Overconfident in Hindsight: Memory, Hindsight Bias and Overconfidence

Matthew B. Welsh (matthew.welsh@adelaide.edu.au)
Australian School of Petroleum, University of Adelaide
North Terrace, Adelaide, SA 5005, Australia

Abstract

Overconfidence and Hindsight Bias are two well-known cognitive biases. Herein, it is argued these biases may be related to one another and human memory limitations; specifically, that memory limitations result in hindsight bias, causing people to recall being right more often than they actually were, which leads to overconfidence as people apply this misremembered confidence to future events. Analyses comparing three types of overconfidence (overestimation, overplacement and overprecision) and hindsight bias confirm strong, positive correlations between the different types of overconfidence – from 0.488 up to .807 and moderate correlations (.331 to .398) between all of these and hindsight bias. Comparisons between bias scores and five broad cognitive abilities (from the CHC model) suggests hindsight bias is more pronounced in people with worse memories and generally, lower cognitive ability. Overall, results are argued to support the proposed links between memory, hindsight bias and overconfidence and future directions are suggested.

Keywords: overconfidence; hindsight bias; memory; cognitive bias.

Introduction

Hindsight bias (Fischhoff & Beyth, 1975) and overconfidence (Lichtenstein, Fischhoff, & Phillips, 1982) are two well-known effects within the heuristics and biases literature describing, in the simplest of terms, peoples tendencies to misremember how often their predictions were correct in the past and overestimate how often they will be right in the future, respectively.

While typically regarded as separate effects, the statement of their basic natures above gives reasons to regard these concepts as fundamentally similar – as both revolve around the misestimation of likelihood of success, differing only in terms of whether they are focused on past or future events. The biases and the proposed links between them are expanded upon below.

Overconfidence

Overconfidence, when considered as a cognitive bias specifically, refers to the tendency for people to overstate or overestimate their degree of knowledge about some aspect of the world. This is confused somewhat by the use of the same terminology for a set of related measures. Moore and Healy (2008), for example, distinguish three types of overconfidence measures commonly used in the bias literature: overestimation, overplacement and overprecision – and this distinction will be used herein for clarity.

While all three types of overconfidence reflect some tendency of people to judge themselves and their own knowledge more favourably than the evidence warrants, there are differences that should be noted. Overestimation tasks measure people's tendency to overestimate how many items in a test they will or have gotten correct whereas overplacement refers to their tendency to rate themselves more highly against other people. Finally, overprecision is the observation that, when giving ranges that they are confident (to some stated level) that future/unknown values will fall within, people tend to give ranges that are too narrow such that observed values fall outside the predicted ranges more often than they expect (Moore & Healy, 2008).

Given these different methods for assessing overconfidence, it is possible to generate multiple measures of an individual's overconfidence that may differ significantly. For example, a person's overprecision can be measured by giving them a set of forecasts to make and comparing their stated confidence level with the observed rate at which their ranges capture the true values. One can also, however, ask people directly to indicate how many of their ranges (as a set) they expect to capture the true value and, despite these being theoretically equivalent (if people are accurately stating their confidence in their ranges), the standard observation is that this overestimation approach produces a differing degree of overconfidence (Juslin, Wennerholm, & Olsson, 1999) - resulting in discussions about the extent to which these effects are caused by the same underlying mechanisms. For example, some argue that overestimation and overplacement are, largely, statistical artefacts (see, e.g., Soll & Klayman, 2004; Winman, Hansson, & Juslin, 2004) whereas long-standing theory suggests overprecision is caused by an anchoring effect (Tversky & Kahneman, 1974) – despite a paucity of supporting evidence (see, e.g., Block & Harper, 1991; Bruza, Welsh, Navarro, & Begg, 2011).

Despite the statistical arguments above, though, the fact that all three seem to reflect a general tendency to overestimate one's own knowledge makes it seem most likely that there is some shared, underlying cause and, thus, that people's performance on one task should align with their performance on the others.

Hindsight Bias

Hindsight bias, by contrast, refers to the tendency for people to 'update' their remembered predictions when given information about how the predicted events actually turned out. Specifically, it seems that, once an outcome has become known, people are more likely to believe that they predicted it all along and, conversely, believe they rated events that did not eventuate as less likely. This was initially demonstrated with people's predictions regarding possible outcomes of Richard Nixon's trip to China and the Soviet

Union - tested prior to and several months after the trip once the outcomes were known (Fischhoff & Beyth, 1975).

This makes sense as an adaptive strategy for operating in the real world – once you know the outcome of an event, it is much easier to understand the causal relationships leading to that event and see *why* it occurred. From the point of view of learning in the real world with limited cognitive resources, it then makes little sense to keep the previous causal story that you now know to be incorrect, and more sense to ‘update’ memories of such predictions to match current knowledge.

More recent work, however, has suggested that there may be more to hindsight bias than this – suggesting that it, too, has three components and that these may be distinct enough that differences observed between different hindsight paradigms might be explained by these differing effects (Nestler, Blank, & Egloff, 2010). These effects are described by Nestler and colleagues as: inevitability impressions, which matches closely the causal reasoning approach above; foreseeability impressions, which are driven by metacognition regarding what you *might* have thought at some other time; and memory distortions, describing errors in encoding, retrieving and reconstructing past estimates. Which of these three types of hindsight bias are seen in an experiment will depend on the particular experimental design but, significantly, this interpretation of the concept means that even designs which avoid manipulations of inevitability or foreseeability can still show a memory-based hindsight bias effect.

Impact of Hindsight Bias and Overconfidence

Both (all) of these effects are important in applied decision making contexts. For example, industry personnel often forecast or estimate unknown quantities to produce data or models on which large investment decisions are made. Overconfidence in these estimates can have significant economic implications – where decision makers put plans in place without contingencies for values much higher or lower than those predicted. For example, overconfidence at the levels commonly seen in oil and gas personnel (30-40%; see, e.g., Welsh & Begg, 2016; Welsh, Bratvold, & Begg, 2005) has the potential to change the valuation of a large development project by hundreds of millions of dollars (Welsh, Begg, & Bratvold, 2007).

Similarly, hindsight bias has the potential to contribute to failures to learn from experience - where decision makers and technical personnel come to believe that they predicted events more accurately than they actually did – with the result that they may not feel a need to change their estimation strategies.

Given this, understanding how and why the effects occur has the potential to improve decision making and suggest ways to limit or avoid the biases.

Linking Hindsight Bias and Overconfidence?

At first glance, these two effects seem distinct and a search of the literature indicates they are not commonly discussed together. In fact, examination of connections between different types of bias in the heuristics and biases field has, in general, been somewhat ad hoc (Ceschi, Costantini, Sartori, Weller, & Di Fabio, 2019).

Consideration of their nature and origins, however, suggests possible connections. For example, hindsight bias could almost be described as overconfidence *despite* evidence. That is, being faced with evidence that should confirm or deny the accuracy of their predictions, people ‘adjust’ their memories in a manner that mimics overestimation – recalling more correct predictions than they actually made.

This suggests the possibility of a causal relationship – where the limitations of human memory (and cognition more generally) cause hindsight bias and this tendency to recall being right more often than they actually were acts to increase people’s confidence in their future estimates – causing overconfidence.

Aims and Objective

This paper is an initial look at this proposed relationship, seeking to establish whether hindsight bias and measures of overconfidence are linked in a meaningful way and how this relates to differences in human cognitive abilities – particularly memory.

Hypothesis 1. The three types of overconfidence will correlate positively with one another.

Hypothesis 2. Hindsight bias will correlate positively with the measures of overconfidence.

Hypothesis 3. Hindsight bias will be more pronounced in people with worse memories (long term memory, retrieval ad working memory).

While these hypotheses are correlational rather than examining causal relationships between the measures, this is regarded as a necessary first step. As noted above, discussions about whether the three forms of overconfidence share an underlying cause have not been resolved but positive correlations between them would make it more difficult to argue that some forms are simple, statistical artefacts.

Likewise, while a potential, causal relationship between hindsight bias and overconfidence is posited above, demonstrating a correlation between them is a necessary first step.

Finally, the relationship of memory and cognitive ability, more generally, to both overconfidence and hindsight bias needs to be understood as this provides a likely mediator linking the two effects.

Method

Participants

The hindsight and overconfidence measures were included in a large study examining relationships between susceptibility to decision making biases and a wide variety of individual traits (including attention, cognitive, personality and decision styles) being conducted under ARC Grant LP160101460. In total 300 participants completed the study (119M, 172F and 9 non-binary or did not say), ranging in age from 18 to 79 (mean = 28.7, SD =12.8). Most participants were native English speakers (n=207) and students/graduates (undergrads=107; bachelor-level graduates=84; post-grads=38; higher degree graduates=26; and vocational qualifications=20) with only 26 participants without any post-secondary study. Participants received a $100 gift card for participating in the study as a whole.

Materials

Overconfidence. Questions were written to test the three types of overconfidence distinguished by Moore and Healy (2008): overestimation, overplacement and overprecision, as described below.

Overplacement. Pre and post overplacement (OC_Pl) questions asked participants to indicate what percentage of participants they expected to do, or thought they had done, better than on the overprecision task described below. Their actual percentile performance on these questions was subtracted from their estimated percentages to determine their level of overplacement at each point, such that higher values indicate greater levels of bias.

Overprecision. Twenty range estimation question asked participants for their best estimate and high and low values that they were 80% confident would contain the true answer to that question. Ten questions were almanac-style questions of fact (e.g., how many tonnes of iron are in the Eiffel Tower) while ten were forecast questions (e.g., what will the price of gold be one week from today) but these are treated together herein. A participant's overprecision (OC_Pr) was calculated as the difference between the percentage of questions expected to contain the true value (80%) and the percentage of their ranges that actually contained the true value, which yields higher values for higher level of the bias.

Overestimation. Immediately after completing the overprecision task, participants were asked to indicate how many of the ranges they had just generated they believed would contain the true value. The difference between this and their observed performance (both as percentages) was calculated as their overestimation (OC_E) score. Again, higher scores indicated greater bias.

Hindsight Bias. This task consisted of the 44 questions used in: the 12-item version of Ravens Advanced Progressive Matrices task (Arthur & Day, 1994); the CAB-I (Comprehensive Ability Battery Inductive Reasoning task; Hakstian & Bennet, 1977), and the Overprecision (OC_Pr) task described above. Specifically, participants were shown a question they had previously answered, along with the correct answer and then asked whether they believed they had answered the question correctly. Their overall score for this measure was difference between the percentage of questions they now recalled getting right and the percentage they actually did.

Cognitive Abilities. Following the Cattell-Horn-Carroll model (CHC; McGrew, 2009), a wide range of cognitive ability measures were included in the larger study, covering eight of the broad abilities: long term memory *Glr*, short term memory *Gwm*, fluid ability *Gf*, crystallized ability *Gc*, processing speed *Gs*, decision speed *Gt,* visual processing *Gv* and quantitative ability *Gq*. Of these, long term retrieval, working memory, crystallized ability, fluid ability and quantitative ability were considered potentially relevant to the hindsight bias and overconfidence tasks. Specifically, Gc, Glr and Gwm all seem to be measures of aspects of human memory – knowledge in long term memory, retrieval from long term memory and working memory, respectively. By comparison, Gf measures people's ability to solve novel problems and Gq their ability to work with numbers – both of which seem requisites for answering overconfidence-style questions.

Table 1 shows the specific tests used to measure each of these factors. Given three measures for each broad ability, an overall factor score was calculated from each set of measures using Principle Factors Analysis (PFA) with Oblimin rotation in SPSS. In each case, this returned a single factor with an eigenvalue above 1, which was used as the overall measure.

Free Recall. Participants were shown a list of thirty words and allowed 2 minutes to memorize it. After a simple algebra distractor task to disrupt rehearsal, they were then allowed up to five minutes to write down as many words as they could recall. Their score was the number of words recalled.

Memory Span Forward/Backward. These tasks were designed in Matlab and presented strings of digits – each digit displayed one at a time for one second. The length of the string increased from 1 to 10 and, at the end of each string, the participant was asked to enter the string in: order of presentation (Forward); or reversed from the order of presentation (Backwards). In both cases, the participant's score was the longest string of correctly recalled digits amongst their responses. (NB, this included the possibility of getting, e.g., the first 8 digits right in a 10-digit string or the last 6 in an 8-digit string.)

Table 1. Cognitive measures by CHC broad ability factor

CHC Ability	Specific Measures
Gc	Spot-the-Word (Baddeley, Emslie, & Nimmo-Smith, 1993), Mill Hill Vocab Test (Raven, 1958), WJ-IV Cog1 (Schrank et al., 2015)

Gf	Ravens APM 12 Item (Arthur Jr & Day, 1994), CAB-I (Hakstian & Bennet, 1977), WJ-IV Cog2
Glr	WJ-IV OL4, WJ-IV OL8, Free Recall
Gq	Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), 12 item Numeracy (adapted from DAT) (Bennett, Seashore, & Wesman, 1989), Subjective Numeracy Test (Fagerlin et al., 2007)
Gwm	WJ-IV Cog10, Memory Span Forward, Memory Span Backwards

Note: WJ-IV refers to the Woodcock-Johnson IV tests of cognitive ability and achievement. Tests without references were developed for this study and are described in text.

Procedure

As noted above, participants were engaged in a larger study – too large to describe in detail here. In brief, it included two online studies, the first including a full, 5-factor personality test, several decision style measures and a novel Subjective Attention Scale. A second, online study included some measures of intelligence, confidence and bias susceptibility. Following completion of the two online surveys, participants were invited to the laboratory for an additional 2.5 hours of intelligence testing, decision bias testing and a final survey. The overplacement, overprecision and overplacement tasks were included in the second survey while hindsight bias was measured during the third survey. The cognitive measures were split between the in-person testing (WJ tests, Free Recall and Memory Span tasks) and the second survey (all other measures).

Results

Outcome Measures

Table 2 shows the mean levels of overconfidence and hindsight bias. Looking at the table, one can see the large differences between the different types of overconfidence – with overprecision (OC_Pr) yielding a bias score of more than 50% while overestimation (OC_E) yields a much smaller 6.7%. The overplacement (OC_Pl pre) score from prior to participants actually starting the task is around 9% but, following the task, people actually underestimate their own performance – presumably because they realized how difficult the task was for them. A key observation, however, is the extreme variability in responses – reflected in the large SDs and wide ranges.

The hindsight bias results similarly show a small average tendency for people to recall being right more often than they actually were. A series of single sample t-tests compared the levels of bias observed with the mean scores of zero that would indicate unbiased responses and indicate that for all of the measures except OC_Pl (post), the sample – as a whole – were biased, $t(299) = 4.44, 4.49, 51.71$ and 3.98, $p<<.001$ (2-tailed) for OC_Pl (pre), OC_E, OC_Pr and Hindsight Bias, respectively. The t-test for OC_Pl (post) showed no significant bias, $t(299) = -1.48$, $p = .14$ (2-tailed).

Table 2. Descriptive statistics for outcome measures

Measure	Mean	SD	Min	Max
OC_Pl (pre)	9.2	36.0	-89.3	91.7
OC_Pl (post)	-3.2	37.0	-93.3	79.0
OC_E	6.7	25.9	-65.0	80.0
OC_Pr	50.6	16.9	-10.0	80.0
Hindsight Bias	3.0	13.0	-38.6	56.8

Note: N = 300.

Overconfidence and Hindsight Bias

Correlations calculated to examine the relationships between the four overconfidence measures and hindsight bias are shown in Table 3.

Table 3. Correlations between overconfidence and hindsight

	1	2	3	4	5
1. OC_Pl 1	-				
2. OC_Pl 2	.709	-			
3. OC_E	.545	.596	-		
4. OC_Pr	.807	.745	.488	-	
5. Hindsight	.375	.331	.398	.365	-

Note: N=300, all correlations significant at <.001 level.

Looking at Table 3, one sees strong evidence in support of Hypothesis 1; that being that the overconfidence measures tend to be strongly correlated with one another – the weakest being the .488 correlation between overprecision and overestimation. While overprecision and overestimation as both are calculated using the participants' actual performance on the 20 overprecision questions and this could explain some of their correlation, the same does not hold true for the overplacement measures, which are estimated distinct from each other and both other overconfidence measures. That is, despite the apparent differences in the strength of the four biases described above, there do seem to be stable tendencies for people who show more of one overconfidence bias to show more of the other forms of overconfidence as well.

Finally, hindsight bias is moderately strongly (and significantly) correlated with all four of the overconfidence measures – supporting Hypothesis 2.

Memory / Cognitive Ability

Correlation analyses compared the five bias measures with the five cognitive abilities described above and results are shown in Table 4.

Looking at Table 4, one sees that all but one of the correlations are negative and sixteen of the twenty-five are significant – suggesting an overall pattern of participants scoring higher on the cognitive measures also scoring lower on the bias measures.

Table 4. Correlations between biases and cognitive abilities

	G_f	G_c	G_{lr}	G_q	G_{sm}
OC_Pl 1	***-.231***	**-.160**	**-.165**	-0.092	-0.107
OC_Pl 2	***-.246***	**-.158**	-0.112	**-.157**	-0.105
OC_E	*-.129*	-0.107	-0.083	0.063	-0.016

OC_Pr	***-.344***	***-.200***	**-.160**	***-.236***	*-.147*
Hindsight	***-.504***	***-.299***	***-.245***	-0.105	*-.125*

Note: N=300, 299 (Glr) or 297 (Gsm) due to missing data. * *italic* = sig. at .05 level; bold = sig. at .01 level; ***bold italic*** = sig. at .001 level. The data above are presented without family-wise correction in order to highlight the overall pattern. Using a Bonferroni correction for 25 comparisons, the ***bold italic*** results remain significant at the .01 or .001 level with the **bold** results just missing the .05 criterion (p ≈ .053-.082, one-tailed).

Looking at a finer level, one can see that fluid ability (Gf) is the best predictor – correlating significantly with all five bias measures and at above 0.5 with hindsight bias. Crystallized ability (Gc) is the next best, correlating significantly with all biases measures except overestimation. Long term retrieval (Glr) correlates with three of the bias measures while quantitative ability (Gq) and short term memory (Gsm) correlate weakly with two each. Viewed along the other axis, overestimation (OC_E) seems to be the odd bias out – with only one, very weak significant relationship with cognitive ability (Gf).

Given that Gc, Glr and Gsm can all be regarded as reflecting aspects of human memory (knowledge in long term memory, retrieval from long term memory and working memory, respectively), their correlations with hindsight bias provide support for Hypothesis 3 – although it is worth noting that fluid ability is the best predictor for all of the biases.

Discussion

The results provide support for all three of the hypotheses considered herein, which offers some support to the proposed relationships between memory, hindsight bias and overconfidence.

Firstly, the correlations between the different overconfidence measures suggest that there is some shared, underlying tendency across the different types of overconfidence task. This is important as it holds true even for the overplacement measure (OC_Pl post) that showed no overall bias. That is, even when group performance suggests that the bias does not exist, differences in the degree to which individuals show the bias are stable and predictable – allowing prediction of their performance in related tasks.

The hypothesized relationship between hindsight bias and overconfidence was also observed, with moderate correlations observed between all of the overconfidence measures and the degree of Hindsight Bias shown by participants. While correlation does not prove causation, the idea that hindsight bias should cause overconfidence certainly has face validity and seems more likely than the converse – overconfidence causing hindsight bias. Of course, this does not rule out both effects being caused by the same underlying tendencies – whether memory limitations or other effects.

This brings us to the observed relationships between cognitive ability and hindsight bias. As would be expected given the proposed relationship between memory, hindsight bias and overconfidence, the correlations observed between the cognitive measures and hindsight bias are stronger than those between the cognitive abilities and the overconfidence measures, suggesting that these relationships could be mediated through hindsight bias. One thing to note, though, is that, while hindsight bias correlated significantly with all three of the memory-related abilities, fluid ability was the best predictor of performance, suggesting that memory alone is insufficient to explain people's degree of hindsight bias – which could connect back to Nestler and colleague's tripartite division of hindsight bias (Nestler et al., 2010).

Caveats and Future Research

As noted earlier, the fact that OC_E and OC_Pr are derived from the same question set, while a feature in that it allows comparisons that avoid questions relating to the difficulty of items (the hard-easy effect), means that the relationship between these biases may be being inflated. Given this, future research could test these biases on different question sets that have been matched for difficulty to remove this confound.

A second concern relating to the overconfidence measures is the fact that simple interpretations of correlational analyses assume that lower bias scores are better. This is, of course, not strictly true. Bias scores of zero reflect the best performance while negative scores (which were observed for some participants in all of the bias measures) reflect *under*confidence. While this is often regarded as less concerning given people's tendencies towards overconfidence, it has the potential to be just as damaging for decision making. Given this, additional analyses need to be conducted in order to determine whether treating negative scores as good is obscuring or fallaciously amplifying relationships between the measures and whether the overall pattern of results resembles something like a Dunning-Kruger effect (Kruger & Dunning, 1999) with the best performing participants underselling their performance.

Regarding the cognitive measures, while all of the factor analyses used to combine individual cognitive ability measures into broad ability scores returned single factors, closer examination suggested that some of the individual measures were less than ideal. For example, the Free Recall task correlated only moderately with the other two Glr measures (but also with the Gsm measures), which suggests that the Glr factor used herein was not pure and may have had reduced predictive power. Similarly, the correlations between the three measures of quantitative ability were only moderate (and the use of the Subjective Numeracy Test as one is open to criticism). Overall, more in depth analyses are required to determine future research directions.

For example, structural equation modelling, used in place of the simpler exploratory factor analytic approach herein could tease apart the relationships between hindsight bias and the different types of intelligence by including all of the variables in a single model, which allows (amongst other things) intercorrelations between the intelligence measures to be controlled for.

A final caveat is that, given the rolling nature of the large study, with participants free to complete the two online studies in their own time before scheduling a time to complete the in-person testing, the delay between participants' completion of the overconfidence measures and the hindsight bias measure varied markedly between participants. This has not been accounted for in current analyses and future work will examine whether (or the degree to which) the delay between the testing sessions affected participant's susceptibility to hindsight bias and how this affects its relationship with the different measures of overconfidence.

Given the results, additional research should focus on testing the proposed relationships experimentally in order to shed more light on causality. Specifically, whether the relationship between hindsight bias and overconfidence is directly causal as proposed or mediated by shared underlying tendencies. For example, an experimental design wherein participants are given a learning task and then divided into two groups – one receiving feedback and one not in order to provoke hindsight bias in the former. The participant's degree of overconfidence on a follow-up task could then shed light on whether hindsight bias has an immediate effect on overconfidence. Given the distinction drawn between three types of hindsight bias (Nestler et al., 2010), there are also questions to be answered about whether these same effects would hold in a paradigm where inevitability and or foreseeability were manipulated in addition to the memory distortion effects that seem likely to underly the hindsight bias shown herein.

Conclusions

While correlational rather than demonstrating causation, the results presented herein are generally supportive of the hypothesized relationships between memory, hindsight bias and overconfidence; specifically, of the idea that memory limitations can increase the likelihood of a person showing hindsight bias and that their tendency to do so predicts their degree of overconfidence. Given this, future, experimental work should aim to more directly test the proposed causal relationship between hindsight bias and overconfidence and whether experiments using the alternative forms of hindsight bias show the same pattern of results.

Acknowledgements

MBW is supported by ARC LP160101460, with support from Santos and Woodside. Thanks to Kylie Davis for her assistance in data collection, Steve Begg for his work on the larger project from which this data has been sourced and three anonymous reviewers for their comments.

References

Arthur Jr, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological measurement, 54*(2), 394-403.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32*(1), 55-65.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1989). *Differential Aptitude Test*. Marrickville, NSW: Psychological Corporation.

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.

Bruza, B., Welsh, M. B., Navarro, D. J., & Begg, S. H. (2011). *Does anchoring cause overconfidence only in experts?* Paper presented at the Annual Meeting of the Cognitive Science Society (33rd: 2011: Boston, USA) CogSci 2011.

Ceschi, A., Costantini, A., Sartori, R., Weller, J., & Di Fabio, A. (2019). Dimensions of decision-making: An evidence-based classification of heuristics and biases. *Personality and Individual Differences, 146*, 188-200.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making, 27*(5), 672-680.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance, 13*(1), 1-16.

Hakstian, A. R., & Bennet, R. W. (1977). Validity Studies Using the Comprehensive Ability Battery (CAB): 1. Academic Achievement Criteria. *Educational and Psychological measurement, 37*(2), 425-437.

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 1038-1052.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121.

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.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. doi:https://doi.org/10.1016/j.intell.2008.08.004

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

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight≠ hindsight: Experimentally induced dissociations between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1399.

Raven, J. C. (1958). *Guide to using the Mill Hill Vocabulary Scale with the Progressive Matrices Scales*: H.K. Lewis & Co.

Schrank, F. A., McGrew, K., Mather, N., LaForte, E., Wendling, B., & Dailey, D. (2015). *Woodcock-Johnson® IV Tests of Early Cognitive and Academic Development*. Itaska: Riverside.

Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(2), 299-314.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Welsh, M. B., & Begg, S. H. (2016). What have we learnt? Insights from a decade of bias research. *APPEA Journal, 56*, 435-450.

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., 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*.

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.