Allegiance Bias in Statement Reliability Evaluations Is Not Eliminated by Falsification Instructions

Citation for published version (APA):

Document status and date:
Published: 01/07/2020

DOI:

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 03 Nov. 2021

Allegiance Bias in Statement Reliability Evaluations Is Not Eliminated by Falsification Instructions

Melanie Sauerland[1], Henry Otgaar[1,2], Enide Maegherman[3,4], and Anna Sagana[1]

[1]Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, The Netherlands
[2]Faculty of Law, Catholic University of Leuven, Belgium
[3]Department of Criminal Law and Criminology, Faculty of Law, Maastricht University, The Netherlands
[4]Department of Psychology, University of Gothenburg, Sweden

Abstract: Are expert witnesses biased by the side (defense vs. prosecution) that hires them? We examined this issue by having students act as expert witnesses in evaluating interviews in a child sexual abuse case (Experiment 1, N = 143) and tested the value of an instruction to counteract such allegiance effects. The intervention concerned an instruction to consider arguments both *for* and *against* the given hypothesis (i.e., two-sided instructions; Experiment 2, N = 139). In Experiment 3 (N = 123), we additionally provided participants with three different scenarios. Participants received a case file regarding a case of alleged sexual abuse. With the file, participants received an appointment letter emphasizing elements of the file that questioned (defense) or supported (prosecution) the veracity of the accusation. Participants displayed allegiance bias (Experiments 1–3), but two-sided instructions were not successful in eliminating allegiance bias (Experiments 2 and 3). The findings underscore the importance of legal safeguards in expert witness work.

Keywords: alternative scenario, partisanship, confirmation bias, expert witness, adversarial allegiance

Psychologists have long recognized that context information can bias expert witnesses' decision-making and impact the outcome of their case evaluations. Being retained by the defense or prosecution can act as such context information and affect expert witnesses' impartiality. Specifically, evidence accumulates that the experts' assessment of a case can drift from objectivity to support the retaining party's opinion. This effect is known as *allegiance bias*.

In the first demonstration of allegiance bias, PhD students who studied a criminal case vignette rated the defendant more often as *not guilty by reason of insanity* when hired by the defendant, compared to the prosecution (Otto, 1989). Defense-appointed participants also rated the defendant's behavior to be more strongly related to his mental state than prosecution-appointed participants. However, this early exploration did not reveal allegiance bias in a civil case condition, or when using mental health professionals (Beckham, Annis, & Gustafson, 1989).

Twenty years later, a wave of field studies consistently reported higher scores on actuarial risk assessment measures in experts appointed by the prosecution, compared to the defense (Edens, Cox, Smith, DeMatteo, & Sörman,

2015; Lloyd, Clark, & Forth, 2010; Murrie et al., 2009). However, these studies have been criticized for the lack of random assignment to conditions. Hence, differences might have been caused by allegiance bias, but could also be the result of attorney or expert selection bias, unique characteristics of the cases, individual differences among experts, or a combination (Guarnera, Murrie, & Boccaccini, 2017). Murrie, Boccaccini, Guarnera, and Rufino (2013) addressed some of these concerns in a more controlled experimental setting and still found allegiance bias.

Although the mechanism behind the effect is still unclear, allegiance bias seems to be most closely related to confirmation bias. Experiencing the uncomfortable feeling of cognitive dissonance, due to conflicting information, can be reduced by achieving consonance (Festinger, 1957). We can achieve consonance by preferring evidence that confirms an existing belief and such preference can result in confirmation bias. Confirmation bias has repeatedly been demonstrated in a wide array of experimental research in the legal psychological literature, including the consideration of evidence (e.g., Ask & Granhag, 2005;

Ask & Granhag, 2007; Marksteiner, Ask, Reinhard & Granhag, 2011; O'Brien, 2009).

Research thus far mainly focused on allegiance bias in the context of risk assessment, whereas little work has been devoted to other domains of expertise (but see McAuliff & Arter, 2016) or possible interventions. Here, we studied allegiance bias for the assessment of witness testimony and a potential technique to alleviate the bias. Student participants acted as expert witnesses and received a case file concerning alleged child sexual abuse, accompanied by a letter from either the defense or the prosecution. The letters emphasized elements of the file that questioned (defense) or supported (prosecution) the veracity of the accusation. We expected this manipulation to result in allegiance bias, with defense condition participants being more skeptical about the reliability of the testimony than prosecution condition participants (Hypothesis 1).

Experiment 2 tested the value of instructions to look at information for and against the hypothesis. We based this intervention on the general scientific principle of falsification (Popper, 1959) and the notion that allegiance bias might be caused by the fact that expert witnesses do not actively search for information that can falsify certain hypotheses (Reese, 2012). Indeed, people who receive a prompt to consider an opposite position have shown to display less bias in social judgment (Arkes, 1991; Lord, Lepper, & Preston, 1984) and considering why a hypothesis might be wrong diminished students' confirmation bias in a criminal investigation (O'Brien, 2009). Studying contradicting information has also been successful in reducing hindsight bias (Slovic & Fischhoff, 1977), and confirmatory hypothesis testing (Wason, 1960). Our instruction advised participants to look for elements in the case file that spoke for or against the reliability of the statements, thereby creating two-sided argumentation from the expert. We expected participants who received standard instructions to be more vulnerable to allegiance bias than those who received instructions to look for information for two sides (i.e., an interaction between retaining party and instructions; Hypothesis 2). In Experiment 3, we strengthened the two-sided instructions by additionally providing participants with specific alternative scenarios. The use of scenarios has been advocated, to enhance reasoning about evidence and causal explanations (Rassin, 2018; Van Koppen & Mackor, 2019), thus providing more guidance for participants.

Method

Participants

In Experiment 1, 143 students enrolled in a legal psychology Bachelor's course at the Law Faculty participated.

Participants did not receive reimbursement. We preregistered Experiments 2 (https://osf.io/db5w8) and 3 (https://osf.io/6e4fn) prior to data collection. Data are available at https://hdl.handle.net/10411/K5YJ9F. We aimed for a minimum sample size of 128 participants in Experiments 2 and 3, based on priori power analyses for power analysis for the standard F tests in a 2×2 between-subject design with power = .80, α = .05, and a moderate effect size f = .25 (Faul, Erdfelder, Lang, & Buchner, 2009). Participants in Experiment 2 (N = 139, 82 female, 12 male, 1 other, 44 non-identified; M_{age} = 23.6, Mdn_{age} = 23, range: 20–34 years) were Legal and Forensic Psychology Master's students at the Faculty of Psychology and Neuroscience (53.2%) and Master's students at the Law Faculty (46.8%). Participants received a candy bar. Participants in Experiment 3 (N = 123, 105 female, 15 male, 3 non-identified; M_{age} = 21.8, Mdn_{age} = 21, range: 19–29 years) were Legal and Forensic Psychology Master's students (42.3%) and 3rd year Psychology Bachelor students (57.7%). Bachelor students received course credit for participating. The research line was approved by the standing ethics committee of the Faculty.

Design

Experiment 1

We randomly assigned participants to one of two bias conditions (defense vs. prosecution). To induce allegiance bias, participants received one of two letters together with the case file. The defense lawyer's letter (n = 59) emphasized elements of the file that questioned the veracity of the accusation, namely that the mother's questioning of the child (Victoria) had been highly suggestive; Victoria's parents were in a divorce battle and her mother tried to influence her; the police interview with Victoria contained suggestive interviewing techniques. The prosecutor's letter (n = 84) emphasized elements of the file that supported the veracity of the accusation, namely that Victoria had given a spontaneous statement; her statement was quite detailed; she had been interviewed in a child-friendly manner in a child investigative interview studio.

Experiments 2 and 3

We added the instruction factor, resulting in a 2 (bias: defense vs. prosecution) $\times$ 2 (instruction: standard vs. two-sided) between-subject design. Participant numbers per condition varied between 33 and 36 in Experiment 2 and between 25 and 34 in Experiment 3. The bias manipulation was identical to Experiment 1.

Two-Sided Instructions Condition

In Experiment 2, participants were asked to look for elements in the case file that spoke for or against the reliability

of the statements and they could use the form to take notes. The form had two columns entitled "Favour" and "Against". In Experiment 3, we additionally specified three possible scenarios (1) Victoria describes events that she really experienced (the abuse took place); (2) Victoria is lying about the events (the abuse did not take place); (3) Victoria did not experience the events, but they were suggested by someone else (the abuse did not take place, but Victoria thinks it did).

Standard Instructions

Standard instructions in all experiments just mentioned that participants could use the form to make notes, but contained no other information.

The main dependent measures consisted of participants' answers to five (Experiment 3: six) questions rated on a 7-point Likert scale: How likely is it that the (1) child was physically abused? (2) child was sexually abused? (3) child's testimony is reliable? (4) suspect is guilty of physical abuse? (5) suspect is guilty of sexual abuse? (6) child is telling the truth? (Experiment 3 only). We created a *case assessment score* by summing items 2, 3, and 5 (+6 in Experiment 3) and a *physical abuse score* (referring to the undisputed physical abuse) by summing items 1 and 4. Across experiments, we observed good to excellent internal consistency (Cronbach's α for the case assessment score varied between .93 and .94 and between .87 and .93 for the physical abuse score).

In Experiment 2, we added seven questions indirectly related to the case, namely: I think (i) Victoria's mother is a liar; (ii) Victoria's mother is a decent person; (iii) Victoria wants attention; (iv) Victoria is spoiled, (v) Victoria's father has temper issues; (vi) Victoria's father was abused as a child; and (vii) I feel empathetic toward Victoria. These items assessed whether allegiance bias would spill over to character assessments of the individuals involved in the case. Results are reported in the Electronic Supplementary Material (ESM 1).[1]

Materials

The case file consisted of 12 pages describing a sexual abuse case against the father of 7-year old Victoria. It included a police interview with Victoria, her mother, and her father. Apart from the serious sexual abuse accusations, Victoria also mentioned that her father was very strict and sometimes hit her (i.e., physical abuse). The father confirmed hitting her in the police interview, but denied sexual abuse. For Experiment 3, we deleted the information about this undisputed physical abuse. Without information given about physical abuse, defense and prosecution-attained participants should not differ in their ratings on the physical abuse score. The case file and the appointment letters are available at https://hdl.handle.net/10411/K5YJ9F.

Procedure

We invited students attending a lecture to participate in the Experiments. In Experiment 1, they read the case file and the appointment letter on their laptop or phones, by means of interactive presentation software *Mentimeter* (https://www.mentimeter.com/). After studying the case file, participants answered the assessment questions. In Experiments 2 and 3, students received a paper version of all materials. Participants had 30 min to complete the task. The debriefing occurred during a lecture, via email or a presentation upon completion of data collection.

Results

Experiment 1: Allegiance Bias

We conducted independent Welch t tests to compare the case assessment and physical abuse scores as a function of bias (defense vs. prosecutor). As expected, defense-appointed participants ($M = 8.26$; $SD = 3.38$) scored lower than prosecutor-appointed participants ($M = 9.91$; $SD = 3.66$) on the case assessment scale, $t(129.3) = 2.78$, $p = .006$, $d = 0.47$. The same was true for the physical abuse scale, $t(132.6) = 2.05$, $p = .042$, $d = 0.35$ (defense-appointed: $M = 10.14$; $SD = 2.45$; prosecutor-appointed: $M = 11.05$; $SD = 2.78$).

Experiment 2: Counteracting Allegiance Bias With Two-Sided Instructions

We conducted 2 (bias: defense vs. prosecution) × 2 (instruction: standard vs. two-sided) ANOVAs for the case assessment and the physical abuse scores.[2] For the case

---

[1] At the end of the questionnaire, we asked 10 memory-belief questions. These were part of an unrelated research project and we report the results elsewhere (Sauerland & Otgaar, 2020).

[2] The analyses deviate from preregistered analyses in that we analyzed two composite variables rather than each item separately. We took this measure to increase power. We report the preregistered analyses in the ESM 1.
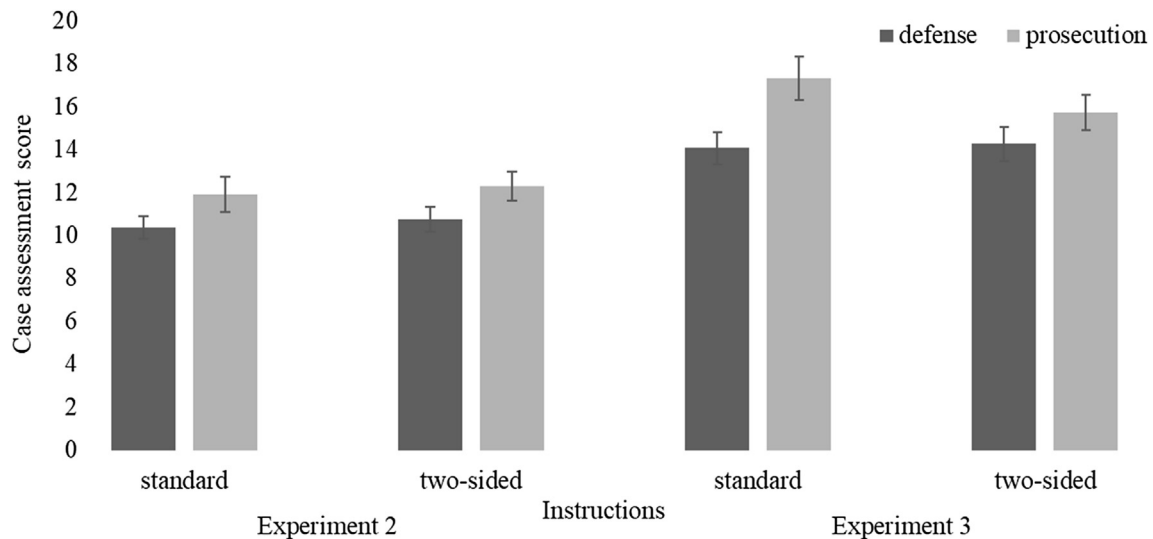
Figure 1. Case assessment scores (and standard error) in Experiment 2 and 3 as a function of bias and instructions.

Table 1. Mean case assessment and physical abuse sores as a function of bias and instruction across three experiments

| | Bias | | | | Instruction[1] | | | |
| | Defense | | Prosecution | | Standard | | Two-sided | |
	M	SD	M	SD	M	SD	M	SD
Experiment 2								
Case assessment	10.56[a]	3.30	12.10[a]	4.31	11.14	4.11	11.49	3.69
Physical abuse[2]	11.46	2.43	11.50	2.39	11.12	2.67	11.78	2.08
Experiment 3								
Case assessment	14.15[a]	4.39	16.43[a]	4.82	15.44	4.91	14.97	4.55
Physical abuse	7.05	2.92	7.24	2.76	7.19	2.74	7.08	2.95

Note. Rows sharing the same superscript letter differ significantly from another. The case assessment score can range between 3 and 21 in Experiments 1 and 2 and between 4 and 28 in Experiment 3. The physical abuse score can range between 2 and 14. [1] Instruction was not manipulated in Experiment 1. [2]See text for means of the significant Bias × Instruction interaction.

assessment score, we found lower scores in the defense than the prosecution condition, $F(1, 135) = 5.59$, $p = .019$, $d = 0.40$. The expected interaction, $F(1, 135) < 0.01$, $p = .994$, $\eta_p^2 < .01$, or main effect of instruction did not emerge, $F(1, 135) = 0.34$, $p = .559$, $d = 0.09$. Figure 1 and Table 1 displays the mean case assessment scores as a function of bias and instruction for Experiments 2 and 3.

For the physical abuse score, we found the predicted Bias × Instruction interaction, $F(1, 135) = 7.83$, $p = .006$, $\eta_p^2 = .055$. Unexpectedly, when using two-sided instructions, participants' ratings were higher in the prosecution ($M = 12.39$, $SD = 1.87$) than in the defense condition ($M = 11.22$, $SD = 2.13$), $F(1, 67) = 5.86$, $p = .018$, $d = 0.58$. The simple effect for standard instructions was statistically nonsignificant, $F(1, 68) = 2.81$, $p = .098$, $d = 0.40$, as were the main effects of bias, $F(1, 135) = 0.02$, $p = .886$, $d = 0.02$, and instructions,

$F(1, 135) = 2.44$, $p = .120$, $d = 0.24$. Figure 2 displays the mean physical abuse scores as a function of bias and instruction for Experiments 2 and 3.

Experiment 3: Counteracting Allegiance Bias With Two-Sided Instructions With Specified Scenarios

For the case assessment score, prosecution-appointed participants scored higher than defense-appointed participants, $F(1, 118) = 7.93$, $p = .006$, $d = 0.49$, as expected. The main effect of instruction, $F(1, 118) = 0.72$, $p = .399$, $d = 0.10$, and the interaction were statistically non-significant, $F(1, 118) = 1.16$, $p = .285$, $\eta_p^2 = .01$. As expected, for the physical abuse score, no significant effects emerged, all $ps \geq .697$, $\eta_p^2 s \leq .01$.
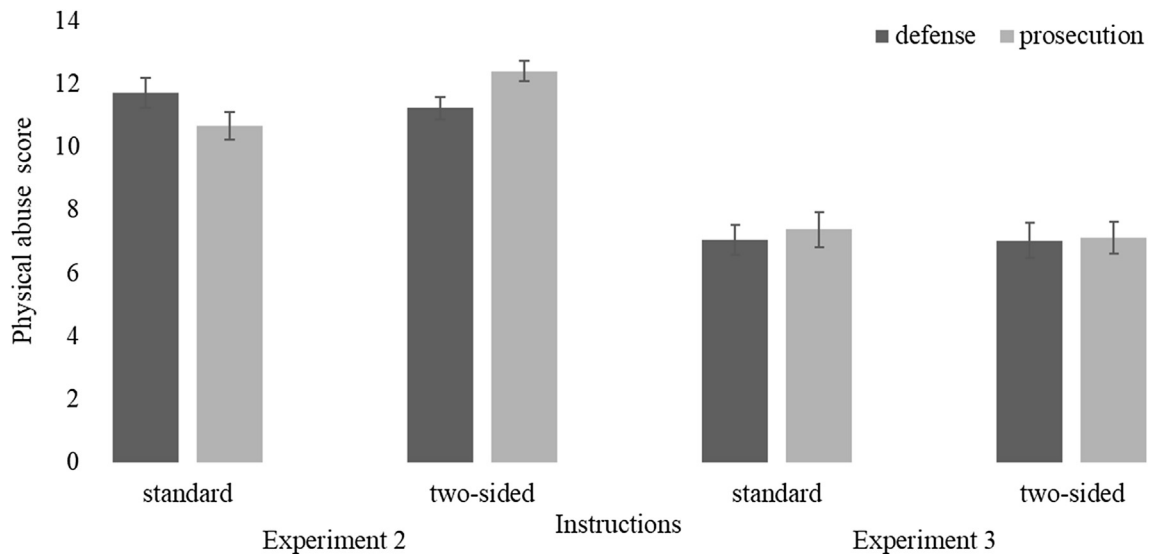
Figure 2. Physical abuse scores (and standard error) in Experiment 2 and 3 as a function of bias and instructions.

Discussion

In three experiments, psychology and law students received identical case files together with an appointment letter that either emphasized elements of the file that questioned (defense) or supported (prosecution) the veracity of a sexual abuse accusation. Following their reading of the file, participants evaluated the case using several case evaluation questions. Providing evidence of allegiance bias, participants' case evaluation adhered with the party that retained them (Hypothesis 1). This shows that prosecution-attained participants interpreted the same information as stronger evidence of guilt than defense-attained participants. Contrary to Hypothesis 2, neither Experiments 2 nor 3 supported the idea that two-sided instructions can eliminate allegiance bias, despite the use of preconstructed alternative scenarios in Experiment 3 (Rassin, 2018; Van Koppen & Mackor, 2019).

The current findings, along with one earlier experiment (McAuliff & Arter, 2016), provide first evidence of allegiance bias for the evaluation of interviews in a child sexual abuse case. This is in line with the accumulating research base demonstrating that experts can be biased by the side that retained them (e.g., Murrie et al., 2013). The findings underscore the importance of legal safeguards such as cross-examination and opposing expert testimony for minimizing the potential effects of allegiance bias (McAuliff & Arter, 2016).

Our two-sided instructions were not successful in counteracting allegiance bias. This is contrary to earlier findings for similar instructions (Lord et al., 1984; O'Brien, 2009, cf. Arkes, 1991). Backfiring might be one possible explanation. This counter-intended effect of debiasing techniques can occur when thinking about alternatives is perceived as effortful, thereby strengthening the original point of view as the only viable solution (O'Brien, 2009; Sanna, Schwarz, & Stocker, 2002).

Predetermined scenarios did not eliminate allegiance bias in Experiment 3. It is possible that our instructions did not (sufficiently) stimulate a *comparison* of these scenarios (Van Koppen & Mackor, 2019). A better approach might be to ask which of the scenarios *best* explains the evidence rather than simply searching for evidence for and against each scenario. Our findings emphasize that using alternative scenarios effectively is a complex task that goes beyond looking for support for different scenarios.

This study is not without limitations. Participants were tested during lectures. It is possible that testing participants in smaller groups would lead to stronger allegiance bias. On the other hand, the current findings demonstrate that allegiance bias occurs despite a lack of personal contact between the retaining party and the "expert," and thus weaker pull to retaining party than in previous studies (Murrie et al., 2013). This suggests that the current research setup might be a cost-effective way for studies on interventions for counteracting allegiance bias. Once promising interventions have been identified, such studies could be followed up by studies with higher ecological validity and expert witness samples.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/2151-2604/a000416

ESM 1. Item-by-item analyses for Experiment 1, pre-registered item-by-item analyses for Experiment 2, and pre-registered analyses on three composite variables for Experiment 3

References

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*, 486–498. https://doi.org/10.1037/0033-2909.110.3.486

Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling, 2*, 43–63. https://doi.org/10.1002/jip.19

Ask, K., & Granhag, P. A. (2007). Motivational bias in criminal investigators' judgments of witness reliability. *Journal of Applied Social Psychology, 37*, 561–591. https://doi.org/10.1111/j.1559-1816.2007.00175.x

Beckham, J. C., Annis, L. V., & Gustafson, D. J. (1989). Decision making and examiner bias in forensic expert recommendations for not guilty by reason of insanity. *Law & Human Behavior, 13*, 79–87. https://doi.org/10.1007/BF01056164

Edens, J. F., Cox, J., Smith, S. T., DeMatteo, D., & Sörman, K. (2015). How reliable are Psychopathy Checklist – Revised scores in Canadian criminal trials? A case law review. *Psychological Assessment, 27*, 447–456. https://doi.org/10.1037/pas0000048

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science, 3*, 143–152. https://doi.org/10.1037/tps0000114

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality & Social Psychology, 47*, 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231s

Lloyd, C. D., Clark, H. J., & Forth, A. E. (2010). Psychopathy, expert testimony, and indeterminate sentences: Exploring the relationship between Psychopathy Checklist-Revised testimony and trial outcome in Canada. *Legal and Criminological Psychology, 15*, 323–339. https://doi.org/10.1348/135532509X468432

Marksteiner, T., Ask, K., Reinhard, M. A., & Granhag, P. A. (2011). Asymmetrical scepticism towards criminal evidence: The role of goal-and belief-consistency. *Applied Cognitive Psychology, 25*, 541–547. https://doi.org/10.1002/acp.1719

McAuliff, B. D., & Arter, J. L. (2016). Adversarial allegiance: The devil is in the evidence details, not just on the witness stand. *Law and Human Behavior, 40*, 524–535. https://doi.org/10.1037/lhb0000198

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, 24*, 1889–1897. https://doi.org/10.1177/0956797613481812

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater(dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, & Law, 15*, 19–53. https://doi.org/10.1037/a0014897

O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology, Public Policy, and Law, 15*, 315–334. https://doi.org/10.1037/a0017881

Otto, R. K. (1989). Bias and expert testimony of mental health professionals in adversarial proceedings: A preliminary investigation. *Behavioral Sciences & the Law, 7*, 267–273. https://doi.org/10.1002/bsl.2370070210

Popper, K. R. (1959). *The logic of scientific discovery.* Abingdon-on-Thames, UK: Routledge.

Rassin, E. (2018). Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *Journal of Investigative Psychology and Offender Profiling, 15*, 227–233. https://doi.org/10.1002/jip.1504

Reese, E. J. (2012). Techniques for mitigating cognitive biases in fingerprint identification. *UCLA Law Review, 59*, 1252–1290.

Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 497–502. https://doi.org/10.1037/0278-7393.28.3.497

Sauerland, M., & Otgaar, H. (2020). *Teaching psychology students to change (or correct) controversial beliefs about memory works.* Manuscript in preparation.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 544–551. https://doi.org/10.1037/0096-1523.3.4.544

Van Koppen, P. J., & Mackor, A. R. (2019). A scenario approach to the Simonshaven case. *Topics in Cognitive Science,*. Advance online publication. https://doi.org/10.1111/tops.12429

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140. https://doi.org/10.1080/17470216008416717

Open Data
The materials and data of all three experiments are available here: https://hdl.handle.net/10411/K5YJ9F

Melanie Sauerland
Section Forensic Psychology
Faculty of Psychology and Neuroscience
Maastricht University
PO Box 616
6200 MD Maastricht
The Netherlands
melanie.sauerland@maastrichtuniversity.nl