Title
Predicting Overprecision in Range Estimation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Kaesler, Matthew
Welsh, Matthew B.
Semmler, Carolyn

Publication Date
2016

Peer reviewed

Predicting Overprecision in Range Estimation

Matthew Kaesler[1] (matt.kaesler@gmail.com)

Matthew B. Welsh[1] (matthew.welsh@adelaide.edu.au)

Carolyn Semmler[1] (carolyn.semmler@adelaide.edu.au)
1. The University of Adelaide, SA 5005, Australia

Abstract

Overprecision (overconfidence in interval estimation) is a bias with clear implications for economic outcomes in industries reliant on forecasting possible ranges for future prices and unknown states of nature - such as mineral and petroleum exploration. Prior research has shown the ranges people provide are too narrow given the knowledge they have – that is, they underestimate uncertainty and are overconfident in their knowledge. The underlying causes of this bias are, however, still unclear and individual differences research has shed little light on traits predictive of susceptibility. Taking this as a starting point, this paper directly contrasts the Naïve Sampling Model and Informativeness-Accuracy Tradeoff accounts of overprecision – seeing which better predicts performance in an interval estimation task. This was achieved by identifying traits associated with these theories – Short Term Memory and Need for Cognitive Closure, respectively. Analyses indicate that NFCC but not STM predicts interval width and thus, potentially, impacts overprecision.

Keywords: confidence; overprecision; need for cognitive closure; STM; informativeness; naïve sampling model.

Introduction

Overprecision, the form overconfidence observed on interval estimation tasks, has been described as the most robust yet least understood form of overconfidence (Moore & Healey, 2008). It occurs where people provide confidence intervals (lower and upper bounds between which they are confident, to a stated degree, that an unknown value lies). If, over a set of questions asking for (e.g.) 90% confidence intervals, objective accuracy levels are lower than 90%, this is deemed overprecision (Alpert & Raffia, 1982; Moore & Healey, 2008). It differs from overestimation, observed on point estimates when post-item confidence judgements exceed accuracy, as point estimate accuracy improves with task experience, while range estimate accuracy does not, implying different underlying processes (Hansson, Juslin & Winman, 2008). People show pronounced overprecision on interval estimation tasks - that is, far fewer ranges contain the true value than expected based on the stated level of confidence (Lichtenstein, Fischhoff & Phillips, 1982). Overprecision is of applied significance in engineering, mining and the oil and gas industry, which all make use of estimates delivered in this form (Moore, Tenney & Haran, 2015), with valuation errors of hundreds of millions of dollars resulting (Welsh, Begg & Bratvold, 2007).

The basic cause of overprecision is people generating intervals too narrow to reflect their degree of subjective uncertainty, failing to capture the true value as often as expected (Alpert & Raffia, 1982). The effect is resistant to debiasing, with participants exhorted to widen their intervals failing to do so enough to achieve good, much less perfect, calibration (e.g. Yaniv & Foster, 1995).

Factors contributing to overprecision, however, are poorly understood and, while task characteristics and debiasing techniques have attracted significant interest, individual differences in performance have received little attention, with only two recent studies relating individual differences to interval estimation performance (Haran, Ritov & Mellers, 2013; Hilton, Regner, Cabantous, Charalambides and Vautier, 2011). This study addresses this deficit by selecting individual differences relating to two promising theories of overprecision, the Naïve Sampling Model (NSM) and the Informativeness-Accuracy Tradeoff (IAT; Juslin, Winman & Hansson, 2007; Yaniv and Foster, 1995; 1997).

Naïve Sampling Model (NSM)

The NSM explains overprecision cognitively - as a consequence of short term memory (STM) capacity. The underlying concept being that, to create an interval, a person calls relevant examples from long term memory (LTM). For instance, if asked to set an interval around the population of Nigeria, one might call to mind populations of other African countries (or, failing that, non-African countries).

This sample, held in STM, is then used as the basis for creating the interval estimate – for example, by taking the 10^{th} and 90^{th} percentiles of that sample and using these as the low and high ends of an 80% confidence interval. The sample drawn from LTM, however, is limited by a person's STM to a small number of instances and, as sampling dispersion is a biased estimator of population dispersion, 80% coverage of the sample does not correspond to 80% coverage of the population – leading to too narrow ranges.

Individual Differences in NSM. NSM holds that better STM decreases overprecision due to increased interval width - as larger samples are less likely to underestimate population dispersion. Support for a reduction of overprecision due to STM was found in a study involving a learning and then a testing phase - but interval width was not explicitly measured (Hansson et al., 2008).

Informativeness-Accuracy Tradeoff (IAT)

The IAT is a motivational explanation of people's preferences when receiving and generating interval estimates (Yaniv & Foster, 1995; 1997). In interval estimation terms, narrow intervals are more informative but, holding all else equal, less likely to be accurate, while wide intervals are less informative and more accurate. Thus, these objectives need to be traded off against one another.

Yaniv and Foster's (1995; 1997) participants showed a general preference for receiving informative (narrow) interval estimates that missed the true value rather than uninformative (broad) estimates containing it, stating that these were more 'useful' than the wider estimates.

Participants also produced intervals containing the true value far less often than the stated 95% confidence level, and when informed of the degree to which their intervals would need expanding to contain the true values often enough, opined that this would render the judgements useless to the receiver (Yaniv & Foster, 1995; 1997).

The regulation of informativeness-accuracy thus follows conversational norms (Yaniv & Foster, 1995; 1997) - in providing a judgement, estimators are attempting to help the receiver. The receiver's purpose in soliciting a judgement should guide the relative informativeness or accuracy of the estimator (Yaniv & Foster, 1995; 1997).

A key contribution of this theory is the decomposition of intervals into width (upper minus lower bound) and absolute error (distance from the interval's midpoint to the true value), with width thought to reflect strategy (how informative or precise an estimator is claiming to be) while absolute error reflects knowledge (Yaniv & Foster, 1997).

Individual Differences in IAT. The conversational norms explanation for regulation of informativeness-accuracy in interval estimation has received limited support, with attempts to manipulate receiver purpose not affecting overprecision. Nor is it clear why communications between experimenter and participant should induce informative rather than accurate responses (Haran, Radzevick & Moore, 2010, cited in Moore, Tenney and Haran, 2015).

A novel explanation is that people may be predisposed to informative or accurate judgments by innate thinking dispositions such as Need for Cognitive Closure (NFCC) - defined as the desire to answer a question rather than sustaining further uncertainty (Webster & Kruglanski, 1994). Wide intervals are necessarily ambiguous, and it may be that high-NFCC participants would produce narrow intervals to avoid ambiguity and attain a feeling of closure.

Additionally, low NFCC is qualitatively similar to high Actively Open-minded Thinking (AOT), which predicted reduced overprecision on a single interval estimate (Haran et al., 2013). NFCC is preferable, however, as it includes discomfort with ambiguity, which AOT does not, and the NFCC scales are better validated than those for the AOT.

Online Confidence

Given uncertainty regarding the causes of overprecision, online confidence (OC) – the average post-item confidence ratings from a 12-item form of the Raven's APM test (Arthur & Day, 1994) - was also used as a predictor in this study. OC is thought to reflect a stable confidence trait across domains, shown to be predictive of performance within a domain - such as, online confidence from an earlier English test predicting end-of-year English grades (Stankov, Lee, Luo & Hogan, 2012) - but not yet across domains.

Hypotheses

1. Better STM will result in wider intervals and reduce overprecision. (NSM hypothesis)
2. Higher NFCC will result in narrowed intervals and increase overprecision (IAT hypothesis).
3. Higher online confidence will predict reduced overprecision (confidence-overprecision hypothesis).

Method

Participants

Participants (n = 49, 29 females, mean age = 31.0 years, SD = 15.5) were drawn from University of Adelaide students and the general population. They were highly educated, with 87% of participants having attempted a Bachelor's degree. Psychology 1B students participated for course credit and others entered in a draw to win one of two $50 gift cards.

Materials

Digit Span (STM capacity) Forward digit span, adapted from the Wechsler Adult Intelligence Scales (Wechlser, 2008), measured STM capacity. To ease data collection, the task was administered in groups. Participants were read lists of numbers as per the standard digit span procedure but provided written rather than verbal responses. The two response methods have been shown to produce similar scores within subjects (Ryan, Townsend & Kreiner, 2014).

There were two trials at each span level, 16 trials in all, ranging from two digits to nine digits in length. Each trial number was read out, e.g. "Trial 1", followed by a four second break, after which the digits were read out at a rate of one per second. Once all participants had finished writing their response, the next trial was announced and the process repeated until all 16 trials had been administered. A participant's score was the last span level at which they were correct on at least one of the two trials.

Need for Cognitive Closure Scale 15-item. A fifteen-item scale derived from the revised 41-item NFCC scale (Roets & Van Hiel, 2007) was used to generate the general NFCC factor (Roets & Van Hiel, 2011). Participants rated statements such as "I don't like situations that are uncertain" on a six-point Likert scale from 1, strongly disagree, to 6, strongly agree. Scores summed from all items give scores from 15 to 90 with higher scores indicating greater NFCC.

Raven's Advanced Progressive Matrices 12-item. Raven's is a 36-item measure of fluid intelligence for use with highly-educated samples but was used here to derive OC. Arthur and Day's (1994) 12-item version was used, untimed, and a participant's score was simply their number of correct answers. After each item, participants provided confidence ratings (0-100%) for the accuracy of their answer, which were averaged as the measure of OC.

General Knowledge Interval Estimates (Overprecision). Twenty general knowledge questions were used to assess overprecision in interval estimation. Participants made 80% confidence estimates in accordance with the oil and gas industry standard (Welsh, Bratvold & Begg 2005), and saw an example question to learn the required response format.

Four topic areas were used; geography, sport, big business and historical events, with five questions from each topic. In theory, question content is unimportant as optimum performance involves scaling intervals to accord with the precision of knowledge, not knowing exact answers to the questions. However, participants often react with frustration when they find questions too difficult or topics unfamiliar (Welsh et al., 2005) and, thus, a range of topics was used.

Bias Score. Bias score was used as the measure of overall overprecision, calculated as the given confidence level (i.e., 0.8) minus the proportion correct. Perfect calibration would result in a bias score of 0, overprecision a positive bias score and underprecision a negative bias score.

Interval Width. Interval width was measured by subtracting an estimate's upper from its lower bound. As width is relative to the scale of the correct answer, participant's ranges were ranked from widest to narrowest (i.e. widest = 1) for each item, using average ranks for ties. Participant's mean rank across the 20 questions was used in analyses. Scores closer to 1 thus indicate participant's intervals that were, typically, wider than other participants.

Absolute Error. This was calculated as the absolute difference between the true value and the interval midpoint. Absolute error was calculated as described for interval width but values were ranked smallest to largest (i.e. least absolute error = 1), ensuring higher interval width and absolute errors both correspond to greater overprecision.

Procedure

Participants first completed the digit span test and gave demographic information before undertaking the interval estimation task. They were told to remind themselves when reading each estimation question that they should be 80% sure that the true value was between the low and high bounds they provided. After the interval estimates, participants completed the NFCC scale and then the Raven's APM. All tests were conducted using pencil and paper and completion time was 30-45 minutes.

Results

Given the small sample, Efron's (1987) BCa bootstrap procedure with 2000 resamples was used to calculate statistics in addition to traditional methods.

Descriptive Statistics

Table 1 shows descriptive statistics for all continuous variables, revealing that participants displayed the expected pattern of overprecision in their interval estimates, with a mean bias score of .51. The degree of miscalibration is on the high end of that observed in other studies, with hit rates around 30% at a pre-stated confidence level of 80%.

Table 1: Descriptive statistics for all continuous variables.

	M	*SD*	*Min.*	*Max.*
Age	31.04	15.45	18	80
Digit Span	6.92	1.35	4	9
NFCC	57.06	10.80	38	82
Online Confidence	67.68	18.34	17.5	100
Raven's APM	7.88	2.76	2	12
Bias Score	.51	.17	.00	.80
Interval Width	25.00	6.64	11.33	39.55
Absolute Error	25.00	5.23	14.23	42.70

Table 2 present a correlation matrix for these variables - excepting between bias score, interval width and absolute error, due to their statistical dependency.

Table 2: Correlation matrix for all continuous variables

		1	2	3	4	5
1	Age	-				
2	DS	-.10	-			
3	NFCC	.04	-.20	-		
4	OC	**-.28**	.13	-.24	-	
5	RAPM	**-.36**	.20	**-.34**	**.67**	-
6	Bias Score	.09	-.06	**.32**	-.10	-.27
7	Int. Width	**.30**	-.27	**.40**	-.24	**-.38**
8	Abs. Err.	-.25	.13	.06	.12	-.01

Bold = $p \leq .05$. DS = digit span; NFCC = need for cognitive closure; OC = online confidence; RAPM = Ravens APM.

Looking at Table 2, one can see three major points relating to our hypotheses. The first is that online confidence is closely related to the people's scores on the Ravens (as would be expected) but not significantly related to any of the three overprecision measures (bias, width or error). The second is that digit span seems similarly unrelated – although the relationship between digit span and interval width, at -0.27, approaches significance. Finally, NFCC is significantly correlated with the bias score and more strongly with interval width but not absolute error.

Beyond our hypotheses, it is worth noting the significant negative relationship between participants Ravens scores and interval widths – due, perhaps, to the fact that age negatively predicts people's online confidence and Raven's scores and is positively related to interval width.

The above examination of correlations suggests that NFCC predicts overprecision bias and interval width whereas digit span and online confidence do not. This piecemeal approach, however, ignores potential relationships between the predictors. Therefore, multiple regressions were conducted as a means of examining all three potential predictors of overprecision simultaneously.

Interval Width

Table 3 displays the results of a multiple linear regression conducted to test the hypotheses that: STM capacity, as measured by digit span, would be related to the production of wider intervals; that NFCC level would be related to the production of narrower intervals; and that online confidence would be related to the production of wider intervals.

The model reached significance, $F(3,45) = 4.17$, $p < .05$, $R^2 = .22$, with NFCC as the only significant predictor. The hypothesis that higher NFCC score would be related to the production of narrower intervals was supported, however the wide bootstrap CI^{95} for the regression coefficient makes the true strength of the relationship difficult to ascertain.

Table 3: Multiple linear regression analysis predicting interval width from digit span, NFCC and online confidence

	B	BCa CI^{95}	*SE*	β	*t*	*p*
DS	-.95	[-2.30, .30]	.72	-.19	-1.43	ns
NFCC	.20	[.07, .32]	.07	.33*	2.39	<.05
OC	-.05	[-.16, .09]	.06	-.13	-.99	ns

In contrast, the hypothesis that high STM would result in wider intervals was not supported. The (n.s.) relationship observed was, in fact, opposite to that predicted. Likewise, online confidence failed to predict interval width.

Overprecision

The above analyses support the supposition that NFCC predicts interval width but the degree to which this results in overprecision bias is, perhaps, more important to know.

Table 4 displays results of a multiple linear regression conducted to test the hypotheses that: STM capacity (i.e. digit span) would be related to reduced overprecision; that NFCC would be related to increased overprecision; and that online confidence would predict reduced overprecision.

The model did not reach significance, $F(3,45) = 1.75$, $p = .17$, $R^2 = .10$. Despite the significant correlation between NFCC and bias score in Table 2, NFCC was not a significant predictor of overprecision in this model - and none of the three hypotheses found support.

Table 4: Multiple linear regression analysis predicting bias score from digit span, NFCC, online confidence and Ravens

	B	BCa CI^{95}	*SE*	β	*t*	*p*
DS	.001	[-.04 .04]	.021	.01	.04	ns
NFCC	.005	[.001, .009]	.002	.32	2.16	ns
OC	.000	[-.003, .002]	.001	-.02	-.14	ns

Discussion

These results go some way in helping to decide between two promising theories of overprecision: the Naïve Sampling Model (NSM, Juslin et al., 2007); and the Informativeness-Accuracy Trade-off (IAT, Yaniv & Foster, 1995; 1997) with some support offered for the latter - at least, to the extent that one accepts that NFCC reflects an intrinsic tendency to prefer informativeness over accuracy.

A third possible predictor of overprecision – online confidence (Stankov et al., 2012) – showed no clear relationship with overprecision, suggesting that these forms of confidence and overconfidence are not closely related.

STM Capacity

The idea that superior STM causes the production of wider intervals and thus reduces overprecision is central to the Naïve Sampling Model (NSM: Juslin et al., 2007). Thus, this study showing no effect of STM on interval width and failing to replicate Hansson et al.'s (2008) finding that STM predicted reduced overprecision presents a challenge to the theory as described (although it does not rule out the possibility of other sampling processes providing a sound explanation of the interval estimation process).

These results appear not to be STM measurement issues, as the digit span test used here produced the expected 'seven plus-or-minus two' pattern of results (Miller, 1956). There may, though, be other explanations for why this study failed to support Hansson et al.'s (2008) assumptions and findings.

Firstly, forward digit span, as used in this study, measures only STM. The digit span task used in Hansson et al. (2008), however, is described as a composite (via an unstated formula) of a passive repeat-back task (presumably forward span) and a sequencing task, which are thought to reflect working memory in addition to capacity (Engle, Tuholski, Laughlin & Conway; 1999). Thus, Hansson et al.'s (2008) scale may have measured working memory and STM, affecting the relationship they observed. The positive correlation Hansson et al. (2008) observed between digit span and Raven's APM was stronger than seen herein, consistent with the notion that the more complex span task may have led to a stronger relationship with working memory, which is more closely related to Gf – as measured by Raven's APM (Wiley, Jarosz, Cushen & Colflesh, 2011).

Secondly, the general-knowledge interval estimation used herein differed significantly from the laboratory learning task used by Hansson et al. (2008) in that it did not control for prior knowledge. Hansson et al.'s (2008) task involved a learning phase of fictitious company data, followed by an estimates phase wherein participants made a point estimate, thought to reflect information successfully stored in LTM if correct, and an interval estimate at the 80% probability level, thought to reflect inference from STM if correct.

Their results suggested that STM capacity as assessed by digit span was negatively related to overprecision but proportion of correctly recalled values from LTM was not. This was interpreted as evidence for the importance of STM capacity as compared to information stored in LTM in generating correct interval estimates (Hansson et al., 2008). However, the strongest predictor of overprecision was the variance of the values correctly recalled from LTM, which could be assessed by comparing the values recalled to the distribution of values from the learning phase. Those in the low-variance group after a median split were almost twice as overprecise as those in the high-variance group, pointing to the possibility of a role for information stored in LTM.

Controlling for prior knowledge is necessary to demonstrate the theorised dissociation between STM capacity and LTM information storage but such a task is not representative of real world estimation tasks.

Of interest - in light of this study's results - is the finding that low-variance LTM representations were associated with increased overprecision (Hansson et al., 2008). It is possible that when prior knowledge is not controlled, as in general knowledge questions as used in this study, between-subjects differences in LTM representations may obscure the relationship between STM capacity and overprecision. However, were STM's predictive power to be realized only in laboratory settings where prior knowledge is controlled, the utility of NSM for applied settings must be questioned.

Need for Cognitive Closure

The hypothesis that higher NFCC would correlate with production of narrower intervals was supported but the effect of this on a person's level of overprecision was mixed – supported in the correlation table but not in the multiple regression. This is, perhaps, unsurprising as overall bias is also affected by differences in individual knowledge and any effect on interval width is diluted by the inclusion of errors in estimation, which NFCC does not predict.

The results offer support for interpreting the NFCC trait as per Webster and Kruglanski (1994), which led to the presumption that it might be linked to a preference for informativeness over accuracy. They also support a qualitative similarity between low NFCC and high AOT, as low-NFCC participants in this study behaved like high-AOT participants in Haran et al. (2013). NFCC could be considered a complimentary construct to AOT given its significant relationship with interval width, which was not predicted by AOT in Haran et al. (2013).

Overall, the fact that a dispositional variable (NFCC) was the best predictor of both interval width and, despite the model not reaching statistical significant, overprecision, argues for an account of overprecision including motivational and strategic aspects rather than a purely cognitive one. It is, thus, broadly supportive of the principles of the Informativeness-Accuracy Tradeoff's overprecision explanation (Yaniv & Foster, 1995; 1997). The theoretical link drawn here between NFCC and IAT thus seems to have been sensible; although the mechanism underpinning this requires elucidation, these findings should stimulate further research into NFCC and, more generally, the role of intrinsic motivators as drivers of the IAT.

Future Directions

Few studies have examined overprecision on interval estimates from an individual differences perspective and, so, these results suggest multiple directions for future research. A first, drawing on results in Table 2, will be to nail down the relationships between age and the various confidence, intelligence and bias measures examined herein.

Naïve Sampling Model This study did not support NSM's ability to explain overprecision – at least in cases where estimates are made in contexts without controlled prior knowledge. Follow up research into the effects of task features on NSM is therefore necessary to clarify its utility. A future study could examine predictions of NSM in a within-subjects design, with conditions having different levels of control over prior knowledge, shedding light on the utility of NSM outside the laboratory. Additional work utilizing separate measures of working memory and STM could also help disentangle possible confounds, while investigating the ability of other sampling processes to explain overprecision would help clarify the worth of further developing sampling-based models of overprecision.

Need for Cognitive Closure As a new construct to the interval estimation field, many aspects of NFCC bear further investigation. Understanding the mechanisms underlying the association between high NFCC, the production of narrower intervals, and a potential relationship with increased overprecision is important. Thus, use of the 41-item revised scale - which includes subscales - (Roets & Van Hiel, 2007) in future research would be a sensible starting point.

Also relevant are moderators that might affect how those high or low in NFCC behave. The link between NFCC and interval estimates posited here was that wide intervals are ambiguous and high-NFCC participants would avoid ambiguity - providing narrower intervals. However, an alternative explanation could be drawn from the literature on NFCC and information search. Specifically, rather than trying to reduce affective discomfort caused by ambiguity, participants may simply not search for enough information to make credible estimates, as the perceived cost of information search is too high. Research on NFCC and information search suggests those high in NFCC search for less information and make faster decisions but the opposite pattern of behaviour can manifest under certain conditions (Choi, Koo, Choi & Auh, 2008; Van Hiel & Mervielde, 2002). Variables moderating the association between NFCC and information search should, therefore, be investigated.

For example, those low in ability to achieve closure, perceived ability to enact strategies to fulfil epistemic needs, and those with a low working memory capacity (WMC) may behave conversely to that suggested by their NFCC level, with high NFCC people prolonging information search and suspending closure and low NFCC people shortening it to achieve rapid closure (Czernatowicz-Kukuczka, Jasko & Kossowksa, 2014; Kossowska & Bar-Tal, 2013). The apparent relevance of WMC to the NSM and to the relationship between NFCC and information search may, thus, provide a path to integrating motivational and cognitive theories of overprecision.

The effect of moderators on the relationship between NFCC and information search paints a complex picture. A more nuanced view of the relationship between NFCC, interval width and overprecision may appear when moderators of NFCC behaviour are also measured.

Conclusion

The key finding of this paper is that NFCC, introduced as a possible measure of a person's preference for informativeness over accuracy in line with Yaniv and Foster's (1995;1997) IAT theory does predict overprecision in interval estimation – or, at least, the interval width aspect of this bias. Equally interesting is its failure to support the central assumption of the NSM account of overprecision. The results, therefore, point to the need to consider motivation and strategy in addition to the potential impact of cognitive processes when examining overprecision.

Acknowledgements

MBW acknowledges the support of the CIBP member companies: BG Group, ExxonMobil, Santos and Woodside.

References

Alpert, M. & Raiffa, H. (1982). A progress report on the training of probability assessors. In Daniel Kahneman, Paul Slovic & Amos Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press 294--305.

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psych. measurement*, (54), 394-403.

Choi, J. A., Koo, M., Choi, I., & Auh, S. (2008). Need for cognitive closure and information search strategy. *Psychology & Marketing, 25*(11), 1027-1042.

Czernatowicz-Kukuczka, A., Jasko, K., & Kossowska, M. (2014). Need for closure and dealing with uncertainty in decision making context: The role of the behavioral inhibition system and working memory capacity. *Personality and Individual Differences, 70*, 126-130.

Efron, B. (1987). Better bootstrap confidence intervals. *J. of the American statistical Association*, *82*(397), 171-185.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J. of experimental psychology: General*, *128*(3), 309.

Hansson, P., Juslin, P., & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *J. of Experimental Psychology: L.M.C., 34*(5), 1027-1042.

Haran, U., Radzevick, J. R., & Moore, D. A. (2010). Audience effects on overprecision in judgment. *Unpublished Manuscript*.

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*(3), 188-201.

Hilton, D., Regner, I., Cabantous's, L., Charalambides, L., & Vautier, S. (2011). Do positive illusions predict overconfidence in judgment? A test using interval production and probability evaluation measures of miscalibration. *J. of Behavioral Decision Making, 24*(2), 117-139.

Juslin, P., Winman, A., & Hansson, P. (2007). The naive intuitive statistician: A naive sampling model of intuitive confidence intervals. *Psych. Review, 114*(3), 678-703.

Kossowska, M., & Bar-Tal, Y. (2013). Need for closure and heuristic information processing: The moderating role of the ability to achieve the need for closure. *British J. of Psychology, 104*(4), 457-480.

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.), *Judgement under uncertainty: Heuristics and biases*, (pp. 306-334). Cambridge, UK: Cambridge University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psych. Review*, *63*(2), 81.

Moore, D. A., Tenney, E. R., & Haran, U. (2015). Overprecision in judgment. In G. Keren & G. Wu (Eds.), *Handbook of Judgment and Decision Making*. New York: Wiley-Blackwell.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psych. Review, 115*(2), 502-517

Roets, A., & Van Hiel, A. (2007). Separating Ability From Need: Clarifying the Dimensional Structure of the Need for Closure Scale. *Personality and Social Psychology Bulletin, 33*(2), 266-280.

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Ind. Diff., 50*(1), 90-94.

Ryan, J., Townsend, J., & Kreiner, D. (2014). Comparison of oral, written, and pointing responses to the WAIS-IV Digit Span. *Applied Neuropsychology: Adult*, *21*, 94-97.

Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences, 22*(6), 747-758.

Van Hiel, A., & Mervielde, I. (2002). Effects of ambiguity and need for closure on the acquisition of information. *Social Cognition, 20*(5), 380-408.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *J. of Personality and Social Psychology, 67*(6), 1049-1062.

Wechsler, David. "Wechsler adult intelligence scale–Fourth Edition." *San Antonio, TX: NCS Pearson* (2008).

Welsh, M.B., Begg, S., & Bratvold, R. (2007). Modelling the Economic Impact of Common Biases on Oil and Gas Decisions. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Welsh, M.B., Bratvold, R., & Begg, S. (2005). Cognitive biases in the petroleum industry: Impact and remediation. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

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

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *J. of Behavioral Decision Making, 10*(1), 21-32.