Number Preference, Precision and Implicit Confidence

Matthew B. Welsh[1], Daniel J. Navarro[2] & Steve H. Begg[1]
([matthew.welsh, daniel.navarro, steve.begg]@adelaide.edu.au)
[1]Australian School of Petroleum, [2]School of Psychology
University of Adelaide, North Terrace
Adelaide, SA, 5005, Australia

Abstract

In elicitation tasks, people are asked to make estimates under conditions of uncertainty but elicitors then interpret these estimates as if the estimator were *certain* of them. An analysis of people's patterns of responding during the elicitation of uncertainty, indicates that there are markers of confidence incorporated into these estimates that can be used to predict the person's true level of confidence. One such marker is the precision (number of significant figures) of the estimate. Analyses of elicited data show the expected positive relationships between accuracy, precision and explicit confidence and, further, that precision offers information beyond that of explicit confidence ratings. We then demonstrate the importance of incorporating this information on an overconfidence task, showing that it can account for a 9% difference in calibration.

Keywords: Number preference, confidence, precision, elicitation, judgment and decision making.

Introduction

Studies of human judgment typically make use of estimates of some quantity given by participants, with a view to assessing the "quality" of these estimates. However, exactly how to make this assessment is not always straightforward as people's estimates can contain more information than just a numerical value. For example, imagine that you have asked two individuals how high Mt Everest is. The first answers "9km"; while the second responds "8925 metres". Later you have the opportunity to check the true answer and find that Mt Everest is 8844.43m high (PRCSBSM, 2005). Which of the two is a better estimate?

One answer, of course, is that the second estimate (8925m) is better as it missed the precisely measured value by only 80.57m whereas the first estimate missed by 155.57m. In terms of human interactions, however, the answer is less clear. While the second answer is closer to the true value than the first, it is also far more precise –stating the height to the nearest metre. The first estimate, by comparison, it is stated only to the nearest kilometre.

The inference a listener might draw from these different levels of precision is that the first speaker is giving an approximate height while the second is giving an exact height – an distinction referred to by Yaniv and Foster (1995) as "graininess". Generally, the less precise an estimate is, then, the less confidence we expect the estimator to have in their estimate being precisely right. This conversational rule mimics the rules of measurement used in the physical sciences where values are given with an error range of ± half the smallest calibration of the measurement device. Thus, a ruler marked in millimetres yields measurements that are presumed to be accurate to ±0.5mm.

Thinking in these, pragmatic terms (Sperber & Wilson, 1986), one could conclude, therefore, that the second estimate is, in fact, worse. This is because it is precise to the nearest metre but the true value lies more than 80 metres beyond the 8924.5 to 8925.5m interval resulting from the addition of an appropriate error. The first estimate, by comparison, implies a range of 8.5 to 9.5km and the true value falls well within this.

The conclusion to be drawn from the above is that the consideration of precision can alter our perceptions of accuracy. Although seemingly unremarkable, this has important implications for the way in which we should interpret estimates given by participants during elicitation procedures, as discussed below.

Elicitation of Uncertainty

The elicitation of uncertainty describes the process of converting a person's subjective beliefs regarding uncertain events into a numerical form to allow easier analysis (Wolfson, 2001). Various techniques designed to do this are used where probabilistic forecasting is required in fields such as Petroleum Exploration (Attanasi & Schuenemeyer, 2002), Hydrology (Krzysztofowicz, 2001) and Meteorology (Morgan & Keith, 1995).

The technique most commonly used in the oil and gas industry, for example, is the elicitation of 80% confidence ranges (see, e.g., Hawkins, Coopersmith, & Cunningham, 2002). Here the elicitee is asked to give a range of values such that they are 80% certain that the true value of whatever parameter they are estimating will fall within it.

Overconfidence

The common observation of people using elicitation techniques, however, is that people are overconfident (Lichtenstein, Fischhoff, & Phillips, 1982) – that is, they give ranges that are too narrow, such that values fall outside their 80% ranges more than the expected 20% of times.

Given this tendency of people to be overconfident in their estimates, it is not surprising that much of the literature on uncertainty elicitation relates directly to mechanisms for overcoming uncertainty or "debiasing" participants. Various techniques from simple advice to widen ranges (Lichtenstein, et al., 1982) through repeated feedback (Murphy & Winkler, 1977) to the use of probabilistic games (Hawkins, et al., 2002) are recommended. The common observation, however, is that such techniques reduce but do not eliminate overconfidence (Morgan & Henrion, 1990).

A particularly interesting observation from the overconfidence literature is that the strength of the effect is greatly impacted by format dependence (Juslin, Wennerholm, & Olsson, 1999). For example, the same participant will give a different answer when asked to *generate* a confidence interval than when asked to *evaluate* that same interval (Winman, Hansson, & Juslin, 2004). Thus, a person who has set ten 80% confidence ranges, when asked how many times the true value will fall within their specified ranges, may answer only “65%”.

This discrepancy is generally taken to indicate a problem with the participant’s understanding of the statistical underpinnings of the processes. If they were accurately setting their 80% confidence intervals, they should expect approximately 80% of values to fall within those ranges but, instead they predict that fewer than this will.

Winman et al (2004) provide a possible explanation for this, where they argue that it results from the statistical naivety of participants, relying on biased estimators of dispersion (Fiedler, 2000). A simulation of the effect, on overconfidence, of using sample variances to estimate population variance can be seen in Welsh, Begg, Bratvold and Lee (2004) where it is shown that sample sizes in the range of human short-term memory limitations do appear to lead to overconfident estimates of population dispersion.

While this approach has the advantage of mathematically corresponding to people’s observed behavior, it also requires that people think in peculiarly statistical ways. Specifically, for overconfidence to be the result of sampling from memory, assumptions must be made about the nature of memory and recall that do not necessarily accord with mnemonic theory and experimental results (for a discussion of this, see, e.g., Bruza, Welsh, & Navarro, 2008).

Precision in Elicited Values

The difference between interval estimation and interval evaluation can also be considered in another way, invoking the concept of precision described above. To understand why this is, first it must be understood that people, when asked to estimate values, answer in a restricted fashion. Specifically, they show number preference (Baird, Lewis, & Romer, 1970; Plug, 1977), preferring to give answers that are integers and also multiples of 5 or 10.

These number preferences are sometimes interpreted as resulting from their ease of use in the decimal system or other psychological preferences (Albers, 1999) and this does seem likely to account for part of the effect, at least, but it is also feasible that people use rounded, imprecise numbers because they are, implicitly, giving imprecise estimates.

Consider a case where a participant gives a range of possible values for the parameter of interest of 100-500. Exactly how reasonable is it to believe that the end-points of this elicited range, which could theoretically take any value, both fell on multiples of 100 by chance? Rather, where participants repeatedly give these rounded numbers, interpreting these estimates as also reflecting an implicit measure of confidence makes it possible to give an alternative explanation of format dependence, as described in our case study below.

Research Aims

The initial aim of this research is to confirm our expectation that people will display number preferences due, in part at least, to a desire (implicit or explicit) to reflect their uncertainty about the magnitude of the value they are estimating. If people are, in fact, using imprecise numbers in this manner, it should lead to a number of observable tendencies. For example, assuming that confidence and accuracy are related, more accurate people should also be more precise in their estimates.

Secondly, and perhaps more importantly, we aim to show why people who use elicited values need to take this additional information into account when examining people’s elicited responses.

Thus, analyses undertaken here tested whether a person’s tendency to use rounded numbers (i.e., multiples of 10, 100, etc) correlated with both the accuracy of, and their stated degree of confidence in, their estimate – the expectation being that people would be more accurate and confident when giving more precise answers. Then we examined a pre-existing data-set to demonstrate how including this effect alters our conclusions about the magnitude of one of the most studied cognitive biases, overconfidence.

Pilot Work

Prior to the current experiment, we ran two pilot studies looking at this effect. The first asked 36 University of Adelaide students (4 male, mean age = 22.7, SD = 4.8) to estimate answers to 20 general knowledge questions. The first 10 of these had no explicit confidence rating while the second 10 did. All questions had 4-digit answers.

This study established that asking for an explicit confidence rating did not alter people’s use of precision but struck significant problems with the levels of confidence observed. People found the questions very difficult and their confidence ratings averaged less than 2 (on a 0 to 10 scale). As a result, while precision correlated with the accuracy of estimates at 0.51, the relationship between confidence and the other measures were very weak at 0.15 and 0.12 for precision and accuracy, respectively.

To avoid this restricted range of confidence, a second analysis examined a small, pre-existing dataset, to test whether number preferences (precision) were observed in a memory task. The data-set was from an unpublished anchoring experiment in which 15 university graduates (5 male, mean age = 31.5, SD = 7.4) had responded to 54 questions (all with numerical answers but of varying magnitudes). This indicated that, when people had previous experience of the facts about which they were later asked, their confidence in their estimates was much higher, but that they still used precision as a marker of accuracy (correlation of 0.34). Precision was also observed to have separate predictive power to the confidence ratings, with the partial correlation between precision and accuracy remaining at 0.25 after controlling for confidence.

Method

Participants

Participants were 40 university students and members of the general public, recruited in and around the University of Adelaide, 27 male, with a mean age of 25.4 (SD = 9.3). Participants were given a $10 book voucher for their participation, with an additional $20 voucher offered as a reward for the most accurate participant.

Materials and Procedure

Materials. 40 almanac-style questions of fact were selected from across a range of topics. All questions had numerical answers that were 4 digits in length (i.e., between 1000 and 9999) and none ended in a zero.

Procedure. Testing was divided into a learning and a testing phase – both computerized and presented via graphical user interfaces (GUIs) designed in Matlab.

During the learning phase, participants were presented with all forty questions, rewritten as statements of fact. They were allowed to look at each of these for as long as they chose before continuing to the next but could not, thereafter, return to look at the same fact again. There was then a two minute break while the experimenter closed the learning GUI and opened the testing GUI.

The testing GUI then presented 20 of the 40 questions (the same 20 for all participants), one at a time, asking participants to enter their answer to the question directly into the GUI and then to indicate how confident they were in that answer using a slider that took values from 0 to 10.

The questions in the learning and testing phases were in the same order for all participants but the two phases had different question orders. Participants were tested individually and most completed the task within 30 minutes.

Results

Measures

Participant performance was measured in three ways. First, their accuracy on the questions was measured – as the absolute percentage error in their estimates. This was assigned a negative value so that higher values correspond to higher accuracy.

Second, their confidence in each estimate was recorded – this being simply their explicit confidence rating from 0 (low confidence) to 10 (high confidence).

Finally, the degree of precision at which they had answered the question was recorded. This measure was simply the number of zeros that their estimate ended in. Thus a fully precise answer, ending in a non-zero digit, was scored ‘0’, whereas an answer that was a multiple of ten scored ‘1’, and a multiple of one hundred ‘2’. Precision scores within the sample ranged from 0 to 3. These values were then inverted such that high values correspond to high precision and vice versa.

Analyses

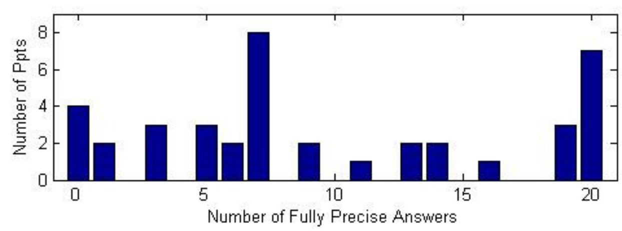
Given that we had 40 participants all complete the same twenty questions, we aggregate data at the participant level and report, for example, the distribution of participants' average accuracy across the 20 questions. Similarly, correlations between our measures of interest are calculated for each participant and the distributions of these discussed.

Number Preferences

The first question we asked of the data was: are people showing number preferences? That is, even in this experiment where *none* of the true answers that the participants saw ended in a zero, would people still report answers ending in zeros or would they, instead, always given fully precise responses?

Figure 1 shows how often participants gave fully specified (to the last digit) responses to the 20 questions. Looking here, one can see that, despite the fact that all questions had answers specified to the last digit, there are strong preferences toward estimates ending in zeros. While there are seven participants in Figure 1 who always gave answers that were precise to the last digit (the peak at the far right), one can see that the majority of people gave some or even all of their answers rounded to the nearest ten (or hundred or thousand).The average number of fully precise answers per participant was 9.75 – slightly less than half.

Figure 1. Histogram showing to how many of the twenty questions participants gave fully precise responses (i.e., no final zeros).



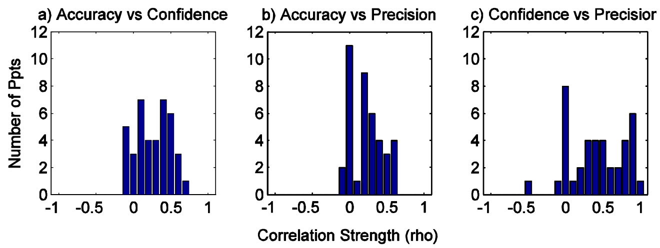
This is, of course, far fewer than one would expect by chance – assuming a 1 in 10 chance of an estimate ending in a zero, the probability of seeing 410/800 estimates ending in zeros is vanishingly small, $p \approx 2.3 \times 10^{-189}$ – so it seems uncontroversial to conclude that number preferences are observed in the sample.

Accuracy, Confidence and Precision

As an initial test of the relationships between accuracy, confidence and precision, we calculated rank-order correlations between the three variables for each participant. These are summarized in Figure 2.

Looking at Figure 2, one sees relationships that are, generally, in the expected directions in all three subplots. Participants whose estimates were more accurate tended to be more confident in those answers (34 of 40 correlations being positive, mean ρ = .27). Similarly, people who gave more precise answers tended to be more accurate (29 of 40 correlations being positive, mean ρ = .22). Finally, confidence and precision are related in a straightforward manner – with high confidence tending to be partnered with high precision (31 of 40 correlations, mean ρ = .42).

Figure 2. Histograms of rank order correlation strengths between each of Accuracy, Precision and Confidence for 40 participants.



In all cases, a sign test indicates that the probability of seeing so many positive correlations in the absence of a genuine effect is very low, p = 6.9×10^{-7}, .001 and 9.1×10^{-5}, respectively.

There are, however, discrepancies in Figure 2 that need further explanation; specifically, the peaks at zero in subplots b) and c). These are caused, primarily, by the minority of people who always gave precise responses and who, therefore, have a zero correlation between their precision scores and both accuracy and confidence.

An important question to ask here, however, is whether these peaks represent those people who remembered the true answers and were, therefore, able to give highly accurate and precise answers or whether they reflect an alternative estimation strategy that avoids the rounded numbers preferred by most people.

To establish this, we examined the accuracy and confidence of the participants within this subset of participants. While statistical analyses on so small a group (7 participants) are extremely unlikely to demonstrate a convincing difference, we noted that the mean error of the 'always precise' subgroup was actually higher than that of the remainder of the sample (23.3% vs 12.6% error) and their confidence was lower (4.7 vs 5.1). That is, the members of the 'always precise' group were both *less* accurate and *less* confident. These effects are very large and very small, respectively, A = 0.95 and 0.52 (this is a non-parametric, probability-based effect size measure which indicates the likelihood of a randomly chosen person from one group outperforming a randomly chosen person from the other group; see Ruscio, 2008, for a full explanation).

Thus, it seems reasonable to conclude that these 'always precise' people are not actually the most accurate but rather seem to have a different estimation strategy from the remainder of the sample. Rather than using rounding to represent their uncertainty, these people engage in what might be thought of as random number entry – entering seemingly precise but actually meaningless final digits when they aren't sure of what the final digit should be.

Of course, an alternative explanation might be that these people simply recognized that there were no values ending in zero in the learning phase. This is argued against by the pilot data, however, where a similar group was seen in experiments where some answers did end in zero. That is, even where zero was a possible value, some people seemed to indicate uncertainty by entering random strings of digits (such as runs across the keyboard – 123, etc).

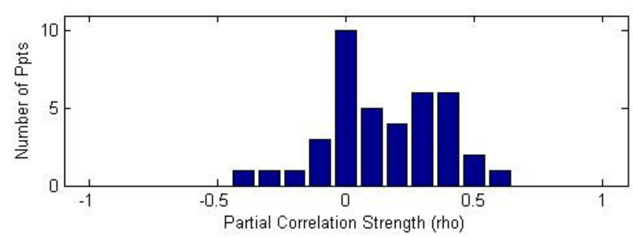
Confidence versus Precision

A secondary question that needs to be asked is whether confidence and precision are, in fact, just two measures of the same thing – the person's underlying confidence in their answer. Even if this were the case, of course, an understanding of how people use precision to flag their underlying confidence in an estimate is useful for those situations where an explicit confidence rating has not been gathered. Of greater interest, though, is whether, even with confidence ratings, examination of people's use of precision adds further information.

The observation above regarding the subgroup who do not use precision at all argues for this conclusion as, within that group, people's confidence scores still varied despite their precision scores all being the same. That is, for at least some people, precision and explicit confidence are different.

To test for any separate relationship between accuracy and precision, partial correlations were calculated, for each participant, between these variables, controlling for the effect of confidence. Figure 3 shows the distribution of partial correlations from our sample of 40 participants.

Figure 3. Histogram showing 40 participants' partial correlations between accuracy and precision, controlling for confidence.



Looking at Figure 3, one sees that, even controlling for confidence, the correlations between accuracy and precision remain mostly positive (27 of 40), which a sign test signals as unlikely in the absence of a positive relationship, p = .008. Excluding those people who never change their precision (the spike at zero in Figure 3), the average partial correlation between accuracy and precision is 0.18. That is, the results suggest a weak but consistent effect. By comparison, the average partial correlation between confidence and accuracy, controlling for precision, is only 0.12 – with or without the 'always precise' subset of participants.

Discussion

The results of this experiment reconfirm those from the pilot work described above. Number preferences, in the form of the precision at which people choose to answer a question, have clear implications for how accurate we should expect that answer to be and how confident a person is in it.

While the strength of the correlations in our results are quite weak, we note that our experiment was an artificial situation where *none* of the true answers that the participants saw prior to testing ended in zero, creating a situation where the effects of precision would be weakest - as this provided the strongest test of the effect's existence.

We therefore expect that, in other experimental designs and, in particular, where uncertainty is greater, the effect will be magnified – as was observed in our first pilot study.

Case Study: Overconfidence

If, as the above results suggest, the majority of people use round numbers in a pragmatic manner to indicate the degree of confidence that they place in an estimate, then this has clear implications for decision making research where people give estimates under uncertainty.

For example, following up on the example given in the introduction, if a person in an overconfidence task has given end points for an estimated range of 100 and 500, exactly how confident should we be that they intend for these values to be interpreted as precise? That is, when they say they are 80% sure that the true value falls between 100 and 500, do they mean precisely that or something closer to "I am 80% sure that the true value falls within a range from *something like 100* to *something like 500*".

Interpreting such responses in line with the second meaning requires a reconsideration of results from previous overconfidence experiments. For example, assuming the pragmatic rule from the natural science – that is, an estimate is good to one-half the smallest specified unit - we would have to acknowledge that a 100-500 range might, in the mind of its generator, include values as low as 50 or as high as 550. Therefore, if we fail to take into account the precision at which responses are given, we may inadvertently inflate overconfidence.

By way of example, we applied this simple rule to a prior dataset (Welsh, Bratvold, & Begg, 2005) looking at overconfidence effects in 80% confidence intervals elicited from 123 petroleum industry professionals. This study initially concluded that the participants were overconfident, with only 42% of estimated 80% confidence ranges including the true value.

Due to the high degree of uncertainty in the participants' estimates, however, more than 95% of estimated ranges in this sample were bounded by imprecise estimates (multiples of 10, 100, 100, etc) and, as a result, when we applied the pragmatic rule of including one-half of the smallest specified unit to each end of the range, most ranges were widened. As a result, calibration increased by 9% (to 51%).

The point of this is not that calibration increased – as it was almost certain to – but rather demonstrating how large an effect this can have in decision making under uncertainty and, thus, that experimenters need to consider this as a source of *apparent* overconfidence. Of particular interest is the fact that this difference is of similar magnitude to that observed by Winman et al (2004) when comparing people's evaluation and generation of confidence intervals in calibration tasks. It, therefore, seems possible, given the above demonstration, that this effect is largely the result of researchers misinterpreting people's responses. That is, if people interpret numbers given to them in an evaluation task as precise (as one might expect given the nature of the task) but naturally generate imprecise end-points for their own ranges, then this might account for the majority of the difference in 'overconfidence' between generated and evaluated ranges.

While a 9% change in calibration seems modest, it should be kept in mind that this can equate to tens of millions of dollars in industrial decision making. Welsh, Begg and Bratvold (2007), for example, discuss the economic significance of overconfidence on an offshore oil and gas development project, noting that even a 5% change in calibration can change cost/profit estimates by more than $22 million.

General Discussion

The data from the experiment described herein (and both pilot studies) offer support for the idea that number preferences, in the form of the precision with which a person answers a question, may reflect that person's underlying confidence in their estimates. More specifically, it seems that the majority of people use the precision of their estimates to convey some sense of how accurate they believe their estimates to be.

Interestingly, the effect of precision, while clearly overlapping the information provided by explicit confidence, also carried additional information in the main study and both of our pilots. That is, even when an explicit confidence rating has been obtained, it remains beneficial to examine peoples' precision if one wishes to understand how good an estimate they believe they have provided.

This is affirmed by the results of our case study, which shows the marked difference that the inclusion of this information makes to the interpretation of data gathered in a typical overconfidence experiment.

Future Research

Given these findings, there are a number of directions that seem worthwhile pursuing. The first of these involves an area that we have skirted here – people's meta-knowledge regarding confidence and their use of precision. That is, are people aware of the way in which they use precision as a marker? At this point, we would predict that the answer is: no; because if people were aware of their use of imprecise numbers then one would expect precision and explicit confidence to be measuring exactly the same thing, which appears not to be the case. Thus, consideration of number preferences seems to offer a method for gaining insight into metacognitive processing in future research.

A secondary question revolves around the use of multiples of 5, which are also known to be disproportionately used in estimation tasks. While we did not include these – as several of our answers ended in "5" - this additional number preference should be taken into account in future to truly nail down the effect.

We also need to look at the opposite face of the pragmatic conversation informed by peoples' use of precision. That is, having established that people use precision as a marker of confidence, we need to confirm whether other people accurately interpret these numbers.

More generally, our results point to a need for consideration of individual differences in decision making. Cognitive biases like overconfidence are often reported as group effects but a reconsideration of what is happening at the individual level can shed light on the processes giving rise to these biases (see, e.g., Welsh & Navarro, 2007).

Conclusion

People use the precision of their estimates as a marker of how accurate they believe they are. Given this, researchers relying on elicited values need to take this effect into account if we are to understand the responses we are given.

In particular, it seems likely that the degree of overconfidence in elicited ranges may have been over-estimated as a result of researchers not paying enough attention to the pragmatic aspects of the communication between researchers and participants.

To paraphrase Inigo Montoya (from *The Princess Bride*): "That number, I do not think it means what you think it means."

Acknowledgments

MBW and SHB are supported through ExxonMobil and Santos' funding of the Improved Business Performance group at the ASP. DJN is supported by an ARC Discovery Grant Fellowship (DP0773794). The authors would like to thank Atanu Saha for his assistance in collecting the data.

References

Albers, W. (1999). Prominence theory as a tool to model boundedly rational decisions. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: the adaptive toolbox*. Cambridge, MA: MIT Press.

Attanasi, E. D., & Schuenemeyer, J. H. (2002). Some aspects of resource uncertainty and their economic consequences in assessment of the 1002 area of the Arctic National Wildlife Refuge. *Natural Resources Research, 11*(2), 109-120.

Baird, J. C., Lewis, C., & Romer, D. (1970). Relative frequencies of numerical responses in ratio estimation. *Perception and Psychophysics, 6*, 78-80.

Bruza, B., Welsh, M. B., & Navarro, D. J. (2008). Does Memory Mediate Susceptibility to Cognitive Biases? Implications of the Decision-by-Sampling Theory. In V. Sloutsky, B. Love & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1498-1503). Austin, TX: Cognitive Science Society.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*, 659-676.

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.

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology, 249*, 2-9.

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.

Plug, C. (1977). Number preferences in ratio estimation and constant-sum scaling. *American Journal of Psychology, 90*(4), 699-704.

PRCSBSM (2005). Mt Qomolangma stands at 8844.43 (Peoples' Republic of China State Bureau of Surveying and Mapping) Retrieved 8th December, 2005, from http://en.wikipedia.org/wiki/Mt_Everest.

Ruscio, J. (2008). A probability-based measure of effect size: robustness to base rates and other factors. *Psychological Methods, 13*(1), 19-30.

Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Blackwell.

Welsh, M. B., Begg, S. H., & Bratvold, R. B. (2007). SPE 110765: Modeling the economic impact of cognitive biases on oil and gas decisions. *Proceedings of the Society of Petroleum Engineers 83rd Annual Technical Conference and Exhibition*.

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*.

Welsh, M. B., Bratvold, R. B., & Begg, S. H. (2005). SPE 96423 - Cognitive biases in the petroleum industry: impact and remediation. *Proceedings of the Society of Petroleum Engineers 81st Annual Technical Conference and Exhibition*.

Welsh, M. B., & Navarro, D. J. (2007). Seeing is believing: priors, trust and base rate neglect. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Meeting of the Cognitive Science Society* (pp. 701-706). Austin, Texas: Cognitive Science Society.

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.

Yaniv, I., & Foster, D. D. (1995). Graininess of judgment under uncertainty: an accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124*(4), 424-432.