncertainty, the content is irrelevant (Capen, 1976).

While this is certainly true, it does ignore the fact that people, in general, feel frustrated answering a series of questions they have no reasonable chance of getting right – even though they are not required to do so. By moving the content of the questions into the field of the respondent, then, it is possible to reduce their discomfort by allowing them to make educated estimates.

To this end, for the questions used herein, all values around which participants were asked to construct 80% ranges were related to the O&G industry – whether to oil reserves, consumption or price – rather than asking geoscientists to place an 80% range around the date *Beowulf* was written, as did Capen (1976).

Experience and Remediation It is known that experts from certain fields are demonstrably less overconfident (i.e. better calibrated) than laypeople and experts from other fields. Generally this is held to be due to area-specific effects such as constant, timely feedback (Morgan & Henrion, 1990; Murphy & Winkler, 1977) but there seems to be a general feeling that experts should be less susceptible to bias due to their greater knowledge and experience in their field. Additionally, it is widely believed that training people in risk, uncertainty and biases reduces the effects of biases (Hawkins et al., 2002; Kahneman & Tversky, 1979; Morgan & Henrion, 1990).

Most of this research, however, looks only at immediate effects rather than the long term impact of such training on people's ability to deal with risk and uncertainty without succumbing to biases. By examining the relationship between lapse since training and bias susceptibility, it will become possible to make statements about people's retention of the concepts that are commonly delivered in industry short courses over a week or less.

In the absence of previous research on long term retention, the prediction must be that the longer the lapse since training, the more susceptible to bias a person will be – as per semantic memory decay (Bahrick, 1984).

Method

Participants

Demographic data was collected for 125 of 187 Oil and Gas industry personnel from whom responses were collected. Of these, 18 were female and the remaining 107 male, with a mean age of 39.8 years (SD 9.4) and an average of 13.5 years of industry experience (SD 9.1).

Questionnaire Design

The biases chosen for study in the questionnaire included anchoring, framing, overconfidence, unpacking and intuitive probability problems but, due to limitations of space, only anchoring and overconfidence are discussed here.

Two versions of the questionnaire were created to test for these biases – these being identical except as noted in the descriptions of the questions, below.

All of the questions were adapted from well established bias batteries (Bazerman, 2002; Plous, 1993), with their focus changed from managerial to O&G specific. Examples of these questions and the hypotheses they were designed to test are described below.

In addition to these bias questions, demographic questions were included, asking participants' gender, age and years of industry experience as well as querying whether they had had formal training in risk, uncertainty and cognitive biases and, if so, when this training had been undertaken. These were gathered to enable testing of hypotheses regarding the impact of experience and specific risk training on bias susceptibility.

Anchoring Question In both versions of the questionnaire, participants were asked to give their best estimate of 2003 world proved reserves**.** The questions differed only in the number shown in the preceding question: which asked whether reserves were greater or less than either 573.9 or 1721.6 Billion Barrels. In this way, the number included in the preceding greater/less than question was expected to act as an anchor, as described by Tversky & Kahneman (1982). These numbers were one-half and one-and-a-half times, respectively, the estimate found on the BP website.

Despite the privileged position of the participants – being asked questions where their expert knowledge could, presumably, be used to make educated guesses – the prediction was that the irrelevant information in the anchor would affect the estimates participants subsequently make. Further predictions were that greater experience would not greatly improve performance but that risk training and, particularly, recent risk training, would.

Overconfidence Questions Both versions asked participants to set 80% ranges around 10 values relating to the Oil and Gas industry – e.g., the United States' share of world oil consumption. A well-calibrated decision maker would, therefore, be expected to get around 8 out of 10 such questions right, whereas an overconfident person would get substantially fewer right (Morgan & Henrion, 1990).

Traditionally, overconfidence has been demonstrated using very general question sets. For reasons described above, this approach is not taken here but the overconfidence effect is expected to be observed despite the familiarity of the participants with the domain of the questions, as the parameter of interest is the participants' ability to judge the limits of their knowledge, rather than how much or how little they know about the topic. Further hypotheses were as described in the anchoring section above: experience is not expected to affect calibration greatly but risk training and recency are.

Recruitment

Participants were invited to take part in the study during visits to a number of Oil and Gas companies and Society of Petroleum Engineers meetings in Europe, South-East Asia, the United States and Australia. Approximately half of each group that agreed to take part was given the first of the two versions of the questionnaire described above, with the remainder receiving the alternate version.

While field research, such as this, results in a largely self-selected sample and tends to produce significant amounts of missing data, it was felt that this was preferable to the smaller sample that could be obtained by restricting recruitment to local companies. Nevertheless, conclusions must be leavened with these known limitations.

Results

Experience

Data from the 125 subjects for whom demographic information was collected indicated that almost half (61) had received training on risk, uncertainty and bias at some point in their careers. Unfortunately, only half of these participants (30) also indicated the time lapse since then.

For this subset, the mean time since training was 7.7 years (SD = 6.6), with answers ranging from 0 to 24 years.

Anchoring

Data was collected from 176 participants, 85 of whom saw the low anchor (573.9 Billion) and 91 the high (1721.6 Billion). Data from 5 participants were excluded as outliers, lying more than 1.5 times the interquartile range above the third quartile of their group. All of these were in the high anchor group with estimates ranging from 10 000 to 100 000 Billion barrels of oil.

Figure 1, below, shows the mean estimates (excluding outliers) for the groups seeing the low and high anchors.

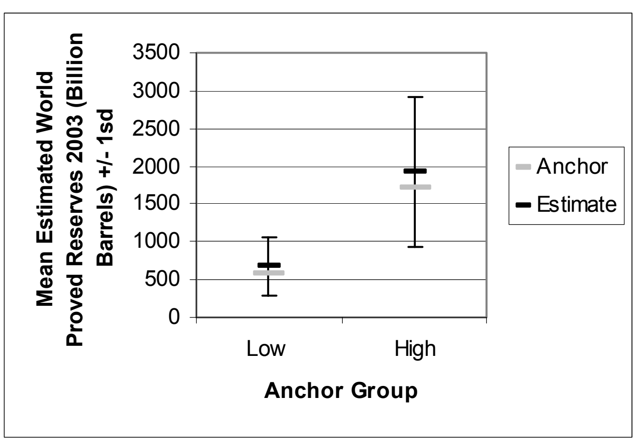

Figure 1: Effect of anchoring on estimated oil reserves.

Figure 1 clearly shows the strong effect that an anchor can have on estimates. The participants who saw the low anchor of 573.9 Billion barrels returned estimates averaging 682 Billion barrels, compared to the mean estimate of 1931 Billion Barrels from the group that saw the high anchor of 1721.6 Billion barrels.

Looking at the difference in standard deviations shown in Figure 1 and the distribution of estimates shown in Figure 2, below, it seemed clear that the two groups had unequal variances, and this was confirmed by an F-ratio, $F(85,84) = 6.84$, $p = 9.05 \times 10^{-17}$.

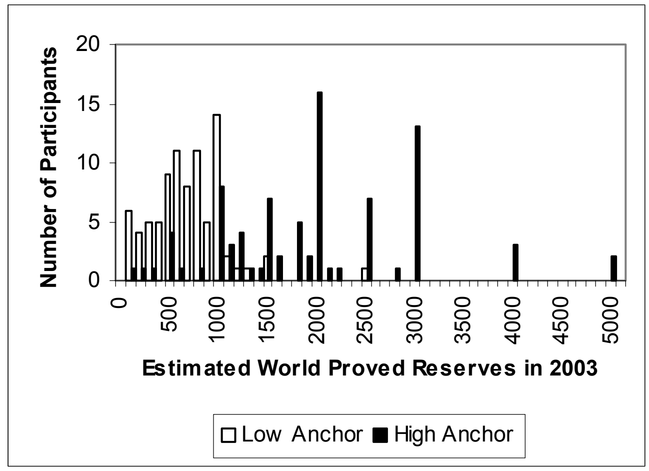

Figure 2: Oil reserves estimates by anchor group.

A heteroscedastic t-test was therefore used to test the hypothesis that anchor group had a significant effect on reserves estimation. This indicated that the difference between groups was highly significant, $t(110) = 10.85$, $p = 2.21 \times 10^{-19}$ two-tailed.

Anchoring and Experience In addition to the raw estimates described above, a measure of susceptibility to anchoring was calculated for each participant, as the absolute difference between the anchor and their estimate, as well as an accuracy measure, indicating the absolute difference between their estimate and the true value.

Correlations were calculated between anchoring susceptibility, anchoring accuracy and years of industry experience for the high and low anchor groups separately. Neither of the correlations between experience and anchoring susceptibility was significant, indicating that, as predicted, experience, in and of itself, is insufficient to offset the bias. The correlations between experience and accuracy for the two groups, however, were both negative and one approached significance, $r(83)=-0.207$ for the low group, $p=0.058$ two-tailed, and $r(84)=-.108$ for the high anchor groups, $p=0.322$ two-tailed, suggesting a possible, weak tendency for experienced people to provide more accurate estimates, as would be expected when asking questions about which participants can make educated guesses.

Dividing the low and high anchor groups according to whether participants had had risk training or not, it was observed that, in each of the four comparisons made (anchoring susceptibility and accuracy within each of the low and high anchor groups) participants with risk training performed better. That is, their estimates were less susceptible to the anchor and closer to the true value. A sign test indicated that the probability of this result arising from chance alone neared significance, p=0.063, but heteroscedastic t-tests showed no significant difference between any individual pairing.

Finally, correlations were calculated between: the time lapsed since risk training; and the measures of accuracy and susceptibility of both groups in the anchoring task. None of these approached significance, nor was there a clear pattern.

Overconfidence

Calibration data was collected from 123 participants as a score out of 10, measuring how often the true value fell within their 80% ranges. Figure 3 compares the distribution of participants' scores with that expected from perfectly calibrated participants. That is, the expected scores of people with an 80% chance of getting any question right.

Looking at Figure 3, one sees that even a perfectly calibrated person is only expected to get exactly 8 questions out of 10 around 30% of the time. Eight from 10 should, however, be the most likely outcome. By contrast, the mode of the distribution of participants' scores is 2 out of 10, scored by more than 15% of respondents.

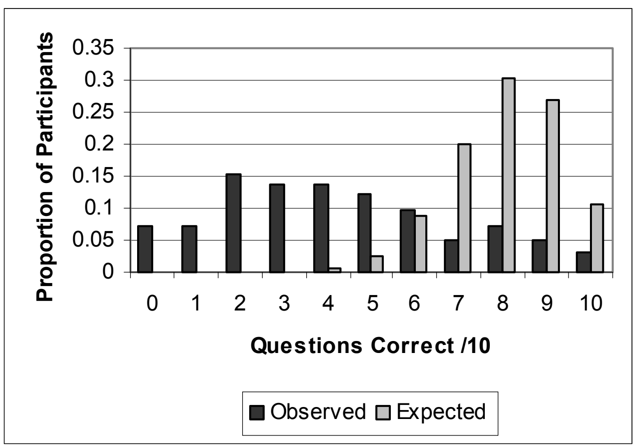

Figure 3: Observed vs Expected correct responses – 80% confidence calibration task.

This is clear evidence of overconfidence, with participants averaging only 4.2 correct from 10 questions and 58% of participants scoring 4 or less, compared to the expected proportion of less than 1%.

Overconfidence and Experience To determine the effect of experience on calibration/overconfidence, a number of calculations were performed.

Firstly, a correlation was calculated between years of experience and calibration score. While not significant, $r(120)=.120$, $p = .192$ two-tailed, the relationship is, again, in the predicted direction with more experienced personnel getting higher calibration scores.

Secondly, the mean calibration scores of the groups with and without risk training were compared by means of a t-test, which showed no significant difference, $t(120)=0.355$, $p = 0.72$ two-tailed. As was the case for anchoring, however, the mean calibration of the risk-trained group was higher than that of the untrained group (4.3 compared to 4.1 out of 10). That is, the relationship, while not significant, was, again, in the expected direction.

Finally, the correlation between time lapsed since risk training and absolute miscalibration (the difference between a participant's score and the ideal of 8 out of 10) was calculated. This indicated a significant relationship between the variables, $r(28)= .432$, $p=0.017$ two-tailed. This indicates that, amongst those people with risk training, those who had undertaken training more recently tended to get closer to 8 out of ten and were thus better calibrated, as shown in Figure 4, below.

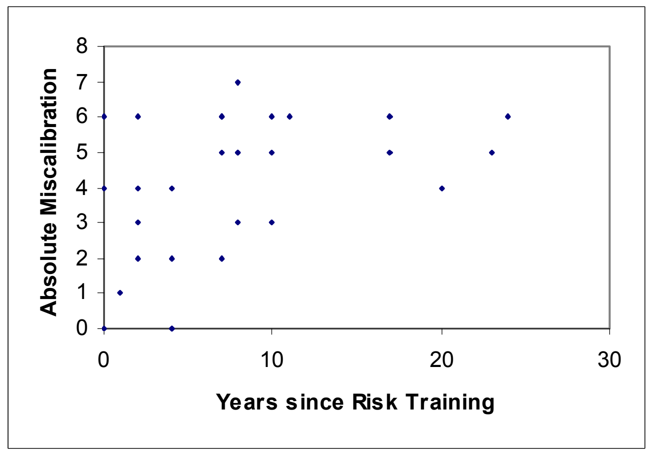

Figure 4: Scatterplot of Absolute Miscalibration and Years since Risk Training

Inter-Bias Relations

Correlations were also calculated between the participants' calibration score and both susceptibility and accuracy scores from the anchoring task. Neither correlation was significant, $r(111)= .002$ and $-.133$, $p = .985$ and $.159$ two-tailed, respectively. This suggests that the two biases have unrelated modes of action – in contrast to the claim by Tversky and Kahneman (1982) that overconfidence might be an effect of anchoring during the setting of bounds.

Discussion

Overview

In general, the predictions made regarding the expected relationships between the biases examined herein and experience, both general and risk-specific, held. While none of the results, in and of themselves, are impressively strong and few are significant, the pattern of results, with experience playing little role in preventing bias and risk training increasingly more as it becomes more recent, is as predicted – with the exceptions noted below.

Anchoring and Experience

There seems little doubt that the anchoring bias is observed in O&G personnel; even when asked questions about which they should be able to make educated guesses.

A possible criticism of this finding is that the anchoring question used herein was not representative of the domains in which most of the participants would be regarded as experts. In order to make it general enough for wide recruitment, it was made a test of O&G general knowledge rather than a serious test of domain specific expertise.

There were, however, weak results suggesting that estimates made by more experienced people were more accurate, which does offer some support for the idea that domain specific industry experience translates into a degree of expertise in the industry generally.

More clearly than this, it was shown that risk training had little effect on participants' susceptibility to anchoring. Although people with such training did perform better on both accuracy and susceptibility than untrained participants, these differences were uniformly small and not significant.

On the basis of this result, it might be suggested that risk training does not, in fact, aid in avoiding or reducing the anchoring bias. This conclusion, however, would be premature. Assuming that the participants who gave a figure for the time lapsed since their risk training are representative of the larger group, the average number of years since training is between seven and eight. It is possible that, if a sample included a larger proportion of more recent trainees, a stronger effect would be seen.

There is, however, a possible reason why risk training might *not* aid anchoring. In most industry literature (Bratvold, Begg, & Campbell, 2002; Clemen, 2001), the only description of anchoring comes as part of the discussion of overconfidence where it is said that thinking of a best guess first will anchor the end points of their confidence range.

There is, in general, no discussion of how to avoid being anchored on a figure when asked a question of the sort used herein. Indeed, given that the person is expected not to know the true answer, it may be hard to avoid anchoring. While it has been demonstrated that people will use irrelevant values as anchors (Chapman & Johnson, 2002), in most cases, in a question of the form used here, it *is* relevant as it sets the expected magnitude of answers. It may, therefore, be that asking the same question without an anchor would result in less accurate answers as the anchor limits answers to an appropriate order of magnitude.

In the experiment described above, for example, the anchors were selected to be *reasonable* possibilities, given that the respondents were known to have at least some knowledge of the range into which the true value should fall. In a test format, such as this, then, a state of tacit cooperation is assumed to exist between examiner and respondent that makes anchoring a rational tactic.

Overconfidence and Experience

There seems no doubt that the participants' calibration shows overconfidence as a general rule and any argument against participant expertise here is less concerning as calibration questions should be indifferent to this. O&G-related questions were used, primarily, to ease participants' concerns regarding their ability to make educated guesses, rather than to ensure questions from their fields of expertise.

As with the anchoring question, the relationship between experience and overconfidence was very weak but in the expected direction. By contrast, the relationship between overconfidence and time since risk training was significant, indicating that participants who had had such training more recently were more likely to be well calibrated. This effect, though, is still only weak to moderate, at best.

Further, while the correlation gives the impression of improvement, the question remains as to *how* the risk training is impacting responses. A minority of participants, for example, give such wide ranges that they end up with 10 out of 10. Hypothetically, astute participants could give 8 wide ranges and 2 narrow ones to get perfect calibration without knowing how to set accurate confidence bounds.

In some cases the use of these tactics is obvious enough –all ranges running from 0.01 to 1,000,000, for example – but, in other cases, it is hard to tell from honest ignorance. A more detailed look at participants' patterns of results over multiple sets of calibration questions would be required to sift out this effect and to determine whether risk training is teaching better confidence bound setting or just tactics to beat the system.

Another question arising from risk training is whether some courses teach counterproductive techniques such as avoiding initial best guesses in an attempt to avoid anchoring. In fact, providing such more often causes what has been called a cushioning effect (Welsh, Begg, Bratvold, & Lee, 2004) where an initial estimate 'pushes' later estimates away from it, leading to a wider range as is commonly observed in experiments (Block & Harper, 1991; Clemen, 2001).

Further Research

Given ongoing data collection, it should be possible to undertake additional enquiries into the nature of biases. For example, including a control group for the anchoring task will allow testing of the hypothesis that anchors are used because they are, in most cases, informative about the magnitude of the answer.

Likewise, additional participants will aid the comparison of calibration patterns between groups of participants with similar levels of recency of risk training to see if improvement results from better understanding or simply learning to beat the system.

Future studies may include more detailed examinations of the anchoring bias and its mode of action, in particular as it relates to the cushioning effect and alternate explanations of the potential effect of anchoring on overconfidence tasks.

Finally, a longitudinal study, tracking participants' susceptibility to bias over time after risk training, would test such training's efficacy. This would also allow tracking of calibration over time to see how training alters patterns of answering. Similarly, tracking the performance of personnel in their daily work would indicate whether the lessons had been incorporated or were only being applied in test situations.

Conclusion

While not providing a glowing endorsement of risk training's effects on susceptibility to anchoring and overconfidence biases, this study has suggested that trained people have a marginal advantage over untrained. Given the average time since training, however, the lack of a stronger advantage may be down to the fact that the difference in knowledge between the trained and untrained groups as defined herein was, itself, small.

There are, however, reasons to suspect that there is more to it than this – including the differences between the results for the two biases and the concerns raised above about the techniques used to train personnel during such courses.

It is also difficult to know, without further research, whether the lessons learnt in risk training courses are actually being transferred to the participant's work.

The conclusion, therefore, is that, while it appears that recent risk training may offer some advantage in bias-prone situations, this can not be pinned to a better understanding or application of principles without more research into these biases and their action in laboratory and real-world settings. It can, however, be concluded that O&G experience, in itself, offers little, if any, reduction in bias susceptibility.

Acknowledgments

I would like to thank Steve Begg and Reidar Bratvold for their advice and assistance during this project and Dan Navarro for his comments on statistical issues. This research was enabled through ExxonMobil and Santos' support of the Centre for Improved Business Performance at the Australian School of Petroleum.

References

Bahrick, H. P. (1984). Semantic memory content in permastore: 50 years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1-29.

Bazerman, M. H. (2002). *Judgment in managerial decision making* (5th ed.). New York: John Wiley and Sons.

Begg, S. H., Bratvold, R. B., & Campbell, J. M. (2002). *The value of flexibility in managing oil and gas investments.* Paper presented at the SPE paper 77586 at 2002 Annual Conference and Technical Exhibition, San Antonio, Texas, Sept 29 - Oct. 2.

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49*, 188-207.

Bratvold, R. B., Begg, S. H., & Campbell, J. M. (2002). *Would you know a good decision if you saw one?* Paper presented at the SPE paper 77509 at the 2002 Annual Conference and Technical Exhibition, San Antonio, Texas, Sept. 29 - Oct. 2.

Capen, E. C. (1976). The difficulty of assessing uncertainty. *Journal of Petroleum Technology*(August), 843-850.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: anchors in judgments of belief and value. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and Biases: the Psychology of Intuitive Judgment* (pp. 857). Cambridge: Cambridge University Press.

Clemen, R. L. (2001). Assessing 10-50-90s: a surprise. *Decision Analysis Newsletter, 20*(1), 2-15.

Hawkins, J. T., Coopersmith, E. M., & Cunningham, P. C. (2002). *Improving stochastic evaluations using objective data analysis and expert interviewing techniques.* Paper presented at the Society of Petroleum Engineers Annual techical Conference and Exhibition, San Antonio, Texas.

Kahneman, D., & Tversky, A. (1979). Intuitive predictions: Biases and corrective procedures. *TIMS Studies in Managment Science, 12*, 313-327.

Morgan, M. G., & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics, 26*(1), 41-47.

Plous, S. (1993). *The Psychology of Judgement and Decision Making* (Vol. 1). New York: McGraw-Hill.

Rolle, F. (2001). Assessing probabilties at DuPont - Maybe not a surprise. *Decision Analysis Newsletter, 20*(1), 3.

Shuenemeyer, J. H. (2002). A framework for expert judgment to assess oil and gas resources. *Natural Resources Research, 11*(2), 97-107.

Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: heuristics and biases. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 555). Cambridge: Cambridge University Press.

Welsh, M. B., Begg, S. H., Bratvold, R. B., & Lee, M. D. (2004). *Problems with the elicitation of uncertainty.* Paper presented at the Annual Technical Conference and Exhibition, Houston, Texas.