

## PUBLISHED VERSION

Welsh, Matthew Brian; Lee, Michael David; Begg, Stephen Hope [More-or-less elicitation \(MOLE\): Testing a heuristic elicitation method](#) Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci 2008), 23-26 July, 2008: pp. 493-498

© the authors

### PERMISSIONS

#### **correspondence from:**

Business Mgr

Cognitive Science Society Inc. [cogsci@psy.utexas.edu]

University of Texas - Austin

Department of Psychology

108 E. Dean Keeton, Stop A8000

Austin

**The copyright for articles and figures published in the Proceedings are held by the authors, not the Society**

<http://hdl.handle.net/2440/54280>

# More-Or-Less Elicitation (MOLE): Testing A Heuristic Elicitation Method

**Matthew B. Welsh (matthew.welsh@adelaide.edu.au)**

Australian School of Petroleum, North Terrace  
University of Adelaide, SA 5005, Australia

**Michael D. Lee (mdlee@uci.edu)**

Department of Cognitive Sciences, 3151 Social Sciences Plaza A  
University of California, Irvine, CA 92697, USA

**Steve H. Begg (steve.begg@adelaide.edu.au)**

Australian School of Petroleum, North Terrace  
University of Adelaide, SA 5005, Australia

## Abstract

Elicitation of people's knowledge is a central methodological challenge for psychology, with important impacts in many technical disciplines and industrial settings. The need to convert an expert's beliefs into a useable format is of particular importance when judgments and decisions are made under uncertainty. Simply asking a person for their best estimate or to estimate a range is subject to many biases – e.g., overconfidence – and methods for eliciting information that avoid these effects are required. This paper presents a heuristic-based elicitation method, More-Or-Less Elicitation (MOLE) which, rather than requiring people make absolute judgments, asks them to make repeated relative judgments. MOLE uses these, along with confidence statements, to construct probability distributions (pdfs) representing a person's beliefs. We evaluate MOLE by comparing these subjective pdfs with ranges elicited using traditional methods. The central finding is that use of MOLE greatly improves the accuracy and precision of elicited ranges, thereby reducing overconfidence. The benefits of this and other possible heuristic-based methods of elicitation are discussed.

**Keywords:** Elicitation, Uncertainty, Overconfidence, Heuristic.

Many technical disciplines share with psychological research the problem of eliciting information from people; that is, translating peoples' beliefs into useable data (Wolfson, 2001). Of particular interest is how to best achieve this under uncertainty, where there is no single, correct answer but rather the "correct" response for an individual to make will vary according to their own level of knowledge about the topic (Morgan & Henrion, 1990).

The reason so much interest is vested in this area is that, despite elicitation's ubiquity, argument continues about the best way to elicit information and people are still subject to many biases, so that elicited responses are less accurate than elicitors would wish (see, e.g., Hawkins, Coopersmith, & Cunningham, 2002; Lichtenstein, Fischhoff, & Phillips, 1982; Welsh, Begg, & Bratvold, 2007).

This paper discusses elicitation and some of its problems. Further, it proposes, as a possible solution, the use of *heuristic*-based methods – that is, elicitation methods based on simple judgments, such as which of two options is closer to the true value (i.e., the value the participant believes the stimulus takes). Such a method, called More-Or-Less

Elicitation (MOLE), is tested and discussed herein.

## Elicitation

The elicitation of uncertainty is the conversion of individual or group's beliefs into a probability distribution. Generally, this is done not for its own sake but to, for example, predict future outcome ranges, or provide inputs for forecasting models (Morgan & Keith, 1995).

In order to be of benefit, elicited values need to be accurate. Accuracy, in this case, however, refers to two separate ideas. The first sense, which we might call objective accuracy, is the one that naturally springs to mind: elicited values need to accurately reflect the probability of an event occurring. Equally important, however, is subjective accuracy: how well elicited values map onto an elicitée's beliefs. The problem for elicitors is that the two are not easily separated. Instead we have to rely on relatively crude measures like overconfidence/calibration scores (Lichtenstein et al., 1982), which primarily measure objective accuracy even though, from the elicitor's point of view, a measure of subjective accuracy is being sought.

## Problems for Elicitation

Standard findings in the elicitation literature are that people's best guesses are anchored by previously seen values and that they are overconfident, producing too narrow ranges of possible outcomes (Tversky & Kahneman, 1974). There is also evidence, however, that this is not entirely due to inaccuracies in people's beliefs. Specifically, different elicitation techniques result in different responses; Winman, Hansson and Juslin (2004), for example, demonstrate that having people evaluate a range rather than produce one leads to less overconfidence in their responses.

There are also concerns about the effect of question order within an elicitation task. These date back to Tversky and Kahneman's (1974) paper, where they suggested that anchoring on an initial best guess might be a cause of overconfidence. Research into this idea, however, has been mixed with, for example, Russo and Schoemaker (1992) finding the predicted effect but Block and Harper (1991) and Juslin, Wennerholm and Olsson (1999) finding the opposite. To complicate matters further, there are concerns regarding the level of control over question order in some of

these studies. For example, Block and Harper (1991) used answer booklets which, while having questions in a set order, could not insure they were answered in that order.

### Debiasing Elicitation

Given these problems, significant work has gone into attempts to debias elicited values. Early work (summarised in Morgan & Henrion, 1990), however, indicated little success in reducing overconfidence and none for anchoring.

As noted above, however, there are some techniques known to reduce overconfidence, including Winman et al's (2004) use of interval assessment, and the use of long-term repeated feedback (Lichtenstein et al., 1982). Additionally, the use by expert elicitors of counterintuitive examples (lying outside the initial range) has been shown to be effective in reducing overconfidence (Hawkins et al., 2002). This remedy, though, requires an expert elicitor to be on hand to ask the right sorts of questions and leaves open the question of whether simply drawing people's attention to regions of the possibility space outside their initial range is helpful in the absence of expertise.

Regardless, none of these techniques eliminates overconfidence – excepting specific cases such as weather forecasting, where repeated feedback seems to have resulted in good calibration (Murphy & Winkler, 1977).

### Heuristic Elicitation

Given the problems with elicitation and the observation that question format has a large impact on the elicited responses, it is worth considering more radical departures from the standard elicitation methods. For example, the work of Gigerenzer and others (Gigerenzer & Selten, 2001; Gigerenzer & Todd, 1999) on bounded rationality has yielded insights into the sorts of questions that the human mind seems most comfortable working with.

One observation is that people are better at making relative judgments than absolute ones (Gigerenzer & Selten, 2001). This is consistent with Winman et al's (2004) observation that people are better at evaluating than generating ranges. Combining this insight with the observation that counter-intuitive examples can reduce overconfidence (Hawkins et al., 2002) leads to the idea that asking a series of questions covering the range of possibility, rather than allowing a person to hone in on a small region of outcomes, thereby excluding other possibilities, may yield better results.

The idea of such a heuristic-based elicitation method – using relative judgments – was first explored in Welsh, Begg, Bratvold and Lee (2004). This found a benefit but relied heavily on assumptions about the underlying distribution required to create a probability distribution from the relative judgments. The current goal was, thus, to create an elicitation method that makes minimal assumptions in a principled manner to produce the final distribution.

We aim for an elicitation method that is less subject to overconfidence than alternative methods requiring direct estimation of values. We are also interested in whether requiring a best guess first reduces or increases the width of

estimated ranges, and whether simply drawing people's attention to values outside their initial range is sufficient to widen those ranges.

## Method

### Participants

Participants were 40 undergraduate students from the University of Adelaide. Four, however, were excluded due to computer errors during testing leaving 36 (10 male and 26 female) with a mean age of 20.1 ( $SD = 1.9$ ).

### Materials

Four graphical user interfaces (GUIs) were developed to enable automated testing of participants using each of the elicitation methods chosen for examination. All of the GUIs displayed an array of circles, from 100 to 300 (determined randomly at each trial) and elicited the participant's beliefs regarding the number of circles - in accordance with the varying elicitation techniques.

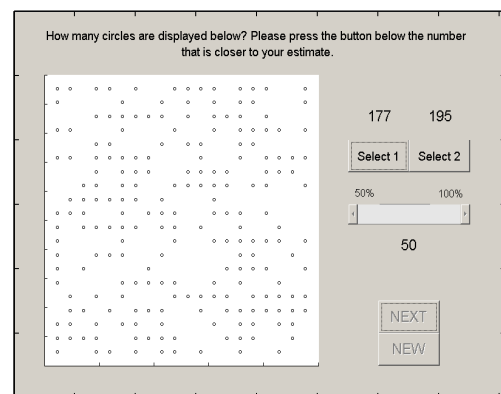


Figure 1. MOLE GUI

For each of the elicitation techniques, the same basic GUI layout was used, with only the questions being asked and the buttons that could be used to respond being different. For example, Figure 1 shows the layout as seen during More-or-Less Elicitation (MOLE) condition, asking participants to select which of two values is closer to their estimate. The GUI controls were sequentially locked and unlocked to ensure that participants answered each question before continuing to the next. This ensured that participants completed the questions in the prescribed order.

### Procedure

Participants were tested on four elicitation methods, described below. Participants, over the course of an hour, completed ten trials under each condition after being sorted at random into four groups to allow counterbalancing for possible order/learning effects as shown in Table 1.

### Simple Elicitation

In this condition, participants were asked to provide a

minimum and maximum value for the number of circles. Following this, they indicated how confident they were that their range contained the true value. This was done using a slider similar to the one seen in Figure 1 but capable of taking any integer value from 0 to 100%.

Table 1. Ordering of Elicitation Methods

Group	Elicitation Methods			
A	1	2	3	4
B	4	3	2	1
C	2	4	1	3
D	3	1	4	2

Note: 1=Simple, 2=Triangular, 3=Iterative, 4=MOLE

### Triangular Elicitation

In this condition, participants were asked to provide a best guess prior to giving their minimum and maximum values – thereby providing sufficient information to produce a triangular distribution. Again, after making estimates, they were asked to indicate their confidence on a 0-100% scale.

### Iterative Elicitation

In this condition, participants were asked to provide an initial range as in the Simple Elicitation condition but then shown values for the minimum and maximum that lay outside their own range - which were described as having been elicited from “previous participants” but which actually were always calculated by the program to lie outside the initial range (60% of the initial minimum and 140% of the initial maximum). Participants were then given the chance to adjust their estimates of the minimum and maximum. Once happy with their estimates, they were asked to indicate their level of confidence that the true value would fall inside their range on a 0 to 100% range.

### More-Or-Less Elicitation

In the MOLE condition, participants did not directly estimate values. Rather, they selected which alternative in randomly generated pairs of values from a range from 0 to 400 was closer to their estimate. After each choice, participants were asked to indicate their confidence that their selection was actually closer to the true value than the alternative on a 50% (guessing) to 100% (certain) range.

This process was repeated 10 times during each trial and the final range of feasible values recorded (i.e., those the participant’s answers did not rule out). Additionally, the confidence ratings were used to create a subjective PDF as described below.

Whenever a confidence rating of 100% was given, any values lying closer to the unchosen value were excluded from the experimental range and then weight of 0.5 was added uniformly across the remaining range. If, however, the confidence level was less than 100%, weight was added to each end of the range separately according to the level of confidence. Figure 2 shows how two stages of this process might progress, starting with a range of possible value from 90 to 150. In the top half of the figure, the person has been shown two values: 135 and 150 (highlighted) and stated with 100% confidence that 135 is closer to the true value.

This means that values above 142.5 (the midway point of 135 and 150) will no longer be considered. An equal weight of 0.5 is then applied across the entire remaining range, indicating ignorance about where in that range the person believes the true value lies.

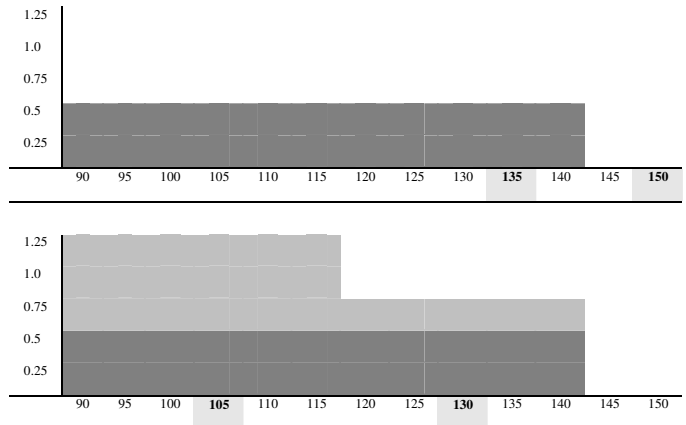


Figure 2. Subjective PDF construction process.

In the lower half of Figure 2, the person is then shown the values of 105 and 130 and states that they are 75% confident that 105 is closer to the true value. This results in a weight of 0.75 being applied from the current minimum up to the midpoint of the two values (117.5) and a weight of 0.25 from the midpoint up to the current maximum – reflecting the fact that the person’s confidence statement indicates that they believe a value closer to the lower option is three times as likely as one closer to the high option.

In this way, over the course of a trial, a PDF was built up. At the end of each trial, this PDF was corrected by removing all weight from areas outside the final feasible range and then adjusted by subtracting 99% of the lowest weight from all remaining areas. Finally, the Beta-distribution that minimized summed squared differences from the resultant PDF was calculated.

## Results

As described above, while overconfidence is generally used as the primary measure of the efficacy of an elicitation method of the sorts used herein, this can be further divided into the accuracy and the precision of the elicited responses. Results relating to the primary hypothesis are therefore described below in terms of all three concepts: overall overconfidence, precision and accuracy.

### Overconfidence

Overconfidence, in terms of elicited ranges, is measured from the degree of ‘coverage’ achieved (i.e, how often the elicited range was correct - that is, contained the true value) and participants’ stated levels of confidence. Table 2 shows this data for each of the four conditions.

It is clear from Table 2 that all three techniques requiring participants to estimate absolute ranges resulted in less than 30% coverage, despite the stated confidence of the

participants averaging more than 70%. By comparison, the MOLE, with its assumed 100% confidence level, resulted in 90.6% coverage. (The confidence level is ‘assumed’ as participants in the MOLE condition did not directly rate the likelihood of the true value falling within their final range, rather it was assumed that their final range contained all of the values they considered feasible.)

Table 2. Coverage and mean confidence rating by condition.

Condition	Correct	Trials <sup>#</sup>	Coverage	Confidence
Simple	92	340	27.1%	75.5%
Triangular	81	340	23.8%	72.6%
Iterative	97	340	28.5%	72.7%
MOLE	308	340	90.6%	100%*

# - 20 of the 360 trials were excluded as individual analyses indicated that participants had either misinterpreted the experimental instruction or were deliberately answering incorrectly in order to limit their participation time. \* - assumed confidence level.

To determine whether the differences between the methods were statistically significant, a repeated measures ANOVA was conducted. The first indicated that there were significant differences between the number of hits achieved by participants under the four conditions,  $F(3, 83) = 123.8$ ,  $p < .001$ . Paired sample t-tests with Bonferroni corrections were used for each unique pair of elicitation methods to determine which conditions differed from the others and these indicated that only the MOLE condition differed significantly,  $t(35) = 14.1$ , 15.7 and 12.4 (from the Simple, Triangular and Iterative, respectively),  $p < .001$  in all cases.

The question remained, however, as to whether the improvement in calibration in the MOLE data resulted from an improvement in precision, accuracy, or both.

### Precision

Precision reflects the subjective aspect of accuracy (i.e., how well a response matches the person’s knowledge and beliefs). The primary measure of precision in an elicited range is its width. Figure 3 shows the mean range width under each of the methods described above, along with the variability, as measured by the standard deviation.

Looking at Figure 3, it seems clear that an improvement in the appropriateness of participants’ levels of precision plays a significant role in the observed reduction in overconfidence. Specifically, participants’ responses to the MOLE technique are far less precise, giving much wider ranges on average than in any of the other three conditions.

The implication of this is that participants in the other conditions were *too* precise. That is, their ranges were far narrower than their level of knowledge warranted. A repeated measures ANOVA, run using the participants’ mean ranges in each condition, confirmed that the difference was highly significant,  $F(2, 59) = 75.9$ ,  $p < .001$ . Once again, paired sample t-tests with Bonferroni corrections were used post-hoc confirming that only the MOLE results differed from the other conditions,  $t(35) = 8.6$ , 12.1 and 9.4

for comparisons with the Simple, Triangular and Iterative methods, respectively,  $p < .001$  in all cases.

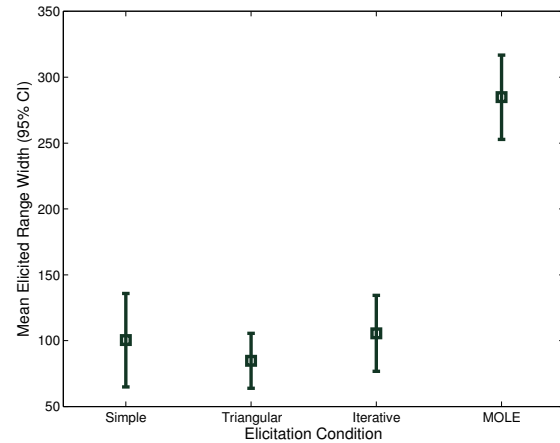


Figure 3. Mean width of range by elicitation method.

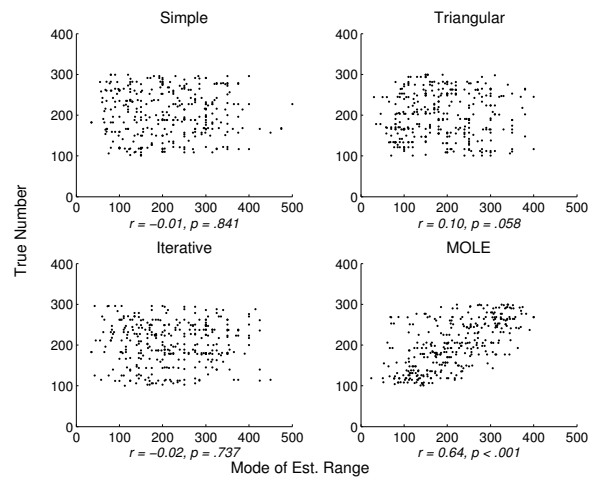


Figure 4. Scatterplots comparing number of objects with the estimated mode for each elicitation condition.

### Accuracy

To assess the objective accuracy of participants’ responses under each elicitation condition, the mode of each elicited range was compared with the true value. For the Simple and Iterative elicitation conditions, a uniform distribution was assumed and thus the mean (midpoint) was substituted for the mode. For the Triangular, the mode was the “most likely” value given by the participant. Finally, for the MOLE, the mode was calculated using the  $\alpha$  and  $\beta$  values calculated from the beta distribution most closely fitting a participant’s subjective PDF, giving  $M = (\alpha - 1) / (\alpha + \beta - 2)$ . Scatterplots showing these data are shown as Figure 4.

Figure 4 suggests that only in the MOLE condition did participant estimates accurately track the number of objects in the stimuli. The correlation between the means of the estimated range and true values was moderately high and highly significant,  $r(338) = 0.64$ ,  $p < .001$ , whereas correlations between the true values and the remaining elicited means were all extremely low,  $r(338) = -0.01$ ,  $-0.10$  and  $-0.02$  for the Simple, Triangular and Iterative method

respectively,  $p > .05$  in all cases.

## Other Findings

### Best Guesses and Overconfidence

One of our initial research questions asked whether requiring participants to give a best guess prior to fixing their confidence interval's end-points would affect its width and thus their levels of overconfidence.

Looking at the data in Table 2 and Figure 3, one sees little difference between the ranges provided in the two conditions of interest (Simple and Triangular). While participants in the Triangular condition gave, on average, narrower ranges ( $M = 84.7$ ,  $SD = 61.8$ ) than they did in the Simple condition ( $M = 100.3$ ,  $SD = 105.0$ ) the confidence intervals in Figure 3 indicate no significant difference between these values. A repeated measures ANOVA, similarly, compared the mean levels of confidence indicated by participants in the three conditions where this was directly assessed (all but the MOLE) and this found no significant differences between the conditions,  $F(2,67) = 2.26$ ,  $p = .112$ . Even were the differences significant, however, overconfidence would not be greatly affected as the 3.3% decrease in the number of hits is offset by the 2.9% decrease in stated levels of confidence.

### Iterative Elicitation

The final research question related to whether an automated system would be effective in prompting participants to reconsider and widen their ranges. Looking again at Table 2 and Figure 3, one sees that there seems to be a weak effect in line with expectations. Participants' ranges in the Iterative condition were wider than in the Simple condition ( $M = 105.5$ ,  $SD = 85.0$  compared to  $M = 100.3$ ,  $SD = 105.0$ ) but not significantly so as examination of the CIs in Figure 3 shows. Similarly, the difference in confidence, while noticeable in Table 2, is not significant – as the repeated measures ANOVA described above indicated.

## Discussion

Our results show a clear benefit to the use of the MOLE heuristic elicitation technique in terms of both the precision and the accuracy of elicited ranges. We found little support, however, for the role of initial best guesses or simplistic emphasis on counter-intuitive values in improving elicitation. These results are discussed in greater depth below.

### Heuristic Elicitation

It seems reasonable to conclude that elicitation techniques enabling people to use the well-honed, heuristic judgment and decision tools already at their disposal are powerful tools for reducing bias. The degree of overconfidence observed in the MOLE responses was much smaller than in the other conditions, particularly given that when asked for a wide confidence interval (80% plus), people tend to give ~50% intervals (Morgan & Henrion, 1990) and the MOLE generated a 100% interval.

Of greater interest is the fact that this method works not just by causing people to consider more values, thereby including a wider range of possibilities (i.e., increasing subjective accuracy by helping people realize the limits of their knowledge) but also by improving their objective accuracy. That is, while participants in the other conditions proved poor at estimating the true number of objects displayed, repeatedly judging which option was closer to the true value resulted in participants in the MOLE condition having a better idea of what that true value was.

This supports the idea from the bounded rationality (Gigerenzer & Selten, 2001) literature that enabling people to answer questions in formats they are adept with is a good way to avoid bias in judgment and decision making.

### Limitations

There are several caveats, however, regarding the current MOLE method. It could, for example, be argued that the MOLE gave participants an unfair advantage in that it limited the range of values that the contrast values could be selected from to between 0 and 400 (remembering that the true value was always between 100 and 300). Figure 4 shows that, in the non-heuristic conditions, a number of estimates lie beyond this range, meaning that participants had the opportunity to be more inaccurate than in the MOLE. That said, the vast majority (98.1%) of estimates from all other conditions fell within that range – it having been chosen as a reasonable estimate of the range of responses – so any effect from this would be limited. The other relevant result for this issue is the qualitative difference in the scatterplots, which shows better accuracy in the MOLE condition across the full range of stimuli.

Additionally, the need for bounds limits the usefulness of the MOLE, as currently formalized, to situations where limits can be put on what people might believe (although these limits can be very broad due to the MOLE's iterative narrowing of the range to exclude infeasible values). There remains, however, a risk of limiting the outcomes a person is allowed to choose. That said, this should not be a problem in many applied domains where expert knowledge is being sought - where there are, often, known limits on outcomes.

Finally, the MOLE requires participants to spend more time observing the stimulus and thus some of the effect may simply be noise reduction – although this would seem only to explain improvements in accuracy, not precision. This also has the effect of increasing the effort required per trial with the resultant problem that more participants gave nonsensical answers indicative of random button pushing in the MOLE (18 of the 20 excluded trials). This is, however, unlikely to cause a problem in applied setting as large numbers of values tend not to be elicited simultaneously.

### Future Directions

Given our findings, it seems worthwhile to continue looking at heuristic-based elicitation as a method for avoiding bias in elicited responses. In addition to extending its use to non-visual elicitation tasks, an obvious direction is to refine the MOLE procedure such that it can automatically determine if

a person has reached the limits of their certainty, rather than requiring a set number of questions.

The application of this approach to other biases that impact on elicited responses such as anchoring would also be of interest - given how resistant to debiasing anchoring has proved (Wilson, Houston, Etling, & Brekke, 1996).

Finally, while we believe the focus on relative judgments is an important advance in developing elicitation methods, it should be possible to improve the way in which subjective PDFs are generated. The current method was chosen so as to minimize the assumptions needing to be made, but remains somewhat ad hoc in nature. One interesting possibility is to follow the recent lead of Sanborn and Griffiths (in press), who apply modern computational Bayesian sampling algorithms, based on Markov-Chain Monte Carlo methods, as experimental procedures for understanding the subjective probability distributions people use to represent mental categories. Applying the same principled ideas to the problem of value elicitation is a promising direction for future research.

### Other Issues

While previous work has found estimated ranges to be either narrowed (Russo & Schoemaker, 1992) or widened (Block & Harper, 1991) by initial best guesses, the present study, despite stronger control over question order, has found no clear effect - although a weak trend was seen towards narrowed ranges resulting from initial best guesses. As such, this remains an open research question, with further work required to tease out the intricacies of this variable effect.

Neither did we find any evidence that simply indicating to people that other people had made estimates well outside their own range had any impact on revisions of those estimates. It could, however, be that participants realized that these "other participants" were computer generated and that future research will determine what sort and how much counter-intuitive evidence people need to provoke them into changing their mind - or whether this only occurs with the presence of a known expert (Hawkins et al., 2002).

### Conclusion

Heuristic-based elicitation methods seem to be a worthwhile addition to the arsenal of researchers interested in reducing the impact of bias on elicited responses. While the fine detail still requires further refinement, the basic premise, of using relative rather absolute judgments, is strongly supported by the findings herein and the concept seems well placed to contribute to our understanding of how our cognitive abilities give rise to bias and to aid in improving the accuracy of forecasting in a variety of areas.

### Acknowledgments

MBW is supported by ExxonMobil and Santos through their support of the CIBP at the Australian School of Petroleum. The authors wish to thank Ben Schultz for his assistance in data collection, Dan Navarro for his assistance with the code and the anonymous reviewers for their comments.

### References

- 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.
- Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded Rationality: the adaptive toolbox* (Vol. 84). Cambridge: Massachusetts: MIT Press.
- Gigerenzer, G., & Todd, P. M. (Eds.). (1999). *Simple heuristics that make us smart*. Oxford, UK: Oxford University Press.
- 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 78th Annual Technical Conference and Exhibition, San Antonio, Texas.
- Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 1038-1052.
- Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Morgan, M. G., & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.
- Morgan, M. G., & Keith, D. W. (1995). Subjective judgements by climate experts. *Environmental Science and Technology*, 29(10), 468A-476A.
- Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics*, 26(1), 41-47.
- Russo, E. J., & Schoemaker, P. J. H. (1992). Managing Overconfidence. *Sloan Management Review*, 33, 7-17.
- Sanborn, A. N., & Griffiths, T. L. (in press). Markov Chain Monte Carlo with People. In B. Scholkopf, J. Platt & T. Hoffman (Eds.), *Advances in Neural Information Processing* (Vol. 20). Cambridge: MA: MIT Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Welsh, M. B., Begg, S. H., & Bratvold, R. B. (2007). Efficacy of bias awareness in debiasing oil and gas judgments. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Meeting of the Cognitive Science Society* (pp. 1647-1652). Austin, Texas: Cognitive Science Society.
- Welsh, M. B., Begg, S. H., Bratvold, R. B., & Lee, M. D. (2004). SPE 90338: Problems with the elicitation of uncertainty. *Proceedings of the Society of Petroleum Engineers 80th Annual Technical Conference and Exhibition, Houston, Texas: SPE*.
- Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 125(4), 387-402.
- Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: how to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(6), 1167-1175.
- Wolfson, L. J. (2001). Elicitation of probabilities and probability distributions. In E. Science (Ed.), *International Encyclopedia of the Social Sciences* (pp. 4413-4417): Elsevier Science.