Routledge
Taylor & Francis Group

'Anyone who commits such a cruel crime, must be criminally irresponsible': context effects in forensic psychological assessment

Eric Rassin

Erasmus School of Social and Behavioural Sciences, Rotterdam, The Netherlands

In recent years, it has become clear that expert opinion can be biased. It has been argued that forensic psychologists may also be susceptible to bias. In the present study, the vulnerability of forensic psychological evaluation of the suspect's mental health to the context effect (i.e. the influencing of the expert opinion by irrelevant information) was tested. Master students in forensic psychology were asked to interpret test scores of a suspect in a fictitious double murder case. Some participants received a version of the case in which the description of the murders was neutral. Others received a more explicit version. Whereas the explicitness should not affect the forensic psychological evaluation, it was found that participants in the latter condition seemed more concerned about the suspect's mental health than those in the former. It is concluded that training programmes in forensic psychological assessment should devote attention to bias.

Keywords: bias; context effect; criminal responsibility; decision making; diagnostics; explicitness of evidence; fallacies; forensic psychology; mental health; undue influence.

Introduction

While criminal procedures differ between countries and legal systems, by and large, in any case, a crucial question for the judge (or jury) is to determine whether or not the suspect has committed the crime of which he is suspected. In some instances, for example when there is conflicting evidence, this can be a difficult task. Many legal systems provide the judge with the opportunity to seek assistance of all sorts of expert witnesses. Intuitively, one may think that these experts will present scientifically-scrutinised, and thus objective, insights. However, expert witness reports have been shown to be susceptible to mistakes and bias (Saks & Koehler, 2005).

Indeed, there is much literature on bias in expert witnesses. Obviously, particularly in adversarial systems, experts may be exposed to allegiance bias – that is, the tendency to produce reports that are favourable to the party that retained them (Murrie et al., 2013). But beyond that, experts may suffer from forensic confirmation bias, which can be defined quite broadly as 'the class of effects through which an individual's preexisting beliefs, expectations, motives, and situational context influence the collection, perception, and interpretation of evidence during the course of a criminal case' (Kassin et al., 2013, p. 45). In essence, this definition encompasses all reasons to come to a flawed (false positive)

conclusion, including prior beliefs of the expert, resistance to changing an initial impression, the tendency to *see* evidence of a (fingerprint) match, merely because one is asked to test whether two samples match, and influences of contextual information such as prior decisions in the case.

The last component of the forensic confirmation bias is referred to as the context effect – that is, an undue influence of information that should be considered irrelevant for the decision at hand. For example, Dror et al. (2021) found, in a sample of 133 forensic pathologists, that the gender and age of a caretaker of a child found dead with a skull fracture determine the extent to which they considered this death suspicious. Particularly, if the caretaker was an African-American male boyfriend of the child's mother, pathologists were five times more likely to label the child's death as a homicide (as opposed to an accident) than when it was the Caucasian grandmother of the child.

Dror et al. (2005) provided another interesting example of the context effect. They asked 27 undergraduates to perform match decisions on 96 pairs of fingerprints. Part of the pairs were obvious matches or nonmatches whereas others were ambiguous. Some pairs were presented blindly – that is, without any context information. Others were accompanied by non-emotional information (e.g. a picture of a neutral part of the crime scene); a third group of pairs was presented together with emotional context information (e.g. a bloody picture of the victim), and the last class of pairs was accompanied by emotional context and subliminal messages such as *same* or *guilty*. Whereas the decisions regarding the unambiguous pairs were not affected by context, those regarding the ambiguous pairs were indeed influenced: matches were concluded in 46% of the pairs without context, in 49% of the non-emotional context pairs, in 58% of the emotional context pairs, and in 66% of the emotional

context plus subliminal messages pairs ($p <$ .001). Hence, these findings suggest that context information can indeed increase perception of matches in fingerprint analyses. The same research group also found evidence to suggest that factual matches can be made to be missed, by providing negative context information. In this particular study, a small group of professional fingerprint experts were included (Dror et al., 2006).

Admittedly, while experts may be expected to deliver high-quality contributions to legal decision making, in fairness, experts are just as susceptible to bias as any other individuals. Indeed, by now, context effects and other biases have been documented in the work of experts from all sorts of domains, including DNA analysis, fingerprint analysis (dactyloscopy), bone analysis (anthropology), bite marks (odontology), bloodstain analysis, handwriting and voice analysis, toolmark and bullet analysis (ballistics), and pathology (see for a review, Cooper & Meterko, 2019). Ironically, experts themselves may have a blind spot for their own susceptibility to bias. Kukucka et al. (2017) asked a mixed sample of 403 professional experts from the field of DNA, fingerprint, handwriting, toxicology and ballistics whether they believed that bias is a cause for concern in forensic science as a whole. This question was answered affirmatively by 71% of the participants. When asked whether bias is a cause for concern in their own specific domain, 52% replied confirmatively. But when asked whether their own judgments are influenced by bias, only 26% confirmed the question.

There is reason to argue that forensic psychologists are also susceptible to bias, even though they generally operate not within the domain of fact finding, but in the domain of determining the appropriate sentence. That is, forensic psychological analyses concern, among other topics, risk assessment, fitness to stand trial and treatment evaluation. Forensic psychologists also seek to assist the judge in

answering the question whether the suspect is criminally responsible, and thus fit for imprisonment, or criminally irresponsible, and thus in need of treatment in a forensic psychiatric clinic. In this role, forensic psychologists assess the suspect's mental health. Neal and Grisso argued that the topic of 'the effects of subtle but powerful biases in forensic mental health assessment is ripe for discussion, as research evidence that challenges our objectivity and credibility garners increased attention both within and outside of psychology' (2014, p. 200).

Indeed, Chevalier et al. (2015) found that forensic psychologists display an allegiance bias, in that, when performing risk assessment, they tend to report their conclusion in a way that benefits either the defence or the prosecution, depending on which party retained them. In a qualitative study in American forensic psychologists, Neal and Brodsky (2016) found that practitioners were somewhat aware of the potential of bias in their work, but also displayed a bias blind spot in that they considered themselves less vulnerable to bias than their peers. Further, the respondents had some good ideas on how to reduce bias (e.g. by keeping up with scientific literature), but also conjured up some less validated strategies such as introspection.

Zapf et al. (2018) found a bias blind spot, in their sample of 1099 forensic mental health practitioners. When asked whether bias is a problem for forensic mental health assessment in general, 79% answered affirmatively, but when asked whether they themselves were susceptible to bias, only 52% endorsed the question. In addition, most of the respondents (i.e. 88%) erroneously believed that they could set aside the effects of bias by mere willpower. Likewise, Zappala et al. (2018) found a bias blind spot, in that respondents considered themselves to be less vulnerable to various biases than their peers, in a sample of 80 forensic mental health professionals.

Hence, while allegiance bias and bias blind spot have been observed in forensic psychology, it is unknown to what extent forensic psychologists are susceptible to contextual information such as the explicitness of the evidence and the appearance of the suspect. The goal of the present study was to test whether context effects occur in the domain of forensic psychology, particularly forensic mental health assessment. Participants were given an assignment to judge the mental state of a suspect. In the assignment, various context effects were included (i.e. information that is irrelevant to the evaluation of the mental state of the suspect). Based on previous research suggesting that context effects occur in the decision making of experts in various fields (e.g. Cooper & Meterko, 2019; Dror et al., 2006, 2005), as well as in professional judges (e.g. Rassin, 2020), it was hypothesised that forensic psychologists too are sensitive to contextual information. That is, it was expected that participants who were exposed to more explicit evidence of the crime would be more concerned about the suspect's mental health than their colleagues who were given a version of the case that included less explicit evidence.

Method

Participants

Sixty master students in forensic psychology (52 women, 87%) participated in this study. The mean age in this sample was 22.87 years ($SD = 1.65$). Students participated in the course of an exercise of which the goal was to interpret the scores on various forensically relevant tests. They were given extra course credits upon completion of the assignment. Data were collected in accordance with national legislation. The sample was split into two groups. These groups did not differ with respect to gender, $\chi^2(1) = 0.58$, $p = .448$;

$BF_{10} = 0.57$ (BF = Bayes factor), or age, $t(58) = 0.31$, $p = .757$; $BF_{10} = 0.27$.

Measures and procedure

Participants were given the following short fictitious case description. 'A man, age 32, is suspect of a double murder. Two victims (female aged 24 and a male aged 25) were found by the police, in the bushes, nearby the suspect's home (i.e. < 10 km). It is believed that the suspect attacked the two (they were a couple), hit them unconscious, and took them somewhere where he finally killed them. The motive for the attack is yet unclear, because the suspect refuses to make a statement, at this point. There is ample evidence that the suspect is indeed the perpetrator (for example, a witness identified the suspect as the person who he had seen carrying two large plastic bags and placing them into a car; this witness also identified the suspect's car; technical evidence proves that the suspect's car has recently been in immediate proximity of the location where the victims were found). While the suspect denies, he has no alibi, and does have a criminal record of violent crimes. It should be noted that when the police found the bodies, they were both stripped and mutilated with a knife. Their clothes have never been found. A forensic psychological evaluation of the suspect is needed. Ideally, four questions are addressed in a forensic psychological evaluation: (1) a psychiatric diagnosis, (2) a personality profile, (3) a recidivism risk analysis, and (4) an advise on criminal responsibility. The following tests have been administered to the suspect: The Magical Ideation Scale (MIS; Eckblad & Chapman, 1983), the Aggression Questionnaire (AQ; Buss & Perry, 1992), the Psychopathy Checklist (PCL; Hare, 1980), and the Short Dark Triad (SD3; Jones & Paulhus, 2014). Below, you will find the suspect's scores on these scales. Interpret these scores and advise the commissioning judge on the four questions'.

The suspect's scores were (borderline) high (i.e. 18 on the MIS, 95 on the AQ, 34 on the PCL, 3.33 on the SD3-Psychopathy, 3 on the SD3-Machiavellianism and 3.22 on the SD3-Narcissism). Thus, the scoring profile left room for interpretation and variance. Participants were given the pertinent articles to enable them to interpret these scores. These four scales are widely used in forensic mental health assessment in the Netherlands. The MIS measures proneness to delusional thinking. The Aggression Questionnaire taps various aggression-related phenomena including anger and physical and verbal aggression. The PCL measures psychopathic traits. Finally, the SD3 also taps psychopathy in addition to Machiavellianism and narcissism. Participants had been made familiar with these measures prior to this study.

Unknown to participants, there were two versions of the assignment. In one version, a (computer generated) photo of the suspect was included that was adopted from Todorov et al. (2009), as a face that tends to elicit feelings of trustworthiness (see also Wilson & Rule, 2015). This version is referred to as the mild version.

In the second version, the photo of the suspect was a version adopted from Todorov et al. (2009), as a face that tends to elicit feelings of untrustworthiness. Further, instead of the phrase that the victims were 'mutilated', it was mentioned that they were 'slashed up with a knife' (see Edwards & Bryan, 1997). Finally, two pictures of mutilated bodies (adopted from the International Affective Picture System (IAPS; Lang et al., 1997; see also Dror et al., 2005) were included. This version is referred to as the aggravated version.

Rationally, the mild and aggravated versions should elicit similar conclusions as to the forensic psychologically relevant topics, because the suspect's appearance and the gruesomeness of the descriptions have no bearing on the psychological evaluation. However, based on the described context effect, it was hypothesised that the aggravated version would make the participants more concerned about the suspect's mental health.

Table 1.   Mean scores on the forensic mental health assessment variables.

	Mild version	Aggravated version	p	BF_{10}
Psychiatric malfunction	7.32 (1.72)	7.77 (1.50)	.286	0.43
Psychological malfunction	6.37 (2.24)	7.40 (1.48)	.040	1.65
Recidivism risk	7.70 (1.62)	8.50 (1.01)	.026	2.28
Criminal irresponsibility	5.77 (1.73)	6.66 (2.14)	.083	0.96
Overall assessment of mental health	27.16 (5.46)	30.33 (3.82)	.012	4.19

Note: Standard deviations in parentheses. BF = Bayes factor.

After reading the case vignette, participants answered four questions by circling a number between zero and 10: (a) does, in your opinion, given the test results and all other information, the suspect suffer from relevant psychiatric complaints? ($0 = certainly\ not$; $10 = certainly$); (b) does the suspect, judging from the psychological tests, have personality characteristics relevant to the question of criminal responsibility? ($0 = certainly\ not$; $10 = certainly$); (c) how do you estimate the likelihood of recidivism assuming that the suspect is guilty, if the suspect were left untreated? ($0 = extremely\ low$; $10 = extremely\ high$); and (d) what is your advice on the suspect's criminal responsibility? The last question was answered on a 3-point scale (*completely responsible*; *partly irresponsible*; *completely irresponsible*) that was quantitatively transformed into the values 3.33, 6.67 and 10 points, with higher scores indicating that the suspect was judged to be irresponsible. This transformation was performed to make this question weigh the same as the other three. The overall mental health assessment thus consisted of various topics (see also Mercado et al., 2006). Ultimately, the four questions were conjoined into one composite variable, called the overall assessment of mental health (range= 0–40), with higher scores indicating that the suspect was mentally unwell. It should be noted that, in the Netherlands, forensic

psychiatrists and psychologists do not have clear definitions or cut-off scores for the determination of criminal responsibility. This implies that the ultimate advice on the suspect's accountability needs to be determined in a somewhat subjective manner.

Results

The mean scores on the four questions and on the composite variable (overall assessment) as a function of context information are presented in Table 1.

The data were analysed with JASP (free Bayesian software available at www.jasp-stats.org). JASP allows for both inferential null hypothesis significance testing (in this case, independent t tests) and Bayesian analysis. Both are reported below. Crucially, the latter analysis yields a Bayes factor, which represents the likelihood ratio for the fit of the data in the null and in the alternative hypothesis. Values of BF_{10} smaller than 1 indicate that the data fit better in the null hypothesis than in the alternative hypothesis. Values of BF_{10} larger than 1 suggest that the alternative hypothesis predicts the data better. Values of BF_{10} larger than 3 can be interpreted as positive/substantial support for the alternative hypothesis. Values of BF_{10} larger than 10 represent positive/strong support, and values of BF_{10} larger than 20 provide strong support for the alternative

hypothesis (Jarosz & Wiley, 2014). In the current analyses, the prior odds were left undefined and thus set at 1.0.

As can be seen in Table 1, the two groups differed on some of the forensic psychological evaluations. Two of the four answers were significantly different, one tended to be, and the psychiatric diagnosis was not affected. Consequently, the overall assessment of mental health was clearly different between the mild and aggravated groups.

Discussion

The current study set out to explore whether forensic psychological evaluations of the suspect's mental health are susceptible to context effects. The findings preliminarily suggest that this hypothesis can be confirmed. Employing inferential null hypothesis significance testing, two out of four ratings differed between the mild and aggravated versions of the same case vignette. Consequentially, the overall evaluation of the suspect's mental health (i.e. the composite variable bringing all four ratings together) was significantly affected by the contextual information. The Bayes factor of 4.19 confirmed this finding, yielding positive/substantial support for the hypothesised effect. Crucially, the psychiatric evaluation was not affected by context information. This may be so, because our participants were forensic psychology students, not forensic psychiatry students. Indeed, in our experience, (forensic) psychiatrists do not primarily rely on test scores, but have other instruments to reach their diagnoses (e.g. diagnostic interviews). Meanwhile, we cannot conclude whether forensic psychiatric evaluations (when produced by students in forensic psychiatry) are free from bias.

Obviously, the present study has some limitations that deserve attention. First, a comparison was made between a mild case vignette (including a trustworthy facial picture and neutral descriptions) and an aggravated version (in which the suspect looked less trustworthy, the

description of the murder was more explicit, and gruesome pictures of the victims were included). Hence, multiple pieces of contextual information were included. This makes it impossible to determine which information (or combination thereof) caused the observed effect. However, that determination was not the goal of this study. Rather, the goal was to explore whether context effects occur at all. Further, whereas explicit pictures were included in the aggravated version, no complementary (e.g. neutral) pictures were included in the mild version. In future research, the mild condition should include non-explicit crime scene pictures, so as to control for the presence of pictures. Nonetheless, even in absence of pictures in the mild version, it remains striking that group differences in the evaluation of the suspect's mental health were observed. We reiterate that the presence of pictures should not at all affect the forensic psychologist's evaluations. It should also be noted that we did not include manipulation checks, and, hence, it is not completely certain how well participants paid attention to all elements in the stimulus materials. Another limitation is that our participants were master students in forensic psychology. Thus, they were not fully trained forensic psychologists yet. Hence, it remains to be seen to what extent fully trained forensic psychologists are susceptible to context effects.

Notwithstanding the limitations, the findings at least suggest that students in forensic psychology are susceptible to context effects. Evidently, this is an alarming finding. Moreover, there is theoretical reason to argue that in this domain, more biases are luring. For example, a combination of outcome bias and confirmation bias may lead to inflated concern about the suspect's mental health. Particularly, at the time of the assessment, the forensic psychologist is knowledgeable about the (severe) crime(s) that the suspect committed. That knowledge alone may cause the psychologist to over-diagnose mental illness. Note that committing crime is in itself a criterion of

antisocial personality disorder (American Psychiatric Association, 2013). In addition, taking as starting point that the suspect indeed committed the crime may also inflate perception of mental health problems, if the suspect subsequently denies. Koenraadt et al. (2007) argue: 'In the case of a defendant who denies the charges . . . the psychologist . . . will therefore need to stick to the facts as described in the casefile' (p. 117). Unintendedly, this 'presumption of guilt' may result in the denying suspect scoring artificially high on antisocial personality disorder and avoidant personality disorder, merely because denying in the face of this presumption may make the psychologist conclude that the suspect is evading responsibility for the crime he (presumably) committed.

Finally, forensic psychologists who have had training in clinical psychology may have acquired specific clinical attitude that is at odds with forensic demeanour. For example, clinical psychologists tend to be a little critical about the client's narrative, because being critical would harm the empathic therapist–client relationship. Hence, if a client testifies that (s)he has experienced a traumatic event, the clinical psychologist should not start a police investigation to find out whether this event actually happened. Obviously, this empathic attitude is not (always) dictated, and is sometimes even counterproductive in forensic psychology (Greenberg & Shuman, 1997; Rassin & Merckelbach, 1999). Meanwhile, there is some evidence to argue that forensic psychological evaluations of mental health have little inter-rater reliability (i.e. 55%; Gowensmith et al., 2013).

If, in the field of forensic psychology, different sources of bias may occur simultaneously, this may result in what Dror et al. (2017) call a *bias cascade*, referring to the accumulation of error caused by the sequential steps in criminal or forensic investigation and proceeding depending on the outcome of previous ones, or even *bias snowball*, which refers to an actual increase in bias in subsequential steps (see also Dror, 2018). Indeed, Dror (2020) distinguishes eight potential sources of bias, including context effect, but also the to-be-evaluated information itself (e.g. a blood sample clearly suggests that blood was spilled), perceived base rates and heuristics such as positive test strategies. Further, he discusses six fallacies that hinder the reduction of bias, such as the belief that only other people suffer from bias (i.e. the bias blind spot), or the illusion of control (i.e. thinking that one is capable of mentally countering bias).

The present data invite the question of how students in forensic psychology (and practising forensic psychologists) can find protection against bias. Theoretically, the use of well-validated tests should offer some protection against bias (Lockhart & Satya-Murti, 2017). Admittedly though, not all tests used in forensic practice have established high psychometric quality. As to the most often used PCL(–R), Rosenberg Larsen et al. (2020) recently found little or no evidence for many of the currently generally assumed core features and associates of psychopathy as measured with the PCL(–R). These authors found 'no consistent, well-replicated evidence of observable deficits in conscience, remorse, empathy or moral judgments' (Rosenberg Larsen et al., 2020, p. 305). Neither did they find consistent evidence for the idea that psychopaths are difficult to treat, nor that the PCL(–R) is a good measure of risk assessment (for which it was not designed in the first place). Fazel et al. (2012) argued that the most often used risk assessment tools have modest validity. Even if tests are well validated, clear cut-off scores are needed to avoid ambiguity. Note that the suspect's scores on the tests in the fictitious case vignette employed in the current study were purposely set at borderline high levels, to create ambiguity.

Another remedy against bias, particularly the context effect, is blinding (Lockhart & Satya-Murti, 2017). However, blinding is difficult for forensic psychologists engaging in

mental health and risk assessment, because, for example, the PCL(–R) and many risk assessment tools require that the evaluators make themselves familiar with the complete casefile, rendering them explicitly vulnerable to contextual information. However, even though blinding is difficult in some forensic psychological evaluations, Dror et al. (2015) discussed linear sequential unmasking (LSU) as a fruitful alternative way of avoiding bias. Particularly, if experts have to be exposed to case information that may cause bias, this information should be kept away from them as long as possible, and only be presented sequentially. Experts should initially only be exposed to minimal case information, sufficient to carry out their analyses. Given its potential protective effect against context effects, It is important to explore whether linear sequential unmasking is applicable in forensic psychological practice.

In sum, future research is needed to find out how vulnerable forensic psychological evaluations are, and how they can be protected against bias, ultimately increasing overall quality. Meanwhile, there is reason to argue that training programmes in forensic psychology should address bias, in order to make future practitioners aware of the risks and of the necessity to counteract possible sources of bias.

Ethical standards

Declaration of conflicts of interest

Eric Rassin has declared no conflicts of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee [Erasmus School of Social and Behavioural Sciences] and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Data availability statement

Data can be obtained from the author.

References

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders*. 5th edi. APA.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. https://doi.org/10.1037/0022-3514.63.3.452

Chevalier, C. S., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior*, *39*(3), 209–218. https://doi.org/10.1037/lhb0000114

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, *297*(1), 35–46. https://doi.org/10.1016/j.forsciint.2019.01.016

Dror, I. E. (2018). Bias in forensic experts. *Science*, *360*(6386), 243. https://doi.org/10.1126/science.aat8443

Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Analytical Chemistry*, *92*, 7998–8004. https://doi.org/10.1021/acs.analchem.0c00704

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74–78. https://doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., Melinek, J., Arden, J. L., Kukucka, J., Hawkins, S., Carter, J., & Atherton, D. S. (2021). Cognitive bias in forensic pathology decisions. *Journal of Forensic Sciences*, 1–7. https://doi.org/10.1111/1556-4029.14697

Dror, I. E., Morgan, R., Rando, C., & Nakhaeizadeh, S. (2017). The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making. *Journal of Forensic Sciences*, *62*(3), 832–833. https://doi.org/10.1111/1556-4029.13496

Dror, I. E., Péron, A. E., Hind, S. L., & Charlton, D. (2005). When emotions get the better of

us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, *19*(6), 799–809. https://doi.org/10.1002/acp.1130

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, *60*(4), 1111–1112. https://doi.org/10.1111/1556-4029.12805

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal of Consulting and Clinical Psychology*, *51*(2), 215–225. https://doi.org/10.1037/0022-006X.51.2.215

Edwards, K., & Bryan, T. S. (1997). Judgmental biases produced by the instructions to disregard: The (paradoxical) case of emotional information. *Personality and Social Psychology Bulletin*, *23*(8), 849–864. https://doi.org/10.1177/0146167297238006

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24827 people: Systematic review and meta-analysis. BMJ, *345*, e4692–e4692. https://doi.org/10.1136/bmj.e4692

Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2013). How reliable are forensic evaluations of legal sanity? *Law and Human Behavior*, *37*(2), 98–106. https://doi.org/10.1037/lhb0000001

Greenberg, S. A., & Shuman, D. W. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice*, *28*(1), 50–57. https://doi.org/10.1037/0735-7028.28.1.50

Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, *1*(2), 111–119. https://doi.org/10.1016/0191-8869(80)90028-8

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, *7*(1), 2–9. https://docs.lib.purdue.edu/jps/vol7/iss1/2

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief measure of dark personality traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions.

Journal of Applied Research in Memory and Cognition, *2*(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Koenraadt, F., Mooij, A., & van Mulbregt, J. (2007). *The mental condition in criminal law: Forensic psychiatric and psychological assessment in a residential setting*. Dutch University Press.

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: A global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition*, *6*(4), 452–459. https://doi.org/10.1016/j.jarmac.2017.09.001

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). *International affective picture system: Technical manual and affective ratings*. NIMH Center for the Study of Emotion and Attention.

Lockhart, J. J., & Satya-Murti, S. (2017). Diagnosing crime and diagnosing disease: Bias reduction strategies in the forensic and clinical sciences. *Journal of Forensic Sciences*, *62*(6), 1534–1541. https://doi.org/10.1111/1556-4029.13453

Mercado, C. C., Bornstein, B. H., & Schopp, R. F. (2006). Decision-making about volitional impairment in sexually violent predators. *Law and Human Behavior*, *30*(5), 587–602. https://doi.org/10.1007/s10979-006-9055-7

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*(10), 1889–1897. https://doi.org/10.1177/0956797613481812

Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *22*(1), 58–76. https://doi.org/10.1037/law0000077

Neal, T. M. S., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, *20*(2), 200–211. https://doi.org/10.1037/a0035824

Rassin, E. (2020). Context effect and confirmation bias in criminal fact finding. *Legal and Criminological Psychology*, *25*(2), 80–89. https://doi.org/10.1111/lcrp.12172

Rassin, E., & Merckelbach, H. (1999). The potential conflict between clinical and judicial decision making heuristics. *Behavioral Sciences & the Law*, *17*(2), 237–248. https://doi.org/10.1002/(sici)1099-0798(199904/06)17:2<237::aid-bsl342>3.0.co;2-5

Rosenberg Larsen, R., Jalava, J., & Griffiths, S. (2020). Are Psychopathy Checklist (PCL) psychopaths dangerous, untreatable, and without conscience? A systematic review of the empirical evidence. *Psychology, Public Policy, and Law*, *26*(3), 297–311. https://doi.org/10.1037/law0000239

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*(5736), 892–895. https://doi.org/10.1126/science.1111565

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcome. *Psychological Science*, *26*(8), 1325–1331. https://doi.org/10.1177/0956797615590992

Zapf, P. A., Kukucka, J., Kassin, S. M., & Dror, I. E. (2018). Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. *Psychology, Public Policy, and Law*, *24*(1), 1–10. https://doi.org/10.1037/law0000153

Zappala, M., Reed, A. L., Beltrani, A., Zapf, P. A., & Otto, R. K. (2018). Anything you can do, I can do better: Bias awareness in forensic evaluators. *Journal of Forensic Psychology Research and Practice*, *18*(1), 45–56. https://doi.org/10.1080/24732850.2017.1413532