scientific reports

OPEN

Debiasing training reduces confirmation bias in national risk analysts

Bas Heerma van Voss[1]✉, Haewon Yoon[2], Irene Scopelliti[3], Rowena Zweet[4], Ira Helsloot[5] & Carey K. Morewedge[6]

State risk forecasts are crucial for allocating resources to address international and domestic threats such as war, pandemics, and climate change. These risk forecasts largely rely on human judgment, which is often susceptible to cognitive biases. We conducted an experiment involving the majority of national risk analysts in a European country and a matched sample of masters students to compare their susceptibility to confirmation bias and bias blind spot in judgments related and unrelated to national risk. Additionally, we evaluated the effectiveness of a one-shot debiasing training intervention across both samples. We find that analysts exhibit less confirmation bias than students within and outside risk-related judgments. Crucially, a one-shot debiasing training session reduced confirmation bias in both analyst and student groups. These findings suggest that cost-effective debiasing interventions can improve expert judgment in national risk forecasting and provide evidence that experience and expertise reduce cognitive bias more broadly than previously recognized.

Assessments of the risk of a pandemic, war, flood, or nuclear accident guide policy in domains central to the safety and security of nations. Distortions of such risk assessments critically affect national risk analysis. State risk forecasting relies on expert judgment by national risk analysts[1]. Unfortunately, even experts exhibit systematic errors when predicting and forecasting future events[2]. A variety of biases and heuristics influence the complex task of making policy relevant estimates of future risks, including confirmation bias, availability bias, and bias blind spot[3,4]. The impact and direction of some cognitive biases on risk forecasts is predictable. Availability bias, for example, can lead analysts to overweigh the risk of disasters that are easily remembered. Other biases distort judgments in less predictable directions. Confirmation bias, which leads people to search for, interpret, and overweigh information that confirms their prior hypothesis, could lead analysts to underestimate or overestimate a risk, depending on their priors. Because the direction and level of bias is often unknown, forecasts cannot be corrected *ex-post*. Therefore, the influence and reduction of cognitive bias on assessments of national risk analysts are matters of great policy relevance.

Cognitive biases affect the judgments of experts and novices in personal life and in professional domains such as business and finance, medicine and pharmacology, law and forensic sciences, and even philosophy[5–11]. However, the extent to which experts differ from novices in their susceptibility to biases varies considerably across domains, judgments, and cognitive biases. For example, physicians, students, and patients are similarly affected by framing and loss aversion in their preference for medical treatments with differing levels of risk[12]. Conversely, experts exhibit less bias than novices in other domains such as complex subjects and judgments involving repetitive tasks[13].

There is considerable individual-level variation in the ability to assess risks[14], and the impact of cognitive bias in the domain of national risk analysis has been established within specific cases, such as analyses that supported the US government's decision to go to war in Iraq[15]. However, national risk analysts are a rare enough group that, to the best of our knowledge, no published experiments testing the incidence of cognitive biases on this group exist. Considering the stakes of accurate national risk forecasts and their potential impact on national welfare, understanding the impact of cognitive biases on the judgments of national risk analysts is of crucial importance.

[1]Department of Economic Policy & Research, Dutch Central Bank, Amsterdam, The Netherlands, Department of Public Administration, Radboud University, Nijmegen, The Netherlands. [2]School of Psychological and Behavioral Sciences, Southern Illinois University, Carbondale, USA. [3]Bayes Business School, City St Georges, University of London, London, UK. [4]Department of Public Administration, Radboud University, Nijmegen, The Netherlands. [5]Department of Public Administration, Radboud University, Nijmegen, The Netherlands, PEASEC, Technical University Darmstadt, Darmstadt, Germany, Crisislab, Renswoude, The Netherlands. [6]Questrom School of Business, Boston University, Boston, USA. ✉email: b.s.heerma.van.voss@dnb.nl

We examine the influence of confirmation bias on national risk analysts, as well as their susceptibility to the bias blind spot, and whether debiasing interventions are effective with a population with this degree of expertise. Confirmation bias entails the tendency to selectively search for, interpret, and weigh information in ways that support the hypothesis or belief being tested[16,17]. Since early demonstrations by Wason[18], a variety of forms of confirmation bias have been identified[19]. Studies of rule testing strategies, for example, show that people search for evidence that would confirm rather than disconfirm the rule they are testing. There has been debate on whether confirmation bias is an optimal strategy for human decision-making under conditions of bounded computational capacity[19,20]. Indeed, there may be settings in which our tendency to confirm rather than falsify our beliefs yields optimal results[21]. However, qualitative investigations suggest that confirmation bias can profoundly impair the accuracy of national risk analysis. For example, post-mortems on the decision to invade Iraq found that American national risk analysts 'tunneled' early on the conclusion that there were weapons of mass destruction in Iraq. The impact of confirmation bias on the decision to go to war was confirmed by analysts involved in the assessments[15,22]. Confirmation bias could affect the national risk analysis process at different stages. The various sources from which risk analysts acquire information make the true value of each piece hard to ascertain. Risk analysts often make judgments involving uncertainty, such as estimates of the validity of judgments by other experts (e.g., climate researchers) or predicting the future behavior of individuals (e.g., politicians), and they require those uncertainties to be translated into probabilities.

Despite their professional training, expertise, and experience, confirmation bias in risk assessments may escape detection by national risk analysts. People typically have access to the output of their intuitive judgments ("Nuclear power plants feel unsafe") but lack access to the associative process by which their intuitive judgments are made. Because people rely on introspection to detect biases in their judgments, they less readily detect bias in their own judgments than in the judgments of others and algorithms[23,24] whom they instead evaluate based on their behavior[25,26]. This bias blind spot is not attenuated by cognitive sophistication[27], and while it varies substantially among individuals, it does not correlate with intelligence, the actual incidence of bias, or personality traits such as conscientiousness or self-esteem[28]. Although bias blind spot may not directly impact risk assessments, it reduces the efficacy of debiasing interventions[27], and could thus indirectly affect risk analysis by limiting learning effects.

Evidence suggests that even large economic incentives influence the prevalence of cognitive biases in decision making[29,30]. By contrast, the impact of cognitive biases, including confirmation bias and bias blind spot, can be reduced through debiasing interventions such as structured analytical techniques[31–34], training[35,36], and observational learning[37]. Debiasing has been the subject of study for decades[38] and there is evidence that debiasing training can transfer across domains and to the field[35]. The uptake of debiasing techniques by risk analysts is not widespread[32,39,40], possibly due to the controversy regarding their efficacy, the potential for a floor effect limiting their efficacy, or to bias blind spot leading analysts to perceive their own judgments as unbiased.

We conducted an experiment that assessed and attempted to reduce two cognitive biases in a sample of over half of the national risk analyst population in a European country (N = 71) and a comparison sample of students enrolled in master's programs in risk analysis (N = 118). We examined the incidence of confirmation bias and the bias blind spot in risk assessments within the two groups, and tested whether the prevalence of the biases is consistent or differs within and outside domains in which risk analysts are experts (i.e., national risks). We also examined whether a debiasing training intervention of the kind that is effective on novice samples[35,36] can effectively mitigate cognitive biases on such an expert sample.

Experiment, sample, and hypotheses

We conducted an experiment on a group of national risk analysts and a group of master's students in risk-related graduate programs. The national risk analyst group is a unique sample of contributors to a National Risk Assessment (NRA) in a European country, with all contributors selected by its government. As part of our agreement of anonymity, we do not disclose the name of the country. NRAs, as standardized by the OECD, are conducted once every five years to map societal risks. To produce an NRA, national governments group generalist societal risk experts with domain risk experts with expertise in (a subset of) territorial, physical, ecological, economic and social risks. Together, the analysts identify the most significant societal risks and estimate their impact and probabilities. The assessments inform risk prevention and mitigation policy. We conducted our experiment at the concluding conference of an NRA, which is a rare occurrence bringing together a large sample of national level risk analysts in one room.

We compared the expert risk analysts to a sample of students enrolled in master's programs with an emphasis on risk policy. These students are similar in demography to professional national risk analysts in several aspects, such as gender balance and proficiency in English. None of the participants received monetary incentives. For the risk analysts, the experiment and training were integrated into a conference session. For the students, the experiment and debiasing training were integrated into their curriculum. In the preregistration (AsPredicted.org, #94,994), we planned to recruit a sample of 50–70 analysts and a sample of 120 students. Both groups were naturally limited in size, particularly the risk analyst sample, but were the upper bound of the samples achievable in this context. With 71 analysts in total and 56 valid responses, the sample comprised between 50% and 60% of all national level risk analysts within this country.

In the experiment, we administered scales measuring confirmation bias and the bias blind spot to participants in both groups, both before and after a debiasing training intervention. The measures and the intervention were administered in a similar controlled seminar room setting for both groups. The scales assessed biases in judgments both within the risk domain (e.g., flood risk, nuclear safety, and terrorism) and in domains without a link to national risk policy. The confirmation bias measures assessed the tendency to test hypotheses with evidence that would confirm rather than falsify them. The bias blind spot measures assessed the extent to which respondents think they are more or less biased than the average person.

The debiasing intervention consisted of a scripted explanation of decision making biases and their effects, and of a group exercise: Carter Racing, a well-known business case adapted from the decision by NASA to launch the space shuttle Challenger, which illustrates the effects of confirmation bias on decision making (e.g., 41,42). In total, the debiasing intervention took approximately 40 min. A full description of the experiments is available in the methods section, and the materials used in the Supplementary Information.

This paradigm allowed us to answer three questions. First, to what extent do confirmation bias and bias blind spot influence the judgments of national level risk analysts compared to a similar inexperienced sample? Second, do the cognitive biases similarly influence judgments unrelated and related to national risk, a domain in which national risk analysts have more expertise and experience? Third, do the groups similarly benefit from a debiasing intervention?

With regards to confirmation bias, we hypothesized (H1) that professional risk analysts would exhibit confirmation bias to the same extent as students on general domain tests of confirmation bias, but professional risk analysts should exhibit less confirmation bias than students on risk domain tests. We based this hypothesis on literature on expertise and cognitive biases, which demonstrated the benefits of experience and expertise to have limited effects outside of the domain of expertise[13,41,42]. We also hypothesized (H2) that students would exhibit a greater reduction in confirmation bias following a debiasing intervention than analysts within the risk domain, but that they would show similar reductions within general domain tests of confirmation bias. We based this hypothesis on the assumption that lower confirmation bias scores among analysts in the risk domain at pretest would allow less room for improvement (i.e., a floor effect). With regards to bias blind spot, we hypothesized (H3) that at pretest, professional risk analysts should exhibit greater bias blind spot on both general and risk domain tests of bias blind spot than students. We hypothesized (H4) that at posttest, students would show a greater reduction in bias blind spot following a debiasing intervention in both general and risk domain tests of bias blind spot. Both hypotheses are based on the assumption that analysts would be reluctant to admit the impact of cognitive biases on their judgment and decisions, as it could be more threatening to experts, whose professional status depends on accuracy[43,44].

Results

Descriptive statistics can be found in the Supplementary Information. We report means in the paragraphs that follow.

Confirmation bias

We examined the susceptibility of the analyst and student groups to confirmation bias using scales adapted from previous work[28]. In addition to items used in previous work, we created parallel items pertaining to national risk. Items were scored as indicative of confirmation bias[1] or not (0), and these scores were averaged. Averages ranged from 0 (indicating no confirmation bias) to 1 (indicating confirmation bias on all survey items). One-sample t-tests (against 0) showed that analyst and student groups demonstrated confirmation bias in both general and risk domains both at pretests and posttests (analysts; all Ms > 0.351, all ts > 14.052, all ps < 0.001, all ds > 1.895; students; all Ms > 0.453, all ts > 21.245, all ps < 0.001, all ds > 1.973,).

We next compared groups (analysts, students; between-subjects), domains (risk, general; within-subjects) and the effect of the intervention (pretest, posttest; within-subjects) in a three-way mixed ANOVA, which revealed significant main effects of group [$F(1, 169) = 25.928$, $p < 0.001$, $\eta_p^2 = 0.133$] and intervention [$F(1, 169) = 20.744$, $p < 0.001$, $\eta_p^2 = 0.109$], suggesting that students demonstrated more confirmation bias than analysts and that both groups demonstrated a significant reduction in confirmation bias from the pretest before the intervention to the posttest after the intervention. The analysis did not yield a significant effect of domain ($F < 1$). There were no significant interactions (all $Fs < 1$). See Fig. 1. These findings supported our first and second hypotheses (H1 and H2) with regards to confirmation bias being lower for risk analysts than students in assessments of national risk. Contrary to our predictions and the literature, confirmation bias was also lower in risk analysts than students in more general domain assessments.

Contrary to our hypothesis that debiasing training would benefit students more (H2), debiasing training was similarly effective at reducing confirmation bias for analysts and students. In an ANCOVA with pretest scores as a covariate, we found that participants scored higher at pretest than at posttest (general: $F(1, 168) = 104.579$, $p < 0.001$, $\eta p^2 = 0.384$ & risk: $F(1, 168) = 45.999$, $p < 0.001$, $\eta p^2 = 0.215$). Moreover, controlling for pretest scores, analysts and student groups did not show significant differences at posttest (general: $F(1, 168) = 1.213$, $p = 0.272$ & risk: $F(1, 168) = 2.004$, $p = 0.159$).

These results are robust to the inclusion of demographic control variables. Previous knowledge of cognitive biases, levels of education, work experience, proficiency in English, gender, and age did not have a significant effect on pretest scores in a regression analysis (all *P's* > 0.23) and did not eliminate the difference in bias incidence across groups, which was still significant ($P = 0.012$). Similarly, the lack of a relationship between group and the debiasing effect was robust to demographic controls. Gender, however, showed a significant impact on debiasing effects in the regression analysis, with female participants showing more improvement ($p = 0.008$). See the Supplementary Information for full regression results.

Bias blind spot

We examined general susceptibility to the bias blind spot at pretest and posttest on a validated scale ranging from − 6 to 6. We developed a complementary risk-specific bias blind spot scale by creating items in a national risk domain. Negative scores indicate that participants perceived themselves to be more biased than others. Positive scores indicate that participants perceived themselves to be less biased than others. 0 indicates no bias blind spot. With one sample t-tests (tested against 0), we found that analysts and students exhibited bias blind spot in

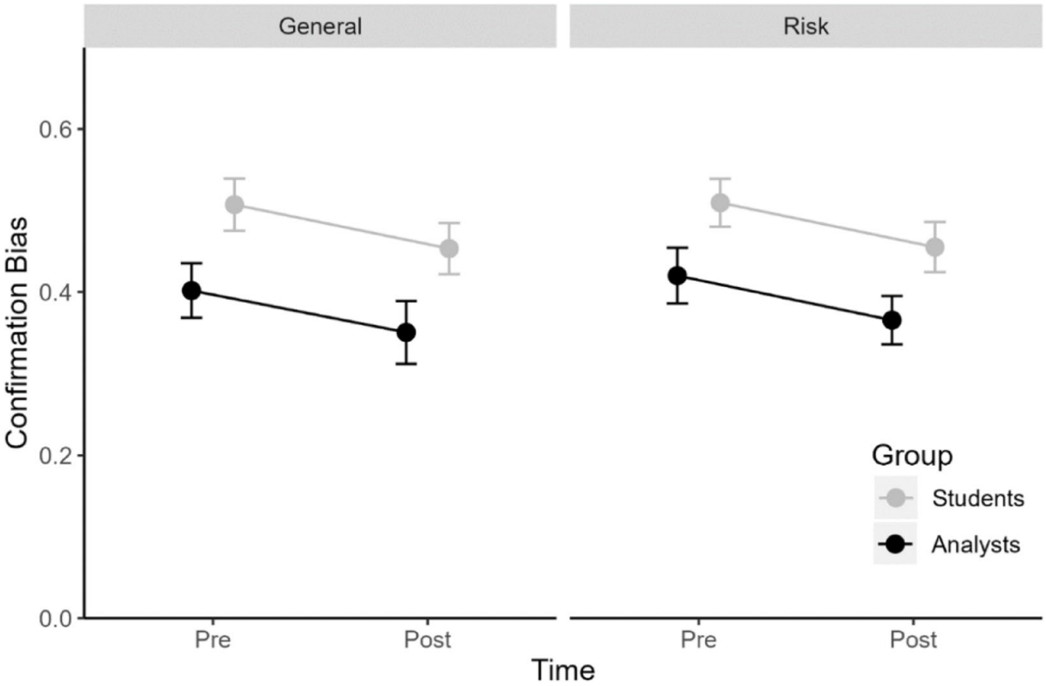

Fig. 1. Confirmation bias in the general and risk domains for student and national risk analyst groups, before and after a debiasing training intervention (error bars represent 95% CIs).

general and risk domains in both pretests and posttests (analysts: all Ms > 0.968, all ts > 6.762, all ps < 0.001, all ds > 0.90 ; students: all Ms > 0.894, all ts > 10.978, all ps < 0.001, all ds > 1.019).

We examined differences between groups and the effects of the intervention on susceptibility to bias blind spot in a three-way mixed ANOVA, as we did for confirmation bias. This analysis revealed main effects of domain [$F(1, 163) = 4.547$, $p = 0.034$, $\eta p^2 = 0.027$] and debiasing intervention [$F(1, 163) = 5.723$, $p = 0.018$, $\eta p^2 = 0.034$], but no difference across groups ($F < 1$). That is, all participants demonstrated more bias blind spot in the general domain than in the risk domain and before the intervention than after the intervention. There were no significant differences between students and analysts. The results suggest that levels of bias blind spot were more similar across experts and novices, and domains of expertise, than the literature suggests (H3 and H4). We did observe a significant domain x group interaction [$F(1, 163) = 6.104$, $p = 0.015$, $\eta p^2 = 0.036$]. A post hoc analysis (Bonferroni-corrected) found that students demonstrated more bias blind spot in the general domain than in the risk domain [$t(163) = 4.140$, $p < 0.001$, $d = 0.324$], but analysts did not. None of the other interactions were significant. See Fig. 2.

Bias blind spot has been unrelated to bias commission in previous work[22] and we found no correlation in our data. In both groups, bias blind spot scores were not correlated with confirmation bias scores in general or risk domains at pretest or posttest [analyst group: all rs < 0.09, all Ps > 0.515; student group: all rs < 0.023, all Ps > 0.808]. In addition, changes in bias blind spot scores from pretest to posttest were not correlated with changes in confirmation bias from pretest to posttest [analyst group: all rs < 0.099, all Ps > 0.495; student group: all rs < 0.032, all Ps > 0.737].

Discussion

Both groups exhibited confirmation bias within and outside their domain of expertise, with analysts exhibiting less confirmation bias than students. While we hypothesized both groups would be susceptible to confirmation bias and that analysts would show less confirmation bias within the risk domain, we did not hypothesize this effect to generalize beyond their domain of expertise (H1). This result stands in contrast to the literature on expertise and cognitive bias[13,41,42], and calls the previous consensus in this literature into question. Our robustness checks suggest this advantage of experience and expertise was not driven by age, higher levels of education, more exposure to research on cognitive biases, or higher proficiency in English (the language in which the study was conducted). As we hypothesized, both analysts and students showed less confirmation bias following a debiasing intervention in all domains. In line with our hypotheses, participants with higher susceptibility to bias at pretest showed greater improvement at posttest (H2). However, contrary to our hypotheses, the effect of the debiasing intervention was not different between groups (H2).

Both analysts and students also showed bias blind spot in all conditions. The similarity in bias blind spot scores across groups, and lack of a correlation between the scores and susceptibility to confirmation bias, suggest that analysts are as unattuned to their bias as students both before and after the debiasing training intervention, contrary to our H3 and H4. Students showed less bias blind spot on risk items than general items. We speculate

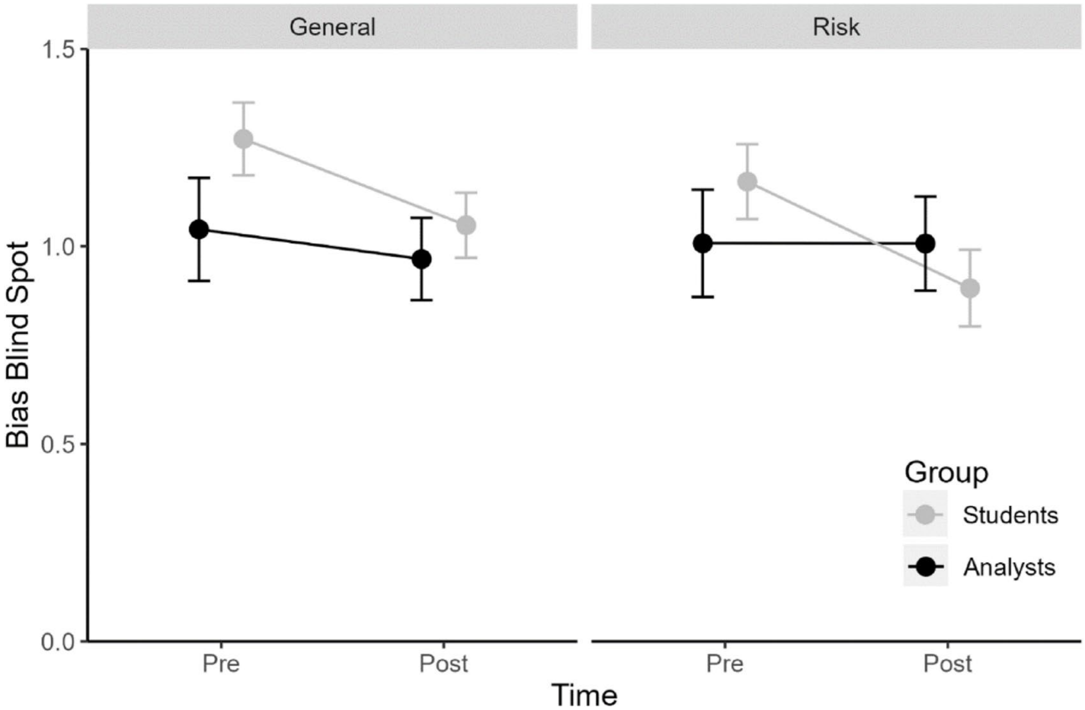

Fig. 2. Bias blind spot in the general and risk domain for student and national risk analyst groups, before and after a debiasing training intervention (error bars represent 95% CIs).

that they perhaps showed more humility regarding their own performance in this domain where the stakes appear higher.

The finding that both groups in all conditions showed confirmation bias and showed a similar improvement after a debiasing intervention is relevant from a policy perspective. In light of research showing that the benefits of debiasing training translate to the field[35], while our data are limited to confirmation bias and bias blind spot, our results suggest that debiasing training interventions is a low-cost intervention with considerable benefits to national risk analysis and the subsequent policy decisions based on their assessments. Bias blind spot was found among all groups in all conditions and did not correlate with confirmation bias scores, suggesting that the participants in both groups were unattuned to their own level of bias. While analysts and students believed themselves to be less biased than average, fortunately, their susceptibility to confirmation bias and bias blind spot still benefitted from debiasing training and all groups showed less confirmation bias and bias blind spot after the intervention. Of course, the cognitive biases were not completely eliminated by debiasing training. Debiasing interventions that reduce bias through changes in choice architecture, like Structural Analytical Techniques[28–34], might provide benefits complementary to training to further improve the risk assessments on which nations rely to prepare for an uncertain future.

Methods

Our methods followed our preregistered plan (AsPredicted.org, #94,994).

Samples

The convenience sample of risk analysts consisted of all participants at a conference in a European country in the fall of 2022 (see Table 1 below for demographics). At the conference, a network of general risk analysts and domain-specific experts came together for the publication of the 5-year-ahead national risk analysis to which they had contributed. The risk analysts were selected by the ministry responsible for delivering the national risk analysis, and worked for the national government, universities, and private risk research institutes. The domain-specific experts were selected for their expertise in analysis of territorial, physical, ecological, economic or social risks, or national risk in general. The analysts spent an average of 27.1% (SD 22.6) of their working hours on risk analysis. The convenience sample of students with which the analysts are compared consists of masters students from three different universities, tested on three different occasions (see Table 1 below for demographics).

We recognize that the small sample of analysts is a limitation of the present study. However, considering the scarcity of national level risk analysts, with the study involving more than half of all contributors to national risk analysis in the country in which the experiment was conducted, and the lack of experimental evidence involving this important professional group in the literature, we consider testing on a smaller sample size warranted.

We examined the sensitivity of the analyses to detect effects and interactions using G*Power. The observed pre-post correlation ($r = 0.675$) and general-risk domain correlation ($r = 0.906$) revealed that the study had sufficient power to detect all bias blind spot interactions. For the Time × Group interaction, the study could detect effects

	Analysts (N=71)	Students (N=118)
Age (M years)	42.9 (12.8)	22.8 (2.0)
Female (frequency, %)	25 (35.2%)	57 (48.3%)
Highest education attained (M; range 0 = high school, 6 = PhD)	4.2 (0.9)	2.2 (1.8)
Work experience (M years)	18.7 (12.6)	1.0 (2.1)
Previous knowledge of cognitive bias (freq, %)	26 (47.3%)	12 (10.3%)
English proficiency (M; range 0–100, self-report)	76.1 (24.8)	77.1 (17.7)
Complete observations (freq, %)	56 (78.9%)	116 (98.3%)

Table 1. Demographics analyst and student samples. Central tendencies and variance are reported with means and standard deviations.

as small as f = 0.088, while the Domain × Group interaction had even better power with a detectable threshold of f = 0.047. The observed Domain × Group interaction effect (f = 0.193, p = 0.015) was more than four times larger than the minimum detectable effect. The observed Time × Group interaction (f = 0.128, p = 0.107) was also well above the detection threshold, indicating that while this effect was not significant, the study had adequate power to detect a significant interaction, reducing the likelihood that the null result was due to insufficient sensitivity. For the Time × Group interaction in confirmation bias, using the pre-post correlation (r = 0.567), the study could detect effects as small as f = 0.100, while the Domain × Group interaction showed reduced power due to the low correlation between general and risk confirmation bias domains (r = 0.190), requiring effects of f = 0.137 to achieve adequate detection. The observed Time × Group effect (f = 0.023, p = 0.762) was only 23% of the minimum detectable effect, while the Domain × Group effect (f = 0.024, p = 0.758) was only 17.5% of its detection threshold. In short, both interactions showed small observed effects well below the respective detection thresholds, supporting the conclusion that confirmation bias was not more effectively debiased between groups and there were no domain-specific effects of expertise.

The experiment was carried out in accordance with relevant guidelines and regulations. Ethics approval was obtained from the Assessment Committee Faculty of Law and Nijmegen School of Management (EACLM, Ref No: 2022.01). The experiment was conducted in controlled classroom settings. As preregistered, at the participant level, we excluded observations from those participants who did not fill in at least partially both the pre and posttest survey, or that demonstrated a clear lack of understanding of the questionnaires, for example by selecting more answers than allowed by a question. In addition, at the questionnaire level, we discarded all questionnaires for which more than 20 percent of the items were not filled in. A total of 56 analysts and 116 students completed both questionnaires successfully. There is a lower share of analysts who completed the questionnaires because several could only attend either the session before or after the debiasing training because of prior work obligations. Informed consent was obtained from all subjects and confirmed in writing.

Measuring bias

In the experiment, participants were tested for bias on the basis of questionnaires containing scale items. The scales were based on measures from Morewedge et al.[36] for confirmation bias and Scopelliti et al.[28] for bias blind spot. We measured confirmation bias using two types of items. The first type (Rule Identification Cards, RIC) was based on Wason's (1960) card selection task (3 items); the second type (Applied Hypothesis Testing, AHT) on Tschirgi's (1980) cause identification paradigm (3 items). We used scores on both types of items to compute an overall confirmation bias score. For both types of confirmation bias scales and for the bias blind spot scales, we created risk-domain scales modeled after the general items. The risk-domain scales were piloted on an online sample (N = 151) to ensure reliability, with the risk-AHT scale consisting of 7 items (α = 0.67) and the risk RIC consisting of 6 items (α = 0.52). This approach resulted in four confirmation bias subscales: risk AHT, general AHT, risk RIC, and general RIC. For bias blind spot, we also had both general domain and risk domain scales.

In the questionnaires, the confirmation bias scales preceded the bias blind spot scales. Risk and general domain questions were administered in a randomized order. In total, there were four different sets of two questionnaires (one for pre-debiasing testing, one for post-debiasing testing). The questionnaires allowed us to calculate confirmation bias and bias blind spot scores for each participant, both in general and in risk domain, before and after the debiasing intervention. See the Supplementary Information for the full scales.

We analyzed these bias scores using 2 (analysts vs. students) × 2 (risk vs. general) × 2 (pre- vs. post-debiasing) mixed ANOVAs. We control for demographic variables using linear regression. One-sample t-tests were used to test whether groups displayed bias at all, i.e., whether mean bias scores were significantly different from 0.

Experimental setup

Experiment sessions took place in seminar rooms. Participants were seated apart so they could not speak. They were not allowed to use online resources during the experiment and had 40 min to fill out the pretest questionnaire. This was followed by the debiasing intervention, which consisted of two phases. First, participants were exposed to a 15-min scripted lecture on cognitive biases. Following a 10-min break, they then collectively completed the Carter Racing Case, a business case that simulates the NASA decision to launch the Space Shuttle Challenger, designed to demonstrate how confirmation bias influenced NASA engineers' decision-making[45]. Following the case, participants had 30 min to finish a second questionnaire, including different but similar scales. The questionnaires were distributed such that half the group had the first set of scales at pretest and the second at posttest, while for the other half of the group the order was reversed. This approach aimed to ensure that variations in scales would not affect the relative performance in pretests and posttests at the group level or influence the effect of the debiasing intervention. See Supplementary Information Sects. 1–3 for the materials used.

Data availability

As part of the agreement of anonymity between the researchers and participants, the data collected for this article will not be made publicly available. They are available for reproduction and research by peers via the corresponding author.

Received: 19 July 2024; Accepted: 12 November 2025
Published online: 24 November 2025

References

1. Nas, I., Helsloot, I. & Cator, E. Of critical importance: Toward a quantitative probabilistic risk assessment framework for critical infrastructure. *Conting. Crisis Mgmt* **1468–5973**, 12427 (2022).
2. Tversky, A. & Kahneman, D. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* **185**, 1124–1131 (1974).
3. R. J. Heuer, *Psychology of intelligence analysis* (Center for the Study of Intelligence, Central Intelligence Agency, 1999).
4. E. Yudkowsky, Cognitive biases potentially affecting judgement of global risks In *Global Catastrophic Risks*, M. Cirkovic, N. Bostrom, Eds. (Oxford University Press, 2008).
5. Fasolo, B., Heard, C. & Scopelliti, I. Mitigating cognitive bias to improve organizational decisions: An integrative review, framework, and research agenda. *J. Manag.* **51**(6), 2182–2211 (2025).
6. Morewedge, C. K. & Kahneman, D. Associative processes in intuitive judgment. *Trends Cogn. Sci.* **14**, 435–440 (2010).
7. Nuijten, A., Benschop, N., Rijsenbilt, A. & Wilmink, K. Cognitive biases in critical decisions facing SME entrepreneurs: An external accountants' perspective. *Adm. Sci.* **10**, 89 (2020).
8. Fisher, K. L. & Statman, M. Cognitive biases in market forecasts. *J. Portf. Manag.* **27**, 72–81 (2000).
9. Segall, M. & Chadwick, A. The risks of subconscious biases in drug-discovery decision making. *Future Med. Chem.* **3**, 771–774 (2011).
10. Cooper, G. S. & Meterko, V. Cognitive bias research in forensic science: A systematic review. *Forensic Sci. Int.* **297**, 35–46 (2019).
11. Dror, I. et al. Cognitive bias in forensic pathology decisions. *J. Forensic Sci.* **66**, 1751–1757 (2021).
12. McNeil, B. J., Pauker, S. G., Sox, H. C. & Tversky, A. On the elicitation of preferences for alternative therapies. *N. Engl. J. Med.* **306**, 1259–1262 (1982).
13. Kynn, M. The "Heuristics and Biases" bias in expert elicitation. *J. R. Stat. Soc. Ser. A Stat. Soc.* **171**, 239–264 (2008).
14. Mellers, B. et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspect. Psychol. Sci.* **10**, 267–281 (2015).
15. Chang, W. & Tetlock, P. E. Rethinking the training of intelligence analysts. *Intell. Nat. Secur.* **31**, 903–920 (2016).
16. Willingham, D. T. Critical thinking: Why is it so hard to teach?. *Arts Educ. Policy Rev.* **109**, 21–32 (2008).
17. Murata, A., Nakamura, T. & Karwowski, W. Influence of cognitive biases in distorting decision making and leading to critical unfavorable incidents. *Safety* **1**, 44–58 (2015).
18. Wason, P. C. On the failure to eliminate hypotheses in a conceptual task. *Q. J. Exp. Psychol.* **12**, 129–140 (1960).
19. Klayman, J. Varieties of confirmation bias. *Psychol. Learn. Motiv.* **32**, 385–418 (1995).
20. Peters, U. What Is the function of confirmation bias?. *Erkenn* **87**, 1351–1376 (2022).
21. Lefebvre, G., Summerfield, C. & Bogacz, R. A normative account of confirmation bias during reinforcement learning. *Neural Comput.* **34**, 307–337 (2022).
22. Stuster, J. D. The Iraq Syndrome. *Foreign Policy 2013* accessed 20 Jan 2023.
23. Ehrlinger, J., Gilovich, T. & Ross, L. Peering into the bias blind spot: people's assessments of bias in themselves and others. *Pers. Soc. Psychol. Bull.* **31**, 680–692 (2005).
24. Celiktutan, B., Cadario, R. & Morewedge, C. K. People see more of their biases in algorithms. *Proc. Nat. Acad. Sci.* **121**(16), e2317602121 (2024).
25. Pronin, E., Lin, D. Y. & Ross, L. The bias blind spot: perceptions of bias in self versus others. *Pers. Soc. Psychol. Bull.* **28**, 369–381 (2002).
26. Pronin, E., Gilovich, T. & Ross, L. Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others. *Psychol. Rev.* **111**, 781–799 (2004).
27. West, R. F., Meserve, R. J. & Stanovich, K. E. Cognitive sophistication does not attenuate the bias blind spot. *J. Pers. Soc. Psychol.* **103**, 506–519 (2012).
28. Scopelliti, I. et al. Bias Blind spot: structure, measurement, and consequences. *Manag. Sci.* **61**, 2468–2486 (2015).
29. Enke, B. et al. Cognitive biases: Mistakes or missing stakes?. *Rev. Econ. Stat.* **105**, 818–832 (2023).
30. Camerer, C. F. & Hogarth, R. M. The effects of financial incentives in experiments: A review and capital-labor-production framework. *J. Risk Uncertain.* **19**, 7–42 (1999).
31. J. R. Cooper, Curing analytic pathologies: pathways to improved intelligence analysis. (2005) (Published online by the Centre for the Study of Intelligence). Accessed on 20 January 2023.
32. I. Belton, M. Dhami, Cognitive Biases and Debiasing in Intelligence Analysis In *Routledge Handbook of Bounded Rationality*, R. Viale, Ed. (Routledge, 2021).
33. Chang, W., Berdini, E., Mandel, D. R. & Tetlock, P. E. Restructuring structured analytic techniques in intelligence. *Intell. Nat. Secur.* **33**, 337–356 (2018).
34. Dror, I. E. & Kukucka, J. Linear sequential unmasking-expanded (LSU-E): A general approach for improving decision making as well as minimizing noise and bias. *Forensic Sci. Int. Synerg.* **3**, 100161 (2021).
35. Sellier, A.-L., Scopelliti, I. & Morewedge, C. K. Debiasing training improves decision making in the field. *Psychol. Sci.* **30**, 1371–1379 (2019).
36. Morewedge, C. K. et al. Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights Behav. Brain Sci.* **2**, 129–140 (2015).
37. Yoon, H., Scopelliti, I. & Morewedge, C. K. Decision making can be improved through observational learning. *Organ. Behav. Hum. Decis. Process.* **162**, 155–188 (2021).
38. B. Fischoff, Debiasing in *Judgment Under Uncertainty: Heuristics and Biases*, In D. Kahneman, P. Slovic & A. Tversky, Eds. (Cambridge University Press, 1982).
39. Tetlock, P. E. & Gardner, D. *Superforecasting: The Art and Science of Prediction* (Broadway Books, 2015).
40. Mandel, D. R. & Tetlock, P. E. Correcting judgment correctives in national security intelligence. *Front. Psychol.* **9**, 2640 (2018).
41. Reeves, L. & Weisberg, R. W. The role of content and abstract information in analogical transfer. *Psychol. Bull.* **115**, 381–400 (1994).

42. Barnett, S. M. & Ceci, S. J. When and where do we apply what we learn?: A taxonomy for far transfer. *Psychol. Bull.* **128**, 612–637 (2002).
43. Kang, S. & Kim, J. W. The fragility of experts: A moderated-mediation model of expertise, expert identity threat, and overprecision. *Acad. Manag. J.* **65**, 577–605 (2022).
44. Campbell, W. K. & Sedikides, C. Self-threat magnifies the self-serving bias: A meta-analytic integration. *Rev. Gen. Psychol.* **3**, 23–43 (1999).
45. J. Brittain, S. Sitkin, Carter Racing Case and Teaching Notes (Stanford Case System No. SOB-24). Stanford, CA: Graduate School of Business, Stanford University (1988).
46. Auer-Rizzi, W. & Berry, M. Business vs. cultural frames of reference in group decision making: interactions among Austrian, Finnish, and Swedish business students. *J. Bus. Commun.* **37**, 264–288 (2000).
47. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 2013).

Acknowledgements
The authors would like to thank Jeroen van Baar for his thoughtful comments on the research design and the main manuscript.

Author contributions
BH came up with the initial idea for the research. CM, IS, and HY refined the experiment and provided resources such as the items used in the questionnaires. CM tested the items on MTurk. IH provided access to the networks of analysts and students for testing. RZ and BH prepared and conducted the experiments. HY and IS analyzed the resultant data, with all authors discussing and commenting on the results. HY, BH and IS drafted the main manuscript text, which was refined by CM.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-28794-w.

Correspondence and requests for materials should be addressed to B.H.v.V.