Valentina Ferretti, Sule Guney, Gilberto Montibeller, and Detlof von Winterfeldt

Testing best practices to reduce the overconfidence bias in multi-criteria decision analysis

Article (Published version) (Refereed)

Original citation: Ferretti, Valentina, Guney, Sule, Montibeller, Gilberto and Winterfeldt, Detlof von (2016) *Testing best practices to reduce the overconfidence bias in multi-criteria decision analysis.* Proceedings of the 49th Annual Hawaii International Conference on System Sciences . pp. 1547-1555. ISSN 1530-1605
DOI: 10.1109/HICSS.2016.195

Testing Best Practices to Reduce the Overconfidence Bias in Multi-Criteria Decision Analysis

Valentina Ferretti
Department of Management
London School of Economics and Political Science
Houghton St., London WC2A 2AE
V.Ferretti@lse.ac.uk

Sule Guney
National Center for Risk and Economic Analysis of Terrorism Events
University of Southern California
University Park, Los Angeles, CA 90089
sguney@usc.edu

Gilberto Montibeller
Management Science and Operations Group
School of Business and Economics
Loughborough University, UK
g.montibeller@lboro.ac.ak

Detlof von Winterfeldt
National Center for Risk and Economic Analysis of Terrorism Events
University of Southern California
University Park, Los Angeles, CA 90089
winterfe@usc.edu

Abstract

This paper explores the effectiveness of several methods to reduce the overconfidence bias when eliciting continuous probability distributions in the context of multicriteria decision analysis. We examine the effectiveness of using a fixed value method (as opposed to the standard fixed probability method) and the use of counterfactuals and hypothetical bets to increase the range of the distributions and to correct possible median displacements. The results show that the betting procedure to correct the median is quite effective, but the methods to increase the range of estimates have only a have small, but positive effect.

1. Introduction

During the past 40 years behavioral research has demonstrated that laypeople and experts have many biases when making probability and utility judgments [1, 2, 3, 4]. During the same time period multi-criteria decision analysis has become a mature field with numerous applications to improve personal, government, and business decisions. Many decision analysts acknowledge the existence and pervasiveness of biases, but argue that analytical tools can reduce or eliminate them. However, some of these biases occur in the judgments required in decision analysis, which can substantially degrade the quality of an analysis.

Biases occur in multi-criteria decision analysis during the assessment of values and utilities and during the assessment of risks and uncertainties. For example, the equal weighting bias occurs when assigning weights to multiple criteria. The overconfidence bias occurs during the elicitation of probability distributions [5]. This paper focuses on biases during the assessment of risks and uncertainties. A multi-criterion example is the assessment of health risks and uncertain costs of health policies. Expert elicitation is often used to quantify these risks and uncertainties, because of a lack of conclusive data, few reliable models, and conflicting evidence. These expert judgments are known to be subject to biases.

In a recent article, Montibeller and von Winterfeldt [5] reviewed a large number of biases and classified them by whether they are easy to correct or hard to correct with debiasing tools. The authors also called for research that tests and evaluates best practices in reducing biases that are hard to correct. Presently, most of these debiasing tools are employed in an ad hoc way, without any assessment of their efficacy or comparison of their relative performance in reducing biases. This paper is the first in a series that examines well-established and extensively employed debiasing techniques.

Specifically, we examine the well-known overconfidence bias and explore the efficacy of several techniques to reduce it. The overconfidence bias has been studied as early as the 1960s [6]. Summaries of the substantial research on this bias up to 1980 can be found in Lichtenstein et al. [7] and a more recent update is provided in Moore and Healy [8]. In short, when asked for probability judgments, subjects generally provide probabilities that are too extreme (when judging the probabilities of binary events) or provide probability distributions that are too tight (when estimating distributions over continuous variables).

This bias is remarkably resistant to debiasing techniques [9, 10]. Improvements appear to occur, when subjects get frequent feedback and when the response mode is changed to asking for probabilities, given fixed values of the uncertain variable. For example, weather forecasters, who have much experience with probability judgments and receive feedback almost daily, are remarkably well calibrated (for a summary, see von Winterfeldt and Edwards [11]). Regarding the response mode, Abbas et al., [12] and Seaver et al. [13] compared the usual fixed probability method (providing values for fixed cumulative probabilities) with a method in which the participants provided cumulative probabilities for fixed values and found better calibration for the latter.

In this paper, we focus on overconfidence when eliciting probability distributions over continuous variables, as this type of judgment is frequently required in multi-criteria decision analysis. We evaluate three techniques to reduce the overconfidence bias, which are widely employed in practice:

1. Using a fixed value technique instead of a fixed probability technique.
2. Using counterfactuals to probe and expand the extremes of the distributions.
3. Using hypothetical bets to expand the extremes and to correct median estimates.

The remainder of this paper is structured as follows. First we provide a brief literature review focusing on overconfidence when eliciting continuous probability distributions. We also briefly discuss the most relevant techniques to reduce this bias. Next, we describe the design and methods of an experiment to test the most promising debiasing techniques, followed by some of the key results. The final section will provide conclusions and guidance for future research.

2. Literature Review

2.1. Overconfidence bias

Overconfidence, which is broadly defined as the excessive certainty that one knows the truth[1], has been considered as the most ubiquitous and potent form of cognitive biases that human judgment is vulnerable to [14, 15, 16, 17].

The existence of the overconfidence bias has been demonstrated by numerous laboratory experiments and field studies in the literature. Overconfidence is measured either in *binary choice tasks* or in *confidence interval tasks*. In the former task people are asked to choose one of the two options for a correct answer, and then estimate how confident they are with their choice. In the latter task, participants are given questions with numerical answers (e.g., "How tall is the Eiffel Tower?") and then asked to estimate a range of values that they are confident at a certain level (e.g., 90%) will include the correct answer [6]. We will be focusing this paper on the confidence interval task, as it is pervasive in decision analysis and more complex than the estimations required in binary choices.

General results in confidence interval tasks show that the percentage of confidence intervals which include the correct answers is lower than their assigned confidence level (i.e., 90% confidence intervals contain the correct answer less than 50% of the time). In other words, elicited confidence intervals are too narrow and exclude too many possibilities – indicating people are too sure that they know the true answer when in fact they do not know it [6, 18, 19]. This pattern has been replicated by numerous studies [14, 20] and is widely observed both in novice and expert judgments [21, 22, 23, 24, 25].

Overconfidence bias has profound consequences in practice, especially in professional and expert judgment. For instance, investors are excessively confident in their predictions on what an asset is worth, which lead them to engage in extraordinarily high trading activity [26, 27, 28]. Physicians and health professionals have been shown to reach their diagnosis too quickly and confidently by overlooking many other possibilities, which eventually results in mistreatment of patients [29 , 30 , 31 , 32]. Within organizations, individuals are overly confident in forecasting [33, 34], which leads them to make fallacious predictions (i.e., base-rate neglect), and ignore decision aids [35].

[1] Moore and Healy [8] differentiated various facets of overconfidence and defined three types of overconfidence: (i) overestimation of one's actual performance; (ii) overplacement of one's performance relative to others; and (iii) overprecision in one's belief (i.e., miscalibration). The respective categorization is widely accepted in the literature, but we continue to use the term "overconfidence" by referring to (iii) in this paper.

Several psychological factors have been offered in the literature to explain why the overconfidence bias occurs. Anchoring is considered as an important factor that gives rise to the overconfidence bias, considering the fact that confidence intervals are set too close to the "best estimate" [1]: The best estimate serves as an "anchor" point from which people might insufficiently adjust their estimations for extreme points in the probability distributions. Another explanation is that the overconfidence bias occurs because, when providing a subjective confidence interval, people prefer to be more precise than to be more accurate in order to adhere to conversational norms (i.e., narrower intervals are preferred to wider intervals, even when the latter would include the correct answers, see [36, 37]. The limited capacity of human working memory is often held responsible for the occurrence of the bias as well [23]. Such a memory constraint leads people to hold only a small set of relevant facts and/or estimates in their memory one at a time, and consequently generate estimates with less variance. Another, somewhat more general, factor that is thought to contribute to the overconfidence bias is related to the difficulty that people have in dealing with probabilistic judgments [38]. However, all these explanations receive only weak empirical support, and thus the underlying psychological mechanism for overconfidence remains unknown (for a detailed literature review, see Moore, Tenney & Haran [39]).

A considerable amount of research has proposed various ways to overcome the overconfidence bias in order to improve judgments and avoid such severe consequences mentioned above. Next we briefly review several debiasing techniques offered in the existing literature, with a specific focus on those that are relevant to the present study.

2.2. Debiasing techniques

Overconfidence has been documented as a hard-to-correct bias [8]. The existing literature adopts different approaches in attempting to reduce the overconfidence bias, which can be broadly classified as follows [39].

The first approach to debias overconfidence focuses on encouraging people to consider more information and/or "an alternative". Soll and Klayman [19] asked participants to determine the cutoffs at the top and bottom ends of the range of possible values in order to engage them to search for more information. So instead of asking participants to determine the ends of an 80% interval, they asked participants to specify a number that is low enough such that there is a 90% chance the correct answer is above it, and a number that is high enough such that there is a 90% chance the correct answer is below it. As a result, overconfidence reduced to some extent. Koriat, Lichtenstein, and Fischhoff [40] showed that the overconfidence bias is reduced when participants are asked to list counter arguments for their estimates before they report confidence levels in the accuracy of their choices. Overconfidence has also been shown to be reduced if people are forced to consider the alternative outcomes [41] or other potential outcomes before they estimate outcome probabilities [42]. Blavatskyy [43] used an incentive compatible method, which presumably encouraged participants to question/re-consider the level of confidence in the accuracy of their estimates since they exhibit less confidence when they were asked to bet on their own knowledge.

A second approach, widely used for debiasing, is concentrated on the presentation and/or elicitation format for the question under investigation. Abbas et al. [12] and Seaver et al. [13] showed that participants produce less confidence in their accuracy when they were asked to provide probabilities for fixed values of the random variables than when they were asked to specify values for fixed cumulative probabilities. Teigen and Jørgensen [44] demonstrated that people have less confidence when they estimate values for a pre-determined confidence level. Winman, Hansson, and Juslin [45] proposed an adaptive elicitation method in which the participants were asked to estimate the probability of a pre-generated interval containing the correct answer, and then adjust this estimated probability until it matched the requested level of confidence. This iterative elicitation method resulted in less overconfidence, in comparison to directly produced intervals. Haran, Moore and Morewedge [46] developed another method in which the participants were able to see the entire range of possible intervals, and then asked to estimate the probability that each of these intervals contains the correct answer such that the sum of all probability estimates is amount to one. This method also led people to produce lesser degrees of overconfidence.

Thirdly, providing people with more feedback is also believed to help reducing overconfidence, since feedback serves as a tool allowing people to correct their errors [47]. Several studies indeed demonstrated the positive effect of feedback on reduction in overconfidence in some cases [48,49]. However this effect is susceptible to various factors, such as task difficulty [50], estimation order (i.e., first estimation vs. following estimations, see Baranski and Petrusic [51], Lichtenstein and Fischhoff [52]), and type of judgment (i.e., high probability vs. low probability, see Baranski and Petrusic [51]).

In the present study, we tested three widely employed debiasing techniques in decision analysis, as outlined in the introduction. They follow the first and second approaches described above, as providing feedback about the choices is not feasible in real-world decision analytic interventions, often characterized by long term horizons and one-off decisions. We detail next the experimental method employed in this study.

3. Methods

3.1. Participants & Materials

One hundred and ten undergraduate students (M_{age} = 21.6) from the Polytechnic University of Turin participated in the study as a part of their activity in the course "Environmental Impact Assessment Procedures". All students had previously taken a statistics course and therefore had basic knowledge of the concepts of probability and probability distributions. Participants were randomly assigned to one of the four conditions described in detail below.

Participants received instructions about the study in which they were asked to provide estimates for each of 10 questions, with five questions about environmental issues (i.e., "What is the expected sea level rise in the next 100 years, according to IPCC 2008 report?"), and five general knowledge questions (i.e., "What is the area of Italy?").

Participants were also given a detailed explanation about the answers they would provide, in particular in terms of probabilistic responses and the way their responses would be evaluated. One of these evaluations was by way of counting surprises, i.e. the number of times the true value (not known to the participants) would fall below the 5th percentile of their probability distribution or above the 95th percentile. They also were shown how the scoring rule by Matheson and Winkler [53] would be used to evaluate their probabilistic estimates. They were told that the participant who scored best in the scoring rule evaluation would be awarded 100 Euros in cash to incentivize accuracy.

3.2. Design & Procedure

Participants' subjective probability distributions were elicited either through the *Fixed Probability* method, or the *Fixed Value* method. Through each elicitation method, either the *Counterfactuals* technique or the *Hypothetical Bets* technique was employed. Therefore we have four experimental conditions: Fixed Probability Counterfactual (FP_C), Fixed Probability Hypothetical Bet (FP_H), Fixed Value Counterfactual (FV_C) and Fixed Value Hypothetical Bet (FV_H). The study consisted of three main phases, as follows:

Phase 1: In the FP_C and FP_H conditions, participants were asked to provide the following values for the question under evaluation, respectively: (a) the lowest number such that they are absolutely sure that the true answer would not be below it, (b) the highest number such that they are absolutely sure that the true answer would not be above it, (c) best guess so that the chances of the true answer falling below or above is 50/50 (d) a low end such that there is a 10% chance that the true answer is between this low end point and their lowest value, and (e) a high end such that there is a 10% chance that the true value answer is between this high end point and their highest value. This sequence of elicitation is fairly standard in decision analysis, as it first establishes the three key points of the distribution (min, max, median) and then identified other points that allow to identify the shape of the distribution.

In the FV_C and FV_H conditions, participants were first asked to provide the following values: (a) the lowest number such that they are absolutely sure that the true answer would not be below it, and (b) the highest number such that they are absolutely sure that the true answer would not be above it. Afterwards, the lowest to highest number interval was divided into 50% of the interval value, then 10% and 90% of the interval value, and subjects were asked to provide the cumulative probabilities for each of the three values in between the lowest and the highest value they gave.

While responding, the cumulative probability distributions were plotted on the side of screen. Once these initial estimations were completed, the participants were allowed to revise their estimations (that stayed on the screen throughout the entire task) before they proceeded to Phase 2.

Phase 2: After providing the initial estimates in Phase 1, several debiasing conditions were applied. In the FP_C and FV_C conditions, participants were asked if they could think of explanations under which the true answer was (i) lower than their initial lowest estimate, and (ii) higher than their initial highest estimate. If they stated they could think of an explanation, they were requested to revise their initial estimates downwards for (i), and upwards for (ii). Otherwise no revision was required.

In the FP_H and FV_H conditions, participants were asked to imagine hypothetical betting situations in which they would have to pay 100 Euros if the true value was (i) below their initial estimate for the lowest value, or receive 1 Euro otherwise; and (ii) above their initial estimate for the highest value, or receive 1 Euro otherwise. If they said they would reject the bets, they were asked to revise their initial estimates downwards for (i), and upwards (ii).

Phase 3: Finally, for all four conditions, participants were asked which side of the median (estimated in Phase 1) they would bet on. In the FP method, the participants provided the median. In the FV method, the median was interpolated from the participants' probability judgments assuming a piece-wise linear cumulative distribution. If they stated to place a bet above the median, they were requested to adjust the median upwards, such that they would be indifferent between betting on the lower and upper sides. If they stated that they would bet below the median, they were asked to adjust the median downwards, again to make them indifferent between betting on either side. If they stated that they had no preference between betting above or below their median, no adjustment was required. Once Phase 3 was completed, participants could see a display of their initial and revised cumulative distributions and could make one more revision of the revised distribution. Subsequently, they moved on to the next question.

4. Results

All 110 participants completed the probability estimates for all ten questions. We eliminated estimates that produced non-monotone cumulative distributions, because this indicated that the participant did not understand the instructions. We also eliminated outlier responses defined as the minimum estimate falling above three times the true value or the maximum estimate falling below 1/3rd of the true value.

The question "What is the inclination (in centimeters) of the Piazza Vittorio Veneto" produced many estimates that would eliminate participants by the outlier criterion. We suspect that many participants responded in meters instead of centimeters, but in retrospect this was difficult to trace and correct. We therefore eliminated this question from further analysis.

After eliminating the Piazza Vittorio Veneto question there were 166 outlier responses: 59 were non-monotone and 107 were too extreme to be credible (not meeting the 3 times and 1/3rd criteria below). The remaining data consisted of 824 pairs of initial and revised cumulative probability distributions for nine questions and 110 participants.

We conducted three analyses: Counts of revisions of low or high estimates in response to the counterfactual and betting treatments, surprises, both for initial estimates and revised ones, and counts of revisions of medians and the direction of these revisions (towards or away from the true value). We also analyzed whether there were any differences between the environmental and the general knowledge questions.

4.1 Revisions

The experiment was designed to determine if best practices in decision analysis aimed at reducing overconfidence had an effect. Table 1 shows the number and percentage of revisions of initial estimates depending on the fixed probability vs. fixed value treatments.

Table 1. Counts and percentages of revisions of one or both tails of the probability distributions for the FV vs. FP method

	Revised	Not Revised	Total
Fixed P	136 (37%)	297 (69%)	433
Fixed V	126 (26%)	265 (74%)	391
Total	262	562	824

Only about 1/3 of the participants made revisions. There were slightly more revisions for the fixed probability procedure than for the fixed value procedure, but the difference is not significant.

Table 2 shows that there are significantly more revisions using the counterfactual method than using the betting method ($p<0.001$)[2].

Table 2. Counts and percentages of revisions of one or both tails of the probability distributions for counter-factuals vs. betting debiasing

	Revised	Not Revised	Total
Counterfact	157 (37%)	266 (63%)	423
Hypobets	105 (26%)	296 (74%)	401
Total	262	562	824

4.2 Surprises

Surprises were instances in which true value fell outside the 90% range of the participant's estimates. Table 3 shows that there were a large number of surprises (50% vs. 10% expected with perfect calibration) and that there was no effect of the treatments overall in reducing surprises.

Table 3. Count and percentages of surprises across all conditions and questions

	Initial	Revised
Surprise	411 (50%)	403 (50%)
No Surprise	413 (50%)	421 (50%)
Total	824	824

Table 4 shows the number of surprises after revision for the fixed probability vs. the fixed value conditions. There are slightly less surprises for the fixed value procedure ($p=0.07$).

Table 4. Count and percentages of surprises for the FP vs. FV debiasing methods

	Fixed Prob	Fixed Value	Total
Surprise	226 (52%)	177 (48%)	403
No Surprise	207 (48%)	214 (52%)	421
Total	433	391	824

[2] The probability values are associated with a Chi-square test for independence.

Table 5 shows the number of surprises for the revised estimates for the counterfactual vs. hypothetical bets procedure. There are slightly less surprises for the counterfactual procedure but this effect is not significant.

Table 5. Count and percentages of surprises for the hypothetical bets vs. counterfactual debiasing procedures

	Counterfact	Hypobets	Total
Surprise	202 (48%)	201 (52%)	403
No Surprise	221 (52%)	200 (48%)	421
Total	423	401	824

4.3 Median Displacement

Table 6 shows the results of the median displacement analysis in terms of the count and percentage of participants who responded affirmatively when asked whether they would bet on one or the other side of their estimated median. The results show that about half of the responses were affirmative with a trend for the FV procedure to elicit more affirmative responses (55%) than the FP procedure (42%) ($p<0.001$).

Table 6. Count and percentages of revisions in response to the median betting question

	Bet	Don't Bet	Total
Fixed P	180 (42%)	253 (58%)	433
Fixed V	215 (55%)	176 (45%)	391
Total	395	429	824

To examine, if the shift in response to the betting question was towards or away from the true value, Table 7 shows the count and percentages of the shifts in direction for those who answered affirmatively to the betting question.

Table 7. Counts and percentages of shifts of the median towards or away from the true value

	Away	No Change	Towards	Total
FixedP	57 (32%)	17 (9%)	106 (59%)	180
FixedV	96 (45%)	24 (11%)	95 (44%)	215
Total	153	41	201	395

For the fixed probability method the shift was predominantly towards the true value, but for the fixed value method there was no noticeable movement towards the true value (p=0.013).

We found no significant differences when we analyzed these patters for the environmental questions vs. the general knowledge questions.

5. Conclusion

This study tested several "best practices" of decision analysts to debias the well-known overconfidence bias in the case of continuous probability distributions. It first re-established, as many studies before it, that the overconfidence bias is strong and persistent with about half of the participants exhibiting overconfidence as defined by surprises.

The debiasing treatment to widen the low and high estimates of participants' probability distributions were not very effective. The fixed value method, which has shown some promise in the Seaver et al. [13] and Abbas et al. [12] studies showed some minor improvement in terms of reducing surprises, but not nearly as much as in these previous two studies. This is probably due to the fact that participants started with their own high low and high estimates, followed by a sectioning of the range they defined. Since the bias is mostly a result of defining the low and high initial estimate too narrowly, a better procedure is to provide these estimates externally. Of course, this may involve other anchoring biases as well as being considered information by the participant.

The effects of the betting and counterfactual procedures were also relatively small, but the counterfactual method appeared to do better than the betting procedure.

There was a strong and significant effect of the betting procedures to correct the initial median estimate. Across questions and participants, about half the responses indicated a preference to bet on one side of the initial median or the other, similar to observations by decision analysts in real applications with experts. On the other hand, the improvements in terms of revisions that are closer to the true value are relatively minor, with the revisions leading to more improvement for the fixed probability procedure.

In conclusion, we were able to demonstrate some positive effects of using best practices in debiasing overconfidence, but the effects were not as large as one would hope. Future research should explore using the fixed value method with externally provided lower and upper bounds, more realistic and convincing techniques to expand the range from the minimum to the maximum of the uncertain variable, as well as the effects of training and practice.

6. Acknowledgements

Detlof von Winterfeldt acknowledges the support by the National Science Foundation under grant number SES-1314644. Sule Guney acknowledges the support by the U.S. Department of Homeland Security through the National Center for Risk and Economic Analysis of Terrorism Events under cooperative agreement number 2010-ST-061-RE0001. However, any opinions, findings, conclusions, and recommendations in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation of the Department of Homeland Security.

7. References

[1] Tversky, A. and Kahneman, D. (1974). 'Judgment under Uncertainty: Heuristics and Biases'. *Science 185(4157),* 1124–31.

[2] Kahneman, D. and Tversky, A. (1979). 'Prospect theory: An analysis of decision under risk'. *Econometrica 47*, 263-291.

[3] Kahneman, D. and Tversky, A. (1982). 'The psychology of preferences'. *Scientific American*.

[4] Gilovich, T., Griffin, D. and Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

[5] Montibeller, G. and Winterfeldt, D. (2015). 'Cognitive and Motivational Biases in Decision and Risk Analysis'. *Risk Analysis*. doi: 10.1111/risa.12360

[6] Alpert, M. and Raiffa, H. (1982). 'A progress report on the training of probability assessors'. In D. Kahneman, P. Slovic and A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

[7] Lichtenstein, S., Fischhoff, B. and Phillips, L. D. (1982). 'Calibration of probabilities: The state of the art in 1980'. In D. Kahneman, P. Slovic and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–333). Cambridge, England: Cambridge University Press.

[8] Moore, D. A. and Healy, P. J. (2008). 'The trouble with overconfidence'. *Psychological Review 115(2),* 502–517.

[9] Fischhoff, B. (1982). ‘Debiasing’. In D. Kahneman, P. Slavic and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422-444). Cambridge, England: Cambridge Univ. Press.

[10] Larrick, R. P., Burson, K. A. and Soll, J. B. (2007). ‘Social comparison and confidence: When thinking you’re better than average predicts overconfidence (and when it does not)’. *Organizational Behavior and Human Decision Processes 102(1),* 76-94.

[11] von Winterfeldt, D. and Edwards, W. (1986). *Decision analysis and behavioral research (Vol. 604)*. Cambridge: Cambridge University Press.

[12] Abbas, A. E., Budescu, D. V., Yu, H. T. and Haggerty, R. (2008). ‘A comparison of two probability encoding methods: Fixed probability vs. fixed variable values’. *Decision Analysis 5(4),* 190-202.

[13] Seaver, D. A., Von Winterfeldt, D. and Edwards, W. (1978). ‘Eliciting subjective probability distributions on continuous variables’. *Organizational Behavior and Human Performance 21(3),* 379-391.

[14] Bazerman, M. H. and Moore, D. A. (2013). *Judgment in managerial decision making* (8th ed.). New York: Wiley.

[15] DeBondt, W.F.M. and Thaler, R.H., 1995. ‘Financial Decision-Making in Markets and Firms: A Behavioral Perspective’. In R. Jarrow, V. Maksimovic and W. Ziemba (Eds), *Handbooks in Operations Research and Management Science* (pp. 385-410). Amsterdam: North Holland.

[16] Kahneman, D. (2011). ‘Don’t blink! The hazards of confidence’. New York Times.

[17] Plous, S. (1993). *The Psychology of Judgment and Decision Making.* New York: McGraw-Hill.

[18] Klayman, J., Soll, J. B., Gonzalez-Vallejo, C. and Barlas, S. (1999). ‘Overconfidence: It depends on how, what, and whom you ask’. *Organizational Behavior and Human Decision Processes 79(3),* 216–247.

[19] Soll, J. B. and Klayman, J. (2004). ‘Overconfidence in interval estimates’. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30(2),* 299–314.

[20] Russo, J. E. and Schoemaker, P. J. H. (1992). ‘Managing overconfidence’. *Sloan Management Review 33(2),* 7–17.

[21] Clemen, R. T. and Ulu, C. (2008). ‘Interior additivity and subjective probability assessment of continuous variables’. *Management Science 54(4),* 835–851.

[22] Henrion, M. and Fischhoff, B. (1986). ‘Assessing uncertainty in physical constants’. *American Journal of Physics 54(9),* 791–798.

[23] Juslin, P., Winman, A. and Hansson, P. (2007). ‘The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals’. *Psychological Review 114(3),* 678–703.

[24] McKenzie, C. R. M., Liersch, M. J. and Yaniv, I. (2008). ‘Overconfidence in interval estimates: What does expertise buy you?’ *Organizational Behavior and Human Decision Processes 107,* 179–191.

[25] Morgan, M. G. and Keith, D. W. (2008). ‘Improving the way we think about projecting future energy use and emissions of carbon dioxide’. *Climatic Change 90(3),* 189-215.

[26] Daniel, K. D., Hirshleifer, D. A. and Sabrahmanyam, A. (1998). ‘Investor psychology and security market under- and overreactions’. *Journal of Finance 53(6),* 1839–1885.

[27] Deaves, R., Lüders, E. and Luo, G. Y. (2008). ‘An experimental test of the impact of overconfidence and gender on trading activity’. *Review of Finance 13,* 555–575.

[28] Odean, T. (1999). ‘Do investors trade too much?’ *American Economic Review 89(5),* 1279–1298.

[29] Arkes, H. R., Wortmann, R. L., Saville, P. D. and Harkness, A. R. (1981). ‘Hindsight bias among physicians weighing the likelihood of diagnoses’. *Journal of Applied Psychology 66(2),* 252–254.

[30] Christensen-Szalanski, J. J. and Bushyhead, J. B. (1981). ‘Physicians’ use of probabilistic information in a real clinical setting’. *Journal of Experimental Psychology: Human Perception and Performance 7,* 928–935.

[31] Hill, L. D., Gray, J. J., Carter, M. M. and Schulkin, J. (2005). ‘Obstetrician-gynecologists’ decision making about the diagnosis of major depressive disorder and premenstrual dysphoric disorder’. *Journal of Psychosomatic Obstetrics & Gynecology 26(1),* 41–51.

[32] Oskamp, S. (1965). ‘Overconfidence in case-study judgments’. *Journal of Consulting Psychology 29(3),* 261–265.

[33] Ben-David, I., Graham, J. R. and Harvey, C. R. (2010). *Managerial miscalibration* (No. w16215). National Bureau of Economic Research.

[34] Du, N., Budescu, D. V, Shelly, M. and Omer, T. C. (2011). ‘The appeal of vague financial forecasts’. *Organizational Behavior & Human Decision Processes 114(2),* 179–189.

[35] Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In *Principles of forecasting* (pp. 495-515). Springer US.

[36] Yaniv, I. and Foster, D. P. (1995). ‘Graininess of judgment under uncertainty: An accuracy-informativeness trade-off’. *Journal of Experimental Psychology: General 124(4),* 424–32.

[37] Yaniv, I. and Foster, D. P. (1997). ‘Precision and accuracy of judgmental estimation’. *Journal of Behavioral Decision Making 10(1),* 21–32.

[38] Gigerenzer, G. and Hoffrage, U. (1995). ‘How to improve Bayesian reasoning without instruction: Frequency formats’. *Psychological Review 102(4),* 684–704.

[39] Moore, D. A., Tenney, E. R. and Haran, U. (in press). ‘Overprecision in judgment’. In C. Keren and G. Wu (Eds.), *Blackwell Handbook of Judgment and Decision Making*. New York: Wiley.

[40] Koriat, A., Lichtenstein, S. and Fischhoff, B. (1980). ‘Reasons for confidence’. *Journal of Experimental Psychology: Human Learning and Memory 6(2)*, 107–118.

[41] McKenzie, C. R. M. (1997). ‘Underweighting alternatives and overconfidence’. *Organizational Behavior and Human Decision Processes 71(2),* 141–160.

[42] Hirt, E. R. and Markman, K. D. (1995). ‘Multiple explanation: A consider-an-alternative strategy for debiasing judgments’. *Journal of Personality and Social Psychology 69(6),* 1069.

[43] Blavatskyy, P. R. (2009). ‘Betting on own knowledge: Experimental test of overconfidence’. *Journal of Risk and Uncertainty 38(1),* 39–49.

[44] Teigen, K. H. and Jørgensen, M. (2005). ‘When 90% confidence intervals are 50% certain: On the credibility of credible intervals’. *Applied Cognitive Psychology 19(4),* 455–475.

[45] Winman, A., Hansson, P. and Juslin, P. (2004). ‘Subjective probability intervals: how to reduce overconfidence by interval evaluation’. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(6),* 1167.

[46] Haran, U., Moore, D. A. and Morewedge, C. K. (2010). ‘A simple remedy for overprecision in judgment’. *Judgment and Decision Making 5(7),* 467–476.

[47] González-Vallejo, C. and Bonham, A. (2007). ‘Aligning confidence with accuracy: revisiting the role of feedback’. *Acta Psychologica 125(2),* 221–39.

[48] Rakow, T., Harvey, N. and Finer, S. (2003). ‘Improving calibration without training: the role of task information’. *Applied Cognitive Psychology 17(4),* 419–441.

[49] Subbotin, V. (1996). ‘Outcome Feedback Effects on Under- and Overconfident Judgments (General Knowledge Tasks)’. *Organizational Behavior and Human Decision Processes 66(3),* 268–276.

[50] Baranski, J. V and Petrusic, W. M. (1994). ‘The calibration and resolution of confidence in perceptual judgments’. *Perception & Psychophysics 55(4),* 412–428.

[51] Baranski, J. V and Petrusic, W. M. (1999). ‘Realism of confidence in sensory discrimination’. *Perception & Psychophysics 61(7),* 1369–1383.

[52] Lichtenstein, S. and Fischhoff, B. (1980). ‘Training for calibration’. *Organizational Behavior and Human Decision Processes 26(2),* 149–171.

[53] Matheson, J. E. and Winkler, R. L. (1976). ‘Scoring rules for continuous probability distributions’. *Management Science 22(10),* 1087-1096.