

Risk in the eye of the beholder

Citation for published version (APA):

Kamorowski, J. (2021). *Risk in the eye of the beholder: Cognitive bias and risk assessment instruments*. [Doctoral Thesis, Maastricht University, University of Gothenburg]. Maastricht University.
<https://doi.org/10.26481/dis.20210614jk>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20210614jk](https://doi.org/10.26481/dis.20210614jk)

Document Version:

Publisher's PDF, also known as Version of record

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](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

RISK IN THE EYE OF THE BEHOLDER:

COGNITIVE BIAS AND RISK ASSESSMENT INSTRUMENTS



JENNIFER KAMOROWSKI

**RISK IN THE EYE OF THE BEHOLDER:
COGNITIVE BIAS AND RISK ASSESSMENT
INSTRUMENTS**

Jennifer Kamorowski



ISBN: 978-94-6416-604-0

Cover photo: Anika Huizinga

Cover & lay-out: Publiss | www.publiss.nl

Print: Ridderprint | www.ridderprint.nl

© Copyright 2021: Jennifer Kamorowski, The Netherlands.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording, or otherwise, without the prior written permission of the author.

Risk in the Eye of the Beholder:

Cognitive Bias and Risk Assessment

Instruments

DISSERTATION

To obtain the degree of Doctor of Philosophy by the University of Gothenburg,
on the authority of the Deputy Vice-Chancellor Dr. Mette Sandoff
in accordance with the decision of the Faculty Board of Social Sciences
and the degree of Doctor by Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans, to be defended in public on,
Monday, June 14, 2021, at 1000 hours in Maastricht

by

Jennifer Kamorowski

Supervisors:

Prof. dr. C. de Ruiter, Maastricht University, the Netherlands

Prof. dr. M. Jelicic, Maastricht University, the Netherlands

Dr. M. Schreuder, Maastricht University, the Netherlands

Co-supervisor:

Prof. dr. K. Ask, University of Gothenburg

Assessment Committee:

Prof. dr. H. Otgaar [Committee Chair], Maastricht University, the Netherlands

Prof. dr. P. van Koppen, Maastricht University, the Netherlands

Prof. dr. I. Dror, University College London, United Kingdom

Dr. V. de Vogel, Hogeschool Utrecht, the Netherlands

Dr. S. Landström, University of Gothenburg, Sweden

This research is supported by a fellowship awarded from the Erasmus Mundus Joint Doctorate Program in the House of Legal Psychology (EMJD-LP) with Framework Partnership Agreement (FPA) 2013-0036 and Specific Grant Agreement (SGA) SGA 532473-EM-5-2017-NL-ERAMUNDUS-EPJD to Jennifer Kamorowski.

Dedication

This dissertation is dedicated to the memory of Dr. Vincent Alfred Memoli

September 25, 1950 – July 12, 2020

Vince, I wish you were here to read this. I'll be missing you.

Dissertation Abstract

Structured risk assessment instruments (SRAIs) are widely used among forensic mental health practitioners, criminal justice actors, and legal decision-makers to predict and manage violent or sexual recidivism risk. In part, SRAIs are relied upon to mitigate the potential for errors in risk assessment due to cognitive bias, which can occur because of external sources or internal characteristics of the evaluator. Chapter 2 of this dissertation reports results from a survey of Dutch forensic mental health practitioners ($N = 110$) about their use of SRAIs, beliefs about sources of cognitive bias in forensic risk evaluations, and opinions of potential strategies to correct for bias. In Chapter 3, we examined variations in clinicians' ($N = 54$) SRAI ratings of violent recidivism risk resulting from negative media about a parricide offender and clinicians' attitudes toward offenders. Media exposure and more positive attitudes toward offenders were associated with some lower risk ratings. An additional purpose of SRAIs is to provide information to legal decision-makers so they can make better-informed decisions. In Chapters 4–6, we examined whether an SRAI sexual recidivism risk estimate in a Sexually Violent Predator (SVP) civil commitment case mitigated the biasing effects of risk-irrelevant information about an SVP respondent among mock jurors and whether SRAI effects were moderated by individual mock juror characteristics. In Chapter 4, an SRAI risk estimate led to significant reductions in mock jurors' ($N = 427$) estimates of sexual recidivism risk and was not moderated by individual need for cognition. Similarly, in Chapter 5, an SRAI risk estimate led to significant reductions in mock jurors' ($N = 376$) estimates of sexual recidivism risk. There was no evidence of asymmetrical skepticism toward the SRAI risk estimate, in that accuracy and reliability ratings were not higher when the risk estimate confirmed the initial impression of the SVP respondent. Finally, in Chapter 6 we examined whether an SRAI sexual recidivism risk estimate mitigated bias and whether mock jurors' ($N = 402$) perceptions of sexual offenders predicted estimates of an SVP respondent's sexual recidivism risk. The SRAI risk estimate was not effective in mitigating bias from risk-irrelevant information about the SVP respondent, the effect of which was moderated by mock jurors' perceptions of sexual offenders. In addition, increased support among mock jurors for harsh sentencing and restrictive management of sex offenders in the community significantly predicted a vote for civil commitment. Overall findings of this dissertation indicate that variations in estimates of violent or sexual recidivism risk are influenced by individual characteristics of the person estimating the risk; in addition, risk-irrelevant information demonstrates a biasing effect on these estimates. There is some support for the hypothesis that SRAIs mitigate the influence of cognitive bias, but the effects tend to be moderated by the individual evaluator's attitudes toward offenders.

Keywords: Cognitive bias, forensic risk assessment, juror decision-making, sex offender civil commitment, structured risk assessment instruments

Table of Contents

DISSERTATION ABSTRACT	v
PREFACE	x
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF ABBREVIATIONS	xiv
CHAPTER 1: GENERAL INTRODUCTION	1
Background	2
Methods of Assessing Violent or Sexual Recidivism Risk	3
SRAIs in Clinical and Forensic Settings	5
Cognitive Bias and Forensic Evaluations of Risk	8
Cognitive Bias and Jurors' Perceptions of Sexual Recidivism Risk	16
Strategies to Mitigate the Effects of Bias in Forensic Risk Evaluations	22
Aims and Outline of the Dissertation	24
CHAPTER 2: FORENSIC MENTAL HEALTH PRACTITIONERS' USE OF STRUCTURED RISK ASSESSMENT INSTRUMENTS, VIEWS ABOUT BIAS IN RISK EVALUATIONS, AND STRATEGIES TO COUNTERACT IT	26
Abstract	27
Introduction	28
Method	33
Results	39
Discussion	52
Limitations and Conclusions	59
CHAPTER 3: NEGATIVE PRETRIAL PUBLICITY, EVALUATOR ATTITUDES TOWARD OFFENDERS, AND RISK ASSESSMENT WITH THE HCR-20^{V3}	62
Abstract	63
Introduction	64
Method	70

Results	77
Discussion	83
Limitations	86
Conclusion	88
CHAPTER 4: “HE SEEMS ODD”: THE EFFECTS OF RISK-IRRELEVANT INFORMATION AND ACTUARIAL RISK ESTIMATES ON MOCK JURORS’ PERCEPTIONS OF SEXUAL RECIDIVISM RISK	90
Abstract	91
Introduction	92
Method	101
Results	110
Discussion	118
Limitations	125
Conclusion	126
CHAPTER 5: ASYMMETRICAL SKEPTICISM TOWARD ACTUARIAL RISK INFORMATION AND MOCK JURORS’ ESTIMATES OF SEXUAL RECIDIVISM RISK	128
Abstract	129
Introduction	130
Method	138
Results	147
Discussion	154
Limitations and Future Research	158
Conclusions	160
CHAPTER 6: WHO’S THE MOST DANGEROUS OF THEM ALL? RISK-IRRELEVANT INFORMATION AND MOCK JURORS’ BELIEFS BIAS SEXUAL RECIDIVISM RISK ESTIMATES	162
Abstract	163
Introduction	165

Method	174
Results	182
Discussion	190
Limitations and Directions for Future Research	194
Implications and Conclusions	196
CHAPTER 7: GENERAL DISCUSSION	197
Introduction	198
Empirical Findings	200
Theoretical Implications	216
Practical Implications	222
Limitations and Directions for Future Studies	231
Conclusion	235
REFERENCES	237
IMPACT STATEMENT	308
SUMMARY	313
SAMMANFATTNING (SWEDISH SUMMARY)	319
ACKNOWLEDGMENTS	325
RESEARCH DISSEMINATION	327
CURRICULUM VITAE	328
APPENDICES	329
Appendix A	331
Appendix B	336
Appendix C	339
Appendix D	341
Appendix E	343
Appendix F	345
Appendix G	348
Appendix H	361
Appendix I	366

Preface

Chapter 1: General Introduction

Kamorowski, J., Schreuder, M., de Ruiter, C., Jelicic, M., & Ask, K. (2018). Risk assessment tools and criminal offending: Does bias determine who is “high risk”? *The Inquisitive Mind*, 38 .<https://www.in-mind.org/article/risk-assessment-tools-and-criminal-reoffending-does-bias-determine-who-is-high-risk>

Chapter 2: Forensic Mental Health Practitioners’ Use of Structured Risk Assessment Instruments, Views About Bias in Risk Evaluations, and Strategies to Counteract It

Kamorowski, J., de Ruiter, C., Schreuder, M., Ask, K., & Jelicic, M. (2021). Forensic mental health practitioners’ use of structured risk assessment instruments, views about bias in risk evaluations, and strategies to counteract it. *International Journal of Forensic Mental Health*. Advance online publication. <https://doi.org/10.1080/14999013.2021.1895377>

Chapter 3: Negative Pretrial Publicity, Evaluator Attitudes Toward Offenders, and Risk Assessment with the HCR-20^{v3}

Kamorowski, J., de Ruiter, C., Schreuder, M., Jelicic, M., & Ask, K. Negative pretrial publicity, evaluator attitudes toward offenders, and risk assessment with the HCR-20^{v3}. *Manuscript in preparation*.

Chapter 4: “He Seems Odd”: The Effects of Risk-Irrelevant Information and Actuarial Risk Estimates on Mock Jurors’ Perceptions of Sexual Recidivism Risk

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. (2021). “He seems odd”: The effects of risk-irrelevant information and actuarial risk estimates on mock jurors’ perceptions of sexual recidivism risk. *Psychology, Crime and Law*. Advance online publication. <https://doi.org/10.1080/1068316X.2021.1909016>

Chapter 5: Asymmetrical Skepticism Toward Actuarial Risk Information and Mock Jurors’ Estimates of Sexual Recidivism Risk

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. Contextual information bias and asymmetrical skepticism toward the Static-99R. *Manuscript in preparation*.

Chapter 6: Who's the Most Dangerous of Them All? Risk-Irrelevant Information and Mock Jurors' Beliefs Bias Sexual Recidivism Risk Estimates

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. Who's the most dangerous of them all? Risk-irrelevant information and mock jurors' beliefs bias sexual recidivism risk estimates. *Manuscript in preparation.*

List of Tables

Table 2.1 Risk Assessment Instruments: Required and Optional Users	42
Table 2.2 Group Mean and Differences in Usefulness Ratings of Most Frequently Used SRAIs	45
Table 2.3 Mean Ratings and Frequencies of Beliefs about Potential Sources of Cognitive Bias in Forensic Risk Evaluations	46
Table 2.4 Means and Frequencies of Effectiveness Ratings for Debiasing Strategies	48
Table 3.1 Demographic Characteristics of Participants (N = 54)	72
Table 3.2 Means and Standard Deviations for HCR-20 Subscales and HCR-20 Total Score	78
Table 3.3 Frequencies and Cross Tabulations of Final Risk Judgments by Media and Control Conditions	79
Table 3.4 Spearman ρ Correlations of ATO Scores, HCR-20 Subscale Scores, and Final Risk Judgments	80
Table 3.5 Results of Hierarchical Multiple Linear Regression Analysis Predicting HCR-20 Total Scores (N = 54)	82
Table 4.1 Risk Factors of the Static-99R	109
Table 4.2 Predictors of T2 Risk Estimates	115
Table 4.3 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' T2 Risk Estimates (No Exclusions for Failed Attention Checks)	117
Table 5.1 Risk Factors of the Static-99R	145
Table 5.2 Means (and Standard Deviations) for T2 Risk Estimates, Accuracy-Reliability Ratings, and Change in Risk Estimates	150
Table 5.3 Accuracy/Reliability Ratings, Static-99R Risk Estimate, and Contextual Information as Predictors of Change in Risk Estimates between T1 and T2	153
Table 6.1 Risk Factors of the Static-99R	180
Table 6.2 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' T1 Risk Estimates	184
Table 6.3 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' Log-Transformed T2 Risk Estimates	186
Table 6.4 Results of Hierarchical Logistic Regression Analysis of T1 Civil Commitment Recommendations	188
Table 6.5 Results of Hierarchical Logistic Regression Analysis of Civil Commitment Recommendations at T2	189

List of Figures

Figure 4.1 Raincloud Plot and Boxplot of T2 Risk Estimate Distribution

113

List of Abbreviations

ARAI	Actuarial risk assessment instrument
ATO	Attitudes Toward Offenders
ATP	Attitudes Toward Prisoners Scale (Melvin et al., 1985)
FRE	Forensic risk evaluation
NFC	Need for cognition
PCSO	Person/people who has/have been convicted of a sexual offense
PSO	Perceptions of Sex Offenders Scale (Harper & Hogue, 2015)
RAI	Risk assessment instruments
SDP	Sexually Dangerous Person
SPJ	Structured professional judgment
SRAI	Structured risk assessment instrument
SVP	Sexually Violent Predator
UCJ	Unstructured clinical judgment

Chapter 1

General Introduction

This chapter is an extended and more theoretically developed version of the publication:
Kamorowski, J., Schreuder, M., de Ruiter, C., Jelicic, M., & Ask, K. (2018).
Risk assessment tools and criminal offending: Does bias determine who is “high risk”?
The Inquisitive Mind, 38.

Background

The risk of violent or sexual recidivism is a key consideration in forensic settings and in legal decision making in the criminal justice system. Since at least 1928, when Ernest W. Burgess first created a scoring method to “predict” success or failure on parole (Burgess, 1928), there have been efforts to create statistical models to estimate the likelihood of criminal reoffending (Werth, 2019). Today, there are more than 200 structured risk assessment instruments (SRAIs) that are used to inform a variety of decisions in clinical forensic settings and in legal and correctional systems worldwide (Singh et al., 2014).

Results from an SRAI may be used to decide whether a person should be released from incarceration pending their trial, known as pretrial release (Bechtel et al., 2017), whether forensic treatment in a secure setting can be terminated to begin reintegration into society (de Ruiter, 2016), and what the nature and length of a criminal sentence should be (Monahan, 2017; Monahan & Skeem, 2016). SRAI results are used to determine an individual’s eligibility for certain types of correctional programming (Desmarais & Singh, 2013; Latessa & Lovins, 2010) and whether an individual should be granted early release on parole (Desmarais et al., 2016). Furthermore, the likelihood that a person with a mental disorder poses an imminent threat to themselves or others is a critical factor in deciding whether they should be subjected to involuntary civil commitment for treatment, and SRAIs are often used for evaluating the risk of harm to self or others (e.g., in sexual offender civil commitment cases). Perhaps most significantly, an SRAI may be used in a forensic evaluation of risk to determine whether an individual should be sentenced to life imprisonment or to death (Cunningham & Reidy, 2001, 2002; Edens et al., 2005).

There were two primary aims to this dissertation. Because SRAIs are often viewed as an objective measure of an individual’s risk of engaging in criminal recidivism, the first aim was to examine whether bias can influence SRAI outcomes. In addition, a risk estimate and risk management information derived from an SRAI is intended to help legal-decision makers make better-informed decisions (Heilbrun, 1997). Therefore, the second aim of this dissertation was to examine the joint effects of biasing information and a sexual recidivism risk estimate derived from an SRAI on mock jurors’ perceptions of sexual recidivism risk and their decisions about sex offender civil commitment.

Methods of Assessing Violent or Sexual Recidivism Risk

The substantial rise in the use of SRAIs in forensic and legal settings is partly attributable to the fact that numerous studies indicate that a violent or sexual recidivism risk estimate

derived from an SRAI tends to be more reliable and accurate than an estimate based solely on *unstructured clinical judgment* (Ægisdóttir et al., 2006; Andrews et al., 2006; Dawes et al., 1989; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009). Unstructured clinical judgment is a method of estimating the likelihood of violent or sexual recidivism that is based solely on an evaluator's experience, intuition, and professional discretion when weighing and combining risk information (Grove et al., 2000). Unstructured clinical judgment has been widely criticized as an inaccurate and unreliable method of assessing criminal recidivism risk that is susceptible to error and bias (Grove et al., 2000; Grove & Meehl, 1996; Quinsey et al., 2006b). In fact, based on a series of five prior studies, Monahan (1981) reported that psychologists and psychiatrists were correct in no more than one out of three cases when they relied solely upon unstructured clinical judgment to predict whether an individual with a mental illness who had committed a violent offense would commit further violent acts if released to the community. Although earlier studies had provided evidence of the poor accuracy of violence predictions based solely on unstructured clinical judgment (e.g., Cocozza & Steadman, 1976, 1977), Monahan's monograph sparked the intervening decades of research aimed at developing SRAs to increase the accuracy and reliability of estimates of violent or sexual recidivism risk (Shepherd & Sullivan, 2017).

SRAIs incorporate factors that are statistically predictive of violent or sexual recidivism among samples of released offenders. One type of SRAI, *actuarial risk assessment instruments* (ARAIs), is based on actuarial models in which risk factors that are most predictive of recidivism are numerically scored, weighted, and summed to arrive at a risk score. A risk score is typically associated with a score-wise risk estimate, or likelihood percentage that is determined by observed recidivism rates of individuals with the same score (Doyle & Dolan, 2007). Early ARAIs typically include only *static risk factors*, which include historical factors that are immutable, such as criminal history (Bonta & Andrews, 2007). The sole use of static risk factors among many early ARAIs renders them of little value for measuring changes in recidivism risk (Bonta & Andrews, 2007; Desmarais & Singh, 2013) or for identifying potential areas of intervention or risk management (Andrews et al., 1990; Wong & Gordon, 2006). However, newer ARAIs, such as the Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1995), incorporate both static and *dynamic risk factors*. Dynamic risk factors are factors that are associated with the likelihood of recidivism and are therefore viewed as suitable treatment targets to reduce the risk of criminal recidivism (for example, substance abuse; Bonta & Andrews, 2007).

A second type of SRAI employs *structured professional judgment* (SPJ). SPJ instruments contain both static and dynamic risk factors that have been shown to be empirically related to recidivism risk. The primary difference between SPJ and actuarial instruments is that SPJ

SRAIs enable the evaluator to use their expertise to determine the relevance and weight of the factors in an individual case (Douglas & Kropp, 2002). In contrast, ARAI risk scores are calculated based on the developer's decisions about how the factors should be scored and weighted. Generally speaking, both ARAIs and SPJ instruments exhibit similar levels of predictive validity based on observed rates of criminal recidivism (Campbell et al., 2009; Yang et al., 2010).

SRAIs in Clinical and Forensic Settings

When conducting a forensic risk evaluation, a forensic mental health professional often uses at least one SRAI to evaluate the examinee's likelihood of violent or sexual recidivism and provides testimony about the results in court (Barbaree et al., 2001; Janus & Prentky, 2003; Murrie et al., 2009). For example, they may testify about the risk of institutional violence if an offender is sentenced to life in prison as opposed to being sentenced to death (Edens et al., 2005). In addition, a forensic mental health expert is nearly always called upon to testify about the risk of future sexual offending in so-called Sexually Violent Predator (SVP) civil commitment hearings (H. A. Miller et al., 2005). The purpose of expert testimony about the risk of violent or sexual recidivism derived from rating an SRAI is to provide judges and jurors with information that is based upon scientifically supported risk factors to make ultimate legal decisions.

The importance of reliable risk estimates is difficult to overstate. The consequences of incorrectly classifying someone as being at a low risk for recidivism who, in fact, is at high risk for recidivism—known as a *false negative*—can be destructive, and even fatal. However, the consequences of a *false positive*—when an individual is classified as having a high risk of recidivism, when they have a low likelihood of recidivism—are also significant. The consequences of a false positive include wasting of limited public resources and the unjustified deprivation of liberty for the individual. The gravest result of a false positive is the potential for a death sentence for a person who could have safely served life in prison (Ewing, 1982, 1991).

While neither a false positive nor a false negative is desirable, the potential damage in each scenario is unlikely to carry equal weight in the mind of the decision-maker. On one hand, a false negative may leave the evaluator open to civil liability, public censure, or damage to their professional reputation (Weinberger & Sreenivasan, 2018). A false negative may also result in the release of an individual who subsequently commits a violent or sexual crime. Acklin (2015) notes, “The greatest fear of examiners and judges is releasing a patient who subsequently commits a highly publicized violent crime” (p. 335). A false negative thereby

poses a risk to potential victims, while also making it more likely that a false negative will be identified. On the other hand, false positives are highly unlikely to be detected (Scurich & John, 2012), and an individual who is classified as “high risk”—correctly or incorrectly—is unlikely to avoid a term of confinement or to be released to the community on parole (Boccaccini et al., 2013; Guy, Kusaj, et al., 2015; Levenson, 2004). As such, a false positive poses very little personal risk to the decision-maker, whereas a false negative represents a personal threat to the decision-maker and a more abstract threat to “public safety.”

Error management theory suggests that when decisions are made under conditions of uncertainty, and where the costs of a false-positive and a false-negative are not equivalent, a bias in one direction can help avoid a more costly mistake in the other direction (Johnson et al., 2013). The tendency to “err on the side of caution,” or avoid a subjectively more costly error, is sometimes referred to as the *error management bias*. The error management bias is the tendency to favor a relatively less costly error (i.e., a false positive) over a more costly one (i.e., a false negative), despite the fact that this tendency is likely to increase error rates (Haselton et al., 2015).

Scholars have suggested that inaccurate estimates of violent or sexual recidivism risk based on unstructured clinical judgment are partly attributable to evaluator bias (Borum et al., 1993; Commons et al., 2004; Elbogen, 2002, 2016; Faust & Ziskin, 1988; Lynett & Rogers, 2000; Monahan, 1981; Mulvey & Lidz, 1984; Quinsey et al., 2006a; S. M. Smith, 2013; Spector, 2001). Hence, it has been suggested that SRAIs mitigate the influence of evaluator bias, specifically by limiting evaluator discretion in considering which factors are relevant to recidivism risk (Bonta & Andrews, 2007; Hannah-Moffat et al., 2009; P. M. Harris, 2006; Krauss & Sales, 2001; Shepherd & Sullivan, 2017; Zapf & Dror, 2017).

Yet, there are a few studies that indicate that the use of an SRAI is no guarantee of either an objective approach to a forensic risk assessment (Chevalier et al., 2015), nor an unbiased risk estimate derived from an SRAI (e.g., Murrie et al., 2009). Even when SRAIs are used to assess criminal recidivism risk, they do not evince complete accuracy (Yang et al., 2010), nor do legal decision-makers always make decisions that align with the results (M. Stevenson & Doleac, 2019; M. T. Stevenson, 2018; for a review, see Viljoen et al., 2019). Importantly, research indicates that criminal justice actors are inclined to overestimate the likelihood of criminal recidivism, or increase the risk level associated with an SRAI through an override, perhaps to avoid what they deem to be a more costly error (i.e., violent or sexual victimization; Chappell et al., 2013; T. H. Cohen et al., 2016, 2020; Guay & Parent, 2018).

Moreover, research shows that judges tend to err on the side of caution (Hamilton, 2020; Scurich & John, 2012), as do forensic evaluators (Bonta & Motiuk, 1990; Weinberger & Sreenivasan, 2018), meaning that they are more inclined to accept the possibility of a false

positive than they are the possibility of a false negative. Specifically, the error management bias would suggest that judges and forensic evaluators may show a preference for incorrectly classifying an individual at high risk of recidivism as opposed to incorrectly classifying an individual as being at low risk for recidivism. However, overall, there is relatively little empirical research that examines how effective the use of an SRAI—or a risk estimate derived therefrom—is in mitigating the influence of cognitive bias among forensic evaluators and legal decision-makers.

Cognitive Bias and Forensic Evaluations of Risk

Kahneman and Tversky (1996; Tversky & Kahneman, 1974) identified predictions and judgments based on intuition as *heuristics*, or a form of mental shortcut. Heuristics are not necessarily negative, and in fact are likely an evolutionary adaptation that helps humans quickly categorize and process information (Haselton et al., 2015). Without this ability, we would quickly become overwhelmed if we had to process every situation as if it were a novel event. However, under conditions of uncertainty, automatic and unconscious processing of information based on heuristics can lead to *cognitive bias*, or systematic errors in judgment (Chaiken, 1980; Tversky & Kahneman, 1974) or suboptimal decisions (Korteling et al., 2018).

The dual-process model of decision-making indicates two primary systems by which information is evaluated. *System 1* (Tversky & Kahneman, 1974), sometimes referred to as experiential processing, is associated with intuitive, preconscious decisions that are influenced by experiences, affect, narratives, and prototypes (e.g., Denes-Raj & Epstein, 1994; Epstein, 1993). In contrast, *System 2* (Tversky & Kahneman, 1974) is associated with conscious and analytical deliberation, and has been referred to as rational processing (Denes-Raj & Epstein, 1994). However, when ‘automatic’ or System 1 processing of information occurs, one example of bias that can lead a person to ignore, devalue, or undervalue the importance of available evidence that contradicts their preexisting beliefs is *confirmation bias* (Charman, 2013; Nickerson, 1998).

The term *bias*, or cognitive bias, is used in this dissertation to refer to the extent to which judgments are based upon factors that are irrelevant to the decision at hand. Although some decision-making research (e.g., Busemeyer & Townsend, 1993) and several forensic science experiments (for a review, see G. S. Cooper & Meterko, 2019) have demonstrated that cognitive bias can affect the perception and interpretation of information, and the relevant decision-making process *without changing the ultimate outcome* in some circumstances (Dror, 2009; Dror & Murrie, 2018; Kukucka, 2014), it “does not negate the existence of bias and its distorting effects [...], especially where the bias pushes and shifts the evidentiary weight past

the decision threshold” (Dror, 2009, p. e17). In fact, Dror and Murrie (2018) have suggested that *irrelevant contextual information*—information that is not necessary to the forensic domain or task—can bias how evidence is collected, the meaning and value attributed to the evidence, and the conclusions drawn therefrom. The critical point is that information that is irrelevant to the forensic practitioner’s task represents a threat to the “independent probative value of forensic examiners’ judgments, even if their judgment is accurate” (Kukucka, 2020, p. 111).

The biasing effect of irrelevant contextual information has been demonstrated through research and discussed extensively in several areas of forensic science (e.g., Dror et al., 2006; Hamnett & Dror, 2020; Kassin et al., 2013; Nakhaeizadeh et al., 2014; Quigley-McBride, 2020). With the exception of a few studies (e.g., Murrie et al., 2013), published research regarding the biasing effects of irrelevant contextual information on observations or conclusions in forensic mental health evaluations of any kind (including forensic risk evaluations) appears to be very limited (Dror & Murrie, 2018; Meyer & Valen  a, 2021; Zapf & Dror, 2017). Hence, the experimental studies in this dissertation are focused on the extent to which task-irrelevant information affects risk judgments among forensic mental health evaluators, and risk estimates and related outcome decisions among mock jurors.

Dror (2020) has proposed a model of factors that affect expert decision-making which includes eight sources of bias that can be grouped into three main categories of: (a) case-specific factors; (b) factors specific to the person conducting the analysis; and (c) bias that arises from how the human brain processes information. Dror’s model may also be useful as a framework for research related to bias that can affect jurors. However, in this dissertation, I refer to *external* and *internal* sources of bias, adopting the terminology of Gowensmith and McCallum (2019) related to threats to objectivity in forensic evaluations. Bias can occur due to the influence of factors external to the evaluator or legal decision-maker (i.e., irrelevant contextual information) but also as a consequence of internal factors, such as the personal characteristics or beliefs of the evaluator or legal decision-maker (Gowensmith & McCallum, 2019).

The term external sources of bias, as used in the present research (e.g., negative media representations of a criminal defendant; positive and negative risk-irrelevant characteristics of a person convicted for a sexual offense), could be understood as Dror’s (2020) case-specific factors. Likewise, internal sources of bias, as referred to in the present research (e.g., attitudes toward offenders), could be considered analogous to Dror’s reference to factors specific to the person conducting the analysis and bias that arises from how the human brain processes information (e.g., need for cognition and asymmetrical skepticism). Potential external and internal sources of bias are explained in more detail in subsequent paragraphs.

In forensic contexts, the people who must interpret evidence, draw conclusions, and make decisions include forensic evaluators, treatment providers, judges, and jurors. Yet, when a forensic expert is exposed to potentially biasing, task-irrelevant contextual information, jurors show a tendency to believe an expert who denies that they have been biased by such information. Moreover, research indicates that jurors themselves are influenced by task-irrelevant contextual information in their decisions (Boyll, 1991; Ford, 1986; LaFree et al., 1985; Reskin & Visher, 1986). In other words, bias has not only the potential to affect the judgment of forensic evaluators, but also jurors' decisions based on an expert's testimony. Therefore, it seems desirable that jurors have access to task-relevant information, where it is available, that presents the most objective assessment of the legal issue in question. Therefore, forensic mental health practitioners and mock jurors were the samples employed in the studies presented in this dissertation.

External Sources of Evaluator Bias in Recidivism Risk Estimates

One well-established example of an external source of evaluator bias is known as *adversarial allegiance*, which is the unconscious tendency to reach conclusions that support the party that retained the evaluator (Boccaccini, Chevalier, et al., 2017; Chevalier et al., 2015; Murrie et al., 2009, 2013). For example, Murrie and colleagues (2008) found evidence suggesting that clinical assessments of psychopathy using a structured measure (the Psychopathy Checklist revised, or PCL-R [Hare, 2003]) varied between evaluators depending upon whether they were hired by the prosecution or the defense. Psychopathy is a constellation of traits that is strongly related to violent and sexual recidivism and is commonly assessed in forensic risk evaluations. Murrie et al. reported that PCL-R scores for the same examinee were significantly higher among evaluators hired by the prosecution than among those hired by the defense, suggesting the existence of an adversarial allegiance effect. In a subsequent study, Murrie et al. (2009) found that in actual SVP civil commitment cases, approximately 23% of the variance in PCL-R scores between the state experts and the defendant experts was attributable to the "side" by which the expert was employed. Although the PCL-R is not an SRAI, *per se*, additional studies provide empirical evidence indicating that SRAI scoring is also affected by whether the expert was retained by the prosecution or the defense (Boccaccini, Chevalier, et al., 2017; DeMatteo et al., 2014; Murrie et al., 2013).

In addition, there are several types of irrelevant contextual information that have been shown to bias evaluators when assessing recidivism risk. Such information includes a victim impact statement (Jackson et al., 2004; Lynett & Rogers, 2000), "morally reprehensible" factors such as the offender's lack of victim empathy or lack of treatment motivation (Huls

et al., 2018), and information that evokes fear or anxiety in the evaluator, such as a graphic accounting of the violent offense (Blumenthal et al., 2010). However, with the exception of the studies on adversarial allegiance (e.g., Murrie et al., 2008, 2009, 2013), external factors identified as having the potential to bias forensic evaluators have rarely been investigated with respect to their effects on a recidivism risk estimate derived from the results of an SRAI (Dror & Murrie, 2018).

Collateral Information. Although an interview of the examinee is important to a forensic risk evaluation, a forensic mental health evaluator cannot properly rely solely upon the examinee to provide the information necessary to conduct a forensic risk evaluation because the examinee may not be willing or able to accurately report on their own conduct (American Psychological Association, 2013; Heilbrun et al., 2009; Melton et al., 2007). Therefore, a forensic evaluator is expected to obtain collateral information from a variety of sources before rendering their analysis of an examinee to inform decision-makers (de Ruiter & Kaser-Boyd, 2015), which may include interviews with third parties and records review (Heilbrun et al., 2009). Yet, there is the potential that collateral sources of information will expose the evaluator to irrelevant information that can bias the forensic risk evaluation. However, what constitutes irrelevant, potentially biasing information in a forensic risk evaluation is not always clear.

For example, increasingly, forensic evaluators often seek information from the Internet about forensic examinees (Coffey et al., 2018; Metzner & Ash, 2010; Neimark et al., 2006; Recupero, 2008, 2010; Vitacco et al., 2018). Particularly in high-profile cases, media about the examinee and the crime the examinee committed is likely to be encountered in an Internet search, and is likely to reflect negatively on them (Bakhshay & Haney, 2018; Grabe et al., 2001; Steblay et al., 1999; Studebaker & Penrod, 1997; Yanich, 2004). Yet, there is very little professional guidance about how the information should be incorporated into a forensic risk evaluation (Coffey et al., 2018; Pirelli et al., 2016, 2018; Vitacco et al., 2018).

Relatedly, it is possible that negative depictions of a forensic examinee in the media will improperly influence the forensic evaluator's perceptions of the examinee. For example, a substantial body of literature indicates that negative pretrial media exposure tends to increase jurors' perceptions of the defendant's guilt and culpability (e.g., Beale, 2006; Daftary-Kapur et al., 2010, 2014; Moran & Cutler, 1991; Otto et al., 1994; Ruva et al., 2007; Steblay et al., 1999). However, there appear to be no published empirical studies that examine the potentially biasing effects of negative media coverage on an evaluator's evaluation of a forensic examinee's risk of recidivism. Whether negative pretrial publicity may influence ratings of recidivism risk using an SRAI is examined in the experimental study presented in Chapter 3 of this dissertation.

Internal Sources of Evaluator Bias in Recidivism Risk Estimates

A forensic evaluator's individual characteristics may affect how they score an SRAI. In what is reportedly the first published exploratory study to examine this possibility, A. K. Miller and colleagues (2011) measured personality characteristics of 22 graduate students and faculty members in a clinical psychology program using the NEO Personality Inventory-Revised (NEO PI-R; Costa & McCrae, 1992). The NEO PI-R is a 240-item measure of the Five-Factor (FFM) model of personality for traits of Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C). The results showed that evaluators who were more agreeable tended to rate offenders as less psychopathic (as measured by the PCL-R), whereas evaluators who were higher in conscientiousness tended to rate offenders as more psychopathic. There were also significant correlations between other aspects of evaluator personality and scores of the four facets of the PCL-R (i.e., Interpersonal, Affective, Impulsive Lifestyle, and Antisocial Behavior). For example, evaluators whose own level of extraversion and excitement-seeking was high, tended to rate offenders as lower on the PCL-R Impulsive Lifestyle facet. Furthermore, some raters consistently assigned higher scores on the PCL-R than did other raters (A. K. Miller et al., 2011).

Moreover, Boccaccini and colleagues (2008) found that differences in how trained clinicians (psychologists and psychiatrists) completed the PCL-R accounted for 30% of variability in scores for the same examinee. As possible explanations for why variations occur when different evaluators score the same examinee, Boccaccini et al. suggested factors such as the evaluator's willingness to seek additional information or their level of adherence to scoring procedures in the PCL-R manual. However, in Boccaccini et al.'s study, there was no examination of characteristics of the evaluators that could have explained the observed variability. Overall, published studies that examine how individual evaluator characteristics affect how they rate recidivism risk using an SRAI are scarce. Nevertheless, the few studies that do exist indicate the need to further explore specific evaluator characteristics in accounting for variability in recidivism risk ratings.

Evaluator Attitudes Toward Offenders. Studies examining whether evaluator attitudes toward offenders affect SRAI ratings appear to be absent from the literature. The closest comparison may be research that examines differences in evaluator ratings based on their feelings toward specific patients, of which there are two examples. In one study, de Vogel and de Ruiter (2004) found that among clinicians who were involved in treatment of forensic psychiatric patients in a secure facility, clinicians' negative feelings toward the patient (e.g., feeling controlled or manipulated) were significantly correlated with higher ratings of the patient's risk of violent recidivism on the Historical-Clinical-Risk Management-20 v.2 (HCR-

20 v.2; Webster et al., 1997). Conversely, when clinicians reported more positive feelings toward the patient, these were correlated with lower ratings of violence recidivism risk on the HCR-20. An earlier study conducted by Dernevik and colleagues (2001) indicated a similar pattern of findings among forensic nurses' HCR-20 ratings of psychiatric patients, although the study was limited by a small sample size of five nurses who rated eight patients.

Clinicians' feelings toward specific patients are different from more generalized attitudes toward offenders or attitudes toward specific subgroups of offenders (e.g., people who have been convicted of a sexual offense). Furthermore, a dual relationship in which a treatment provider acts as a forensic evaluator is generally discouraged (Gordon, 2016; Shapiro & Walker, 2016). Therefore, examining more generalized attitudes toward offenders among evaluators who do not have a clinical relationship with the examinee may provide more ecological validity in examining how evaluator characteristics affect SRAI ratings of recidivism risk. Hence, empirical testing of this question was undertaken in this dissertation in the experimental study involving evaluators who have been trained to complete the HCR-20 (see Chapter 3).

Cognitive Bias and Jurors' Perceptions of Sexual Recidivism Risk

Whether SRAIs are effective in mitigating forensic evaluator bias is but one aspect of SRAIs' effectiveness in producing legal decisions in alignment with an objective measure of recidivism risk. In fact, forensic evaluators and forensic experts who provide testimony about an examinee's risk of violent or sexual recidivism are often relied upon to *inform* legal decision-makers (e.g., judges and jurors). Therefore, whether a risk estimate derived from an SRAI is effective in mitigating bias among judges and jurors should also be examined, although to date, such studies appear absent from the literature. Employing an SRAI to objectively measure recidivism risk is of limited value if the resulting risk estimate does not have the effect of mitigating biased decision-making by the end-users of the information (i.e., judges, jurors).

An example helps illustrate the point: SVP statutes require that SVP civil commitment be limited to people who have been convicted of a sexual offense and who have a mental disorder or abnormality that renders it more likely than not that they will engage in future sexual offending (i.e., recidivism risk) unless they are confined for treatment (*Kansas v. Crane*, 2002; *Kansas v. Hendricks*, 1997). Therefore, the role of a sexual recidivism risk estimate based on an SRAI is particularly important for establishing one of the legal criteria to justify civil commitment. Yet, several studies indicate that jurors in SVP civil commitment

cases tend to give relatively little weight to expert testimony about recidivism risk relative to other case information (Boccaccini et al., 2013; Krauss et al., 2012; Lieberman et al., 2007). Findings from these studies suggest that jurors' decisions in SVP cases may be influenced by factors other than the risk estimate derived from an SRAI. Yet it remains unclear why jurors tend to neglect expert testimony about recidivism risk estimates (S. L. Miller & Brodsky, 2011; Pennycook et al., 2014; Walters et al., 2014).

External Sources of Juror Bias in Recidivism Risk Estimates

When jurors have limited information about the offender, they may rely upon heuristics or stereotypes to evaluate the offender's risk of reoffending (Bodenhausen & Wyer, 1985). The extent to which an offender fits a stereotype associated with a particular type of crime can affect decisions about arrest, prosecution, guilt attributions, conviction, and sentencing for the crime (Sanghara & Wilson, 2006; Skorinko & Spellman, 2013; Stalans, 1993). For example, the public tends to believe that the majority of sexual offenses are committed by strangers to the victim (Wurtele et al., 1992) and by "loners" who have major mental disorders (Levenson et al., 2007) and lack a social life (Borhart & Plumm, 2015). Therefore, the extent to which an SVP *respondent*—the person whom the government seeks to remand to a secure facility for sexual offender treatment—fits stereotypical beliefs about the kind of person likely to commit a sexual offense may affect jurors' perceptions of the respondent's likelihood of sexual recidivism.

Characteristics of the Offender and Case Information. Characteristics, such as the respondent's *social attractiveness* or likability (Alicke & Zell, 2009; Landy & Aronson, 1969; Michelini & Snodgrass, 1980; Richardson & Campbell, 1982) are not relevant to, nor empirically associated with, the risk of sexual recidivism. For example, good deeds or special skills may increase the likability of an individual who has committed a sexual offense, but such facts can be considered *risk-irrelevant contextual information* in that they are not empirically associated with sexual recidivism risk. However, factors that may increase an SVP respondent's likability may also conflict with commonly held stereotypes among the public about people who have committed a sexual offense (PCSOs), thereby affecting jurors' perceptions of the SVP respondent's risk of sexual recidivism. For example, a PCSO who is outgoing and involved in the community (e.g., volunteerism) may conflict with the "loner" stereotype often associated with PCSOs (Borhart & Plumm, 2015; Levenson et al., 2007).

Moreover, bias created by risk-irrelevant contextual information may cause some jurors to overestimate the respondent's risk of sexual reoffending and, thereby, increase the likelihood of voting for civil commitment (Gunnell & Ceci, 2010; Hilton et al., 2005, 2015). In fact, factors that are not empirically related to sexual recidivism risk have been shown to influence

forensic evaluators' recommendations to a court regarding sentencing for PCSOs (McCallum et al., 2017). Yet, there is limited research on whether risk-irrelevant contextual factors may influence jurors' case-specific estimates of reoffending risk or SVP civil commitment decisions (Boccaccini et al., 2013). Further, if risk-irrelevant contextual factors can bias SVP jurors, it is important to examine whether a risk estimate from an SRAI can effectively mitigate the bias so that civil commitment decisions are based on what is presumably the most accurate, risk-relevant information available—that is, information derived from an SRAI.

Internal Sources of Juror Bias in Recidivism Risk Estimates

Preexisting beliefs about the recidivism rates of people who have committed a sexual offense (PCSOs) are likely to influence SVP jurors' perceptions of an SVP respondent's risk of sexual recidivism. In fact, the public tends to grossly overestimate sexual recidivism rates compared to observed sexual recidivism rates (e.g., Levenson et al., 2007) and to perceive PCSOs as a homogeneous group that is not amenable to treatment (Katz-Schiavone et al., 2008; Levenson et al., 2007). Furthermore, the public tends to be highly punitive toward PCSOs (Levenson et al., 2007). These perceptions and attitudes have a substantial impact on public support for imposing harsh punishment and severe restrictions on PCSOs (Rydberg et al., 2018). Despite the public's generally negative opinions of PCSOs, there appears to be no published research examining how individual perceptions of PCSOs may influence SVP jurors' perceptions of sexual recidivism risk and civil commitment recommendations.

Perceptions of People Who Have Committed a Sexual Offense. Harper and Hogue (2015) proposed a model to explain how public perceptions of the recidivism risk of PCSOs, endorsement of PCSO stereotypes, and punitiveness toward PCSOs influence laypersons' decisions about sexual recidivism and support for severe restrictions on PCSOs. Using this model, they created a Perceptions of Sexual Offenders (PSO) scale. This scale contains 20 items that measure three facets of perceptions of PCSOs, including support for harsh sentencing and restrictive management (sentencing and management scale), stereotype endorsement, and risk perception. Higher scores indicate more negative perceptions of PCSOs. The study presented in Chapter 6 employs the PSO scale to examine how mock jurors' perceptions of PCSOs affect their estimates of the SVP respondent's risk of sexual recidivism and civil commitment decisions.

Although states disavow any punitive intent with respect to SVP civil commitment (e.g., *Kansas v. Hendricks*, 1997), it may not always be clear to jurors in SVP cases that it is impermissible to vote for civil commitment for punitive reasons (Carlsmith et al., 2007; Krauss et al., 2018). In fact, the SVP respondent has typically served a term of incarceration in jail or prison for the offense prior to being referred for consideration for SVP civil commitment.

Assuming jurors make their best efforts to follow the law in SVP civil commitment cases, it remains plausible that certain risk-irrelevant qualities or characteristics of the offender may affect jurors' estimates of the SVP respondent's sexual recidivism risk. Furthermore, other factors surrounding the case (e.g., media attention, victim statements)—even though they are irrelevant to evaluating whether the legal criteria for SVP civil commitment are met—could potentially bias jurors' decisions in these cases. Yet, the potentially biasing effects of specific offender characteristics or case factors that are irrelevant to evaluating the legal criteria for SVP civil commitment have received relatively little attention with respect to jurors' perceptions of sexual recidivism risk and decision-making in civil commitment cases.

Individual Differences in Information Processing. There may be other individual differences that affect the extent to which jurors are influenced by a sexual recidivism risk estimate derived from an SRAI. Previous researchers have suggested that a lack of knowledge about science (Batastini, Hoeffner, et al., 2019) or a lack of numeracy (Krauss et al., 2004; Krauss & Lee, 2003; Krauss & Sales, 2001) may account for jurors' tendency to discount expert testimony based on SRAI results. However, others have argued that individual differences may contribute to how information is perceived and integrated in the context of legal decision-making (Gunnell & Ceci, 2010; Krauss et al., 2004; Lieberman, 2002; Lieberman et al., 2007). Therefore, there is a lack of agreement about how individual differences in information processing influence the relative weight of a sexual recidivism risk estimate derived from an SRAI compared to other factors present in an SVP case. One information processing construct that may be relevant in this context is the need for cognition.

Need for Cognition. Individuals differ in the extent to which they enjoy expending cognitive effort to understand a situation, a characteristic known as *need for cognition* (NFC; Cacioppo & Petty, 1982). People with low NFC tend to process information experientially, which means they engage in rapid decision-making that is largely unconscious, emotion-driven and leads to conclusions based on generalizations (Epstein, 1993). They also tend to focus on peripheral cues, such as the attractiveness or likability of the source of information (Cacioppo et al., 1996; Haugvedt et al., 1992). Conversely, people with higher NFC appear more likely to use a rational approach by exerting conscious effort to seek evidence and apply logic in their thinking (Epstein, 1993). Cognitive-experiential self-theory (CEST) suggests people use both rational and experiential modes of processing, but that individual differences and situation-specific factors have a significant effect on which mode of processing predominates (Denes-Raj & Epstein, 1994).

On one hand, when jurors think analytically and rationally, a sexual recidivism risk estimate derived from an SRAI appears to have more influence on their decision-making in SVP cases (cf., Krauss et al., 2012; Lieberman et al., 2007; Lieberman & Krauss, 2009). On the

other hand, when jurors engage in decision-making that is based on emotion, they tend to be less influenced by SRAI risk estimates (Gunnell & Ceci, 2010; Krauss et al., 2004; Lieberman et al., 2007). The existing research suggests that individual levels of NFC may thereby have some influence on the extent to which jurors are influenced by potentially biasing external information as compared to a sexual recidivism risk estimate derived from an SRAI. This hypothesis was examined in the study presented in Chapter 3 of this dissertation.

Asymmetrical Skepticism. A substantial body of research suggests that people tend to interpret information in a way that confirms their preexisting beliefs and preferences (Ditto et al., 1998; Ditto & Lopez, 1992; Kunda, 1990; Pyszczynski & Greenberg, 1987) while also concluding that contradictory information is flawed and invalid (Ditto et al., 2003; Ditto & Lopez, 1992). The tendency to accept preference-consistent information as valid and preference-inconsistent information as faulty is known as *asymmetrical skepticism* (Ditto et al., 2003; Ditto & Lopez, 1992), which is a form of cognitive bias. Given previous research that indicates that asymmetrical skepticism is implicated in interpretation of criminal evidence (e.g., Ask & Granhag, 2007), it is possible that jurors in SVP cases will be motivated to undervalue or disregard a sexual recidivism risk estimate from an SRAI that conflicts with their sexual recidivism risk perceptions based on other factors. This hypothesis was tested in Chapter 5 of this dissertation.

Strategies to Mitigate the Effects of Bias in Forensic Risk Evaluations

Given the research suggesting that SRAIs may be vulnerable to external or internal sources of evaluator bias (e.g., Boccaccini, Chevalier, et al., 2017; de Vogel & de Ruiter, 2004), there is a need to identify strategies that may be effective in mitigating the influence of evaluator bias on forensic evaluations of risk (Neal & Brodsky, 2016). Although the use of SRAIs is commonly believed to be effective in limiting evaluator discretion and thereby increasing the reliability and accuracy of forensic risk evaluations, these conclusions may be premature given the relatively limited research on the effects bias may have on completion of an SRAI, on the overall risk evaluation, or on subsequent risk management-related recommendations (McCallum et al., 2017).

Recently, there has been increasing attention to the issue of bias awareness among forensic mental health evaluators (e.g., MacLean et al., 2019; Zapf et al., 2018; Zappala et al., 2017). Yet, there appears to be a lack of consensus among forensic mental health professionals about sources of bias in forensic mental health evaluations. For example, Zapf et al. (2018) surveyed an international sample of forensic mental health evaluators ($N = 1,099$) about the

scope and potential sources of bias in forensic evaluations. Zapf and colleagues reported that more than half of their respondents agreed about the nature of bias in forensic evaluations in only three out of 10 scenarios. Specifically, a substantial majority of evaluators agreed to some extent that: prior beliefs and expectations can affect (1) how an evaluator analyzes a case (93.8%), and (2) the evaluator's ultimate opinion of the case (91.3%). Moreover, 83.1% of Zapf et al.'s respondents agreed that evaluators sometimes know what conclusion they are expected to reach and 69.2% also agreed that when an evaluator knows what they are expected to find, it affects the conclusions they reach.

Therefore, it is concerning that a substantial majority (87.2%) of Zapf et al.'s (2018) respondents also agreed to some extent that an evaluator who consciously tries to set aside their prior beliefs and expectations is less likely to be influenced by them, despite the fact that introspection has been shown to be an ineffective strategy for mitigating the influence of bias (Nisbett & Wilson, 1977a; Pronin & Kugler, 2007; Wilson & Brekke, 1994). The fact that a substantial number of forensic evaluators agreed that the potential for bias in forensic evaluations exists, combined with the fact that a substantial majority also agreed that introspection is an effective debiasing strategy, highlights the need for research regarding debiasing strategies that may be effective in forensic evaluations.

In fact, although many strategies have been suggested in the literature as potentially effective in mitigating evaluator bias, very few have been empirically tested. For example, it has been suggested that evaluators should examine their own patterns of decision-making to identify potential bias (Murrie & Balusek, 2008; Murrie & Warren, 2005; Parker, 2016), yet there appears to be no empirical research that tests this hypothesis. Numerous other strategies have been suggested (see, e.g., Neal & Brodsky, 2016) as potentially effective in mitigating bias in forensic evaluations, yet almost none have been empirically tested in this context. Nevertheless, forensic evaluators' perceptions of the effectiveness of these strategies may help identify strategies that should be empirically tested in the future. In line with the potential sources of bias in forensic risk evaluations suggested by Zapf et al. (2018) and the debiasing strategies proposed by Neal and Brodsky (2016), Dutch mental health practitioners' beliefs about sources of bias and debiasing strategies are presented in Chapter 2 of this dissertation.

Aims and Outline of the Dissertation

Despite the belief that SRAIs mitigate the potential influence of cognitive bias in forensic risk evaluations, this is not necessarily always the case (Shepherd & Sullivan, 2017). There is a growing body of research indicating that simply using an SRAI does not necessarily mean that an evaluator has arrived at an unbiased recidivism risk estimate, nor that they have

made objective treatment or disposition recommendations based on the results (Boccaccini, Chevalier, et al., 2017; Chevalier et al., 2015; J. Miller & Maloney, 2013; Murrie et al., 2009, 2013). Therefore, the first aim of this dissertation is to examine whether bias can influence the outcome of an SRAI that has been completed by a trained clinical evaluator.

In addition, when jurors must make decisions related to sentencing or treatment, the risk estimate derived from an SRAI is only one piece of information they consider. Therefore, it is important to empirically investigate not only how bias may affect an evaluator who uses an SRAI, but also how bias may affect decision-makers' use of SRAI information in the context of individual difference and case-related factors. Therefore, the second aim of this dissertation is to examine the joint effects of biasing contextual information and a sexual recidivism risk estimate derived from an SRAI on mock jurors' perceptions of sexual recidivism risk and their decisions about sex offender civil commitment. Finally, identifying potential sources of bias and strategies to mitigate bias in forensic risk evaluations may help prompt empirical research that can improve the objectivity of forensic risk evaluations. Therefore, mental health evaluators' perceptions of the potential for bias in forensic risk evaluations and mitigation strategies are also examined.

This dissertation consists of seven chapters. Chapter 2 describes a survey of Dutch forensic mental health practitioners' use and perceived utility of SRAIs and their views on potential sources of bias and bias mitigation strategies in forensic risk evaluations. Chapter 3 is an empirical study of the effects of negative pretrial media exposure and clinicians' attitudes toward offenders on their ratings of violence recidivism risk using an SRAI. Chapter 4 is an experimental study that examines whether irrelevant contextual information about an SVP respondent biases mock jurors, whether sexual recidivism risk information derived from an SRAI mitigates bias, and if mock jurors' level of need for cognition affects their perceptions of sexual recidivism risk. Chapter 5 examines whether mock jurors exhibit asymmetrical skepticism toward a sexual recidivism risk estimate from an SRAI that conflicts with their initial perceptions of an SVP respondent. Chapter 6 tests the joint influence of mock jurors' perceptions of PCOSs and an SRAI-derived sexual recidivism risk estimate on their estimates of an SVP respondent's sexual recidivism risk and their decisions about whether the respondent should be civilly committed. Finally, Chapter 7 provides a summary and discussion of the findings and implications from the studies presented in Chapters 2-6.

Chapter 2

Forensic Mental Health Practitioners' Use of Structured Risk Assessment Instruments, Views About Bias in Risk Evaluations, and Strategies to Counteract It

This chapter is based on the following paper:

Kamorowski, J., de Ruiter, C., Schreuder, M., Ask, K., & Jelicic, M. (2021). Forensic mental health practitioners' use of structured risk assessment instruments, views about bias in risk evaluations, and strategies to counteract it.

International Journal of Forensic Mental Health. Advance online publication.
<https://doi.org/10.1080/14999013.2021.1895377>

Abstract

The use of structured risk assessment instruments (SRAIs) has increased significantly over the past decades, with research documenting variation between countries. The use of SRAIs, their perceived utility and potential for mitigating bias in forensic risk evaluations (FREs) was investigated in a survey of Dutch forensic mental health practitioners ($N = 110$). We found generally positive views regarding SRAI utility. Bias in FREs was of concern to respondents. We found no evidence of a bias blind spot (the belief that oneself is less prone to bias than peers/colleagues). SRAIs were rated as the most effective debiasing strategy, but respondents also endorsed introspection. There were few differences in beliefs about sources of bias or debiasing strategies between respondents who had bias training and those who had not, suggesting the need for development of effective strategies to mitigate bias and training related to bias in FREs.

Introduction

Forensic mental health practitioners play an important role in advising courts about the treatment and management of people who suffer from a mental illness and have been accused or convicted of a crime. Forensic mental health evaluations are a critical component of legal decision making because judges tend to follow the recommendations made by a behavioral expert (Gowensmith et al., 2012, 2014; Leij et al., 2001; Messina et al., 2019). Specifically, an offender's estimated risk of violent or sexual reoffending is a key consideration in determining what conditions are necessary—and legally justified—to minimize the risk of harm to potential victims and to make the best use of limited resources (Harte & Breukink, 2010).

Use of Structured Risk Assessment Instruments in Forensic Risk Evaluations

Research conducted over the past decades indicates that using structured approaches in forensic risk evaluations (FREs)—particularly structured risk assessment instruments (SRAIs)—increases the accuracy of predictions about the likelihood of future violent or sexual offending, as compared to *unstructured clinical judgment* (Ægisdóttir et al., 2006; Andrews et al., 2006; Dawes et al., 1989; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009). Unstructured clinical judgment (UCJ) is a method of FRE in which the clinician uses only their experience and intuition to evaluate an examinee's risk of future violence (Grove et al., 2000). In a seminal monograph, Monahan (1981) revealed that UCJs about the likelihood of violent reoffending were correct in about one out of three cases, which lead to criticisms of UCJ as being inaccurate and unreliable (Douglas & Kropp, 2002).

Since the early 1980s, an astonishing variety of standardized and evidence-based SRAIs have been developed for evaluating the potential for different types of violence (Shepherd & Sullivan, 2017). For example, among a large-scale international survey of forensic mental health professionals ($N = 2,135$), the respondents reported using over 200 different commercially available SRAIs and more than 200 different locally-developed SRAIs (Singh et al., 2014). SRAIs provide information about the probability of recidivism, the severity of the consequences if reoffending occurs, and whether the reoffending risk is imminent or more remote. SRAIs can also help to identify offender treatment needs that, if addressed, may reduce the risk of reoffending (Bonta & Andrews, 2007).

Actuarial risk assessment instruments are one type of SRAI and contain factors empirically related to an increased risk of violent, sexual, or general criminal recidivism (Doyle & Dolan, 2007), depending on the instrument's purpose. Risk factors are scored and combined according to an algorithm determined by the tool developers. The final risk

score thereby provides a recidivism risk estimate based on the recidivism rates of groups of individuals with the same score (Doyle & Dolan, 2007). A second type of SRAIs employs *structured professional judgment* (SPJ), a method in which the evaluator considers the presence or absence of empirically-based risk factors and risk factors they deem relevant based on expertise. The evaluator uses their professional judgment to determine the relative importance of these factors to the individual case (Douglas & Kropp, 2002). Actuarial and SPJ instruments generally exhibit approximately similar predictive validity (Campbell et al., 2009; Yang et al., 2010), but the type of SRAI an evaluator chooses may depend on the setting and purpose of the risk evaluation (B. Brown & Rakow, 2016).

Numerous national and international surveys of forensic mental health professionals indicate that the use of SRAIs to assess the risk of violent recidivism is increasingly common in FREs (Archer et al., 2006; Hurducas et al., 2014; Kelley et al., 2020; Lally, 2003; Neal & Grisso, 2014b; Singh et al., 2014). Findings from more recent surveys also indicate that SRAI use and FRE practices can vary widely across countries (e.g., Canada: McLaughlin & Kan, 2014; Denmark: Nielsen et al., 2015 ; Belgium: Pham et al., 2016; Israel: Singh et al., 2019). Yet there is relatively little country-specific research about which SRAIs forensic evaluators are required to use in practice, which SRAIs they choose to use, and their perceptions about the usefulness of those tools (Hurducas et al., 2014).

Use of SRAIs in Forensic Risk Evaluations in the Netherlands

SRAIs were introduced into the Dutch forensic mental health system in the late 1990s (de Ruiter & Hildebrand, 2007) and are now commonly used in FREs. Dutch forensic mental health experts evaluate approximately four to five thousand criminal defendants a year and produce a written report for a court (Messina et al., 2019). In the Netherlands, a person who has committed a crime and has a mental disorder can be ordered to treatment in a secure psychiatric facility, usually after time served in prison, a disposition known as *maatregel van terbeschikkingstelling* (TBS; de Ruiter & Hildebrand, 2007). Although the goal of treatment under TBS is to successfully reintegrate the offender into the community, a TBS order can entail lifelong mandatory treatment (Bogaerts et al., 2018). SRAIs are therefore used to assess offenders both for potential commitment to TBS, and for regular reviews to determine if the individual's risk has been lowered to a degree sufficient to warrant release (de Ruiter, 2016).

We are unaware of any published surveys of Dutch forensic mental health practitioners about the SRAIs they most commonly use and the perceived utility of those instruments. Two relatively recent studies published in the Dutch language reported on SRAIs commonly used

in forensic settings in the Netherlands (Harte & Breukink, 2010; van Horn et al., 2016), but neither surveyed practitioners about which SRAIs they use in practice, nor the perceived utility of the SRAIs. Therefore, in the current study, we survey forensic mental health practitioners about the SRAIs they are required or choose to use and the perceived usefulness of specific SRAIs.

Bias Awareness and Sources of Bias in Forensic Risk Evaluations

Cognitive bias refers to systematic errors in logic or reasoning that occur outside of conscious awareness (Wilson & Brekke, 1994) and are the result of the mind's automatic processing of information based on experience and prior expectations (Tversky & Kahneman, 1974). Therefore, cognitive bias represents a threat to the objectivity and validity of a forensic mental health evaluation (Neal & Grisso, 2014a; Zapf & Dror, 2017). In addition, the reliability and accuracy of FREs depend to a large extent on the evaluator's ability to minimize the influence of cognitive bias and reach objective conclusions (MacLean et al., 2019).

A number of recently published surveys suggest that forensic mental health practitioners are aware of, and concerned about, the potential for bias in forensic mental health evaluations (Kukucka et al., 2017; Neal & Brodsky, 2016; Neal & Grisso, 2014b; Zapf et al., 2018). However, many remain skeptical about bias affecting their own work as evidenced by a *bias blind spot* (Pronin et al., 2002), that is, the belief that they are less prone to bias than their colleagues (Boccaccini, Chevalier, et al., 2017; Kukucka et al., 2017; Neal & Brodsky, 2016; Zapf et al., 2018; Zappala et al., 2017). For example, Zapf and colleagues (Zapf et al., 2018) surveyed 1,099 mental health practitioners who conduct forensic evaluations and just over half (52.2%) agreed that their own judgments can be influenced by cognitive bias (Zapf et al., 2018).

In order for forensic evaluators to take appropriate steps to counter bias, they must first be aware of it, have accurate perceptions about how bias operates (e.g., Kukucka et al., 2017; MacLean et al., 2019; Zapf & Dror, 2017), and accept the potential for it to affect their work (Dror, 2018; Lilienfeld et al., 2009; Wilson & Brekke, 1994). Unfortunately, it also appears that most forensic evaluators do not receive formal training about how various type of cognitive biases can impact FREs (Zapf et al., 2018). Finally, forensic evaluators need guidance about specific *effective* debiasing strategies that they can employ to increase the objectivity and validity of their conclusions (MacLean et al., 2019; Neal & Brodsky, 2016; Zapf et al., 2018).

Debiasing Strategies

Research related to potential debiasing strategies is substantial in medical, business, and policy applications (Soll et al., 2015). Some progress has also been made in seeking to counteract bias that may occur in the physical forensic sciences (Dror, 2018; Dror et al., 2015; Jeanguenat et al., 2017; Kassin et al., 2013). Yet, despite indications that bias can affect forensic evaluations (e.g., Beckham et al., 1989; Boccaccini et al., 2008; Boccaccini, Chevalier, et al., 2017; Guarnera et al., 2017; Murrie et al., 2013), research regarding forensic mental health evaluators' awareness of various types of cognitive bias and potential strategies to counteract them has only recently emerged (Dror & Murrie, 2018; MacLean et al., 2019; Neal & Brodsky, 2016; Neal & Grisso, 2014a; Zapf et al., 2018; Zapf & Dror, 2017).

In the first study to examine forensic mental health evaluators' ideas for potential debiasing strategies, Neal and Brodsky (2016) conducted interviews with 20 forensic psychologists certified by the American Board of Forensic Psychology. The interview prompts were designed to obtain information about the psychologists' awareness of the potential for bias in forensic evaluations and different strategies they believed could minimize bias in their own work. Qualitative analysis of the interviews revealed 25 strategies the participants believed may be useful to mitigate bias in forensic evaluations (see Table 2.4 for a complete list). In the second stage of their study, Neal and Brodsky asked 351 members of the American Psychological Association to rate the 25 strategies on their usefulness as potential bias correction measures. Overall, participants in the second stage of the study rated 22 of the strategies as useful or very useful. We sought to add to these findings by eliciting Dutch forensic evaluators' beliefs about potential sources of bias and the effectiveness of various strategies to counteract it in FREs (Neal & Brodsky, 2016; Zapf et al., 2018).

The Current Study

In this study, we surveyed Dutch forensic mental health evaluators regarding their use and perceived effectiveness of SRAIs for evaluating violent, sexual, criminal, and intimate partner violence risk. Based on previous research conducted by Zapf et al. (2018), we also measured Dutch evaluators' awareness about the potential for bias in forensic risk evaluations. This study further adds to the previous research conducted by Neal and Brodsky (2016) by examining evaluators' perceptions of the efficacy of various debiasing strategies to reduce the potential for cognitive bias in FREs.

Method

Participants

We aimed to obtain a representative sample of Dutch mental health practitioners who conduct forensic risk evaluations. To be eligible to participate, respondents had to be a mental health professional who works in the Netherlands and conducts risk evaluations. Out of 154 respondents who began the survey, 44 did not complete it (71.4% completion rate). We excluded incomplete surveys from our analyses. Therefore, our sample was comprised of 110 respondents with current or previous experience conducting risk evaluations. About 60.9% of respondents were women and approximately 39.1% men. Most had obtained a Master's degree ($n = 77$; 70.0%) and 25 had a doctoral degree (22.7%). Six respondents held a Bachelor's degree (5.5%), and two had a 2-year degree (1.8%).

Experience in mental health settings among respondents ranged from 0 to 46 years ($M = 18.9$, $SD = 11.8$, $Mdn = 18$). Respondents worked in a variety of settings, including forensic psychiatric hospitals (5.5%), private practice (14.5%), mental health clinics (9.1%), and one person indicated working in a hospital setting. Half of respondents selected the option of "other" for their work environment, with most indicating that they worked in some type of clinical or correctional setting (e.g., forensic outpatient clinic, prison psychiatric center or detention facility), and several who worked solely as a court-appointed evaluator or consultant.

Procedure

This study was reviewed and approved by the Ethics Review Committee for Psychology and Neuroscience at Maastricht University (reference 185_06_11_2017). Data collection took place between late March and early June of 2018. Participants were recruited via advertisements in the newsletter of the *Nederlands Register Gerechtelijk Deskundigen* (NRGD; Dutch Registry of Court Experts) and of the *Expertisecentrum Forensische Psychiatrie* (EFP; Expertise Center for Forensic Psychiatry). The NRGD is a registry of forensic professionals created by statute in 2010. Forensic experts who wish to register with the NRGD are evaluated based on field-specific requirements before they are approved (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). The EFP is a professional organization that facilitates cooperation between researchers and practitioners in the field of forensic psychiatry in the Netherlands (Expertisecentrum Forensische Psychiatrie, n.d.).

We also recruited participants via social media with Facebook posts in groups of former forensic psychology Master's students from Maastricht University, and via personal invitations by e-mail to professional contacts of one of the authors. This same author also shared the survey announcement on their professional LinkedIn page. The survey was also advertised on

KNAPP, which is an online site dedicated to forensic psychiatric care in the Netherlands and permits messages to be posted to members to facilitate collaboration and knowledge-sharing (<https://www.knapp-efp.nl>). In return for completing the study, respondents were offered an opportunity to enter a raffle for a chance to win a voucher in the amount of €50 from an online shopping site. The winner was randomly selected from those who completed the survey and provided their e-mail address.

It is difficult to provide an accurate estimate of the number of potential survey respondents we reached. There were 489 forensic mental health experts listed on the NRGD website at the time we conducted the survey and all of them were sent the NRGD newsletter with the study announcement. To NRGD behavioral experts specializing in either adult or juvenile forensic psychology or psychiatry whose e-mail addresses were available online, we sent a personal invitation by e-mail ($n = 270$) and two follow-up reminders. We have no information about how many potential respondents may have viewed the EFP newsletter online. We estimate approximately 400 people eligible to participate in the survey were reached by Facebook and LinkedIn. Finally, there are approximately 1,880 members on the KNAPP website, although it is unknown how many of them were eligible to participate in the survey. There is also overlap between NRGD experts, professionals who receive the EFP newsletter, the author's professional network, and users of the KNAPP website. We did not have the ability to determine the number of participants obtained from the individual platforms we used for advertisement of the study.

Survey

The survey was conducted online using Qualtrics and was available to respondents in both English and Dutch. We used the cookies-based Qualtrics feature to prevent respondents from taking the survey multiple times. We opted not to provide a back button to prevent respondents from changing their answers after they completed a page.

Respondents were provided with a brief description of the survey and acknowledged their consent to participate. The survey was comprised of five major sections as outlined below.

Demographics, FRE Experience, General Frequency of SRAI Use, Concerns about Errors, and Consulting about FREs

In the first section, we collected demographic information regarding gender, education level, years of experience in mental health settings, and current work environment. We asked respondents to specify how many forensic risk evaluations they had performed over the past two years. Next, respondents were asked to indicate on a 5-point Likert-type scale how

useful they thought SRAIs are in conducting risk evaluations (*1 = not at all, 2 = slightly, 3 = moderately, 4 = very, 5 = extremely*). We also asked how often an SRAI was used when conducting a risk evaluation (*1 = never, 2 = sometimes, 3 = about half the time, 4 = most of the time, 5 = always*).

Next, respondents were asked to rate their concerns about the possibility of false positives and false negatives on two separate 5-point Likert scales (*1 = not at all, 2 = a little, 3 = a moderate amount, 4 = a lot, to 5 = a great deal*). In the survey, we defined a false negative as occurring when an individual is classified as having a low risk to reoffend, when in reality he/she has a high risk. We defined a false positive as occurring when an individual is classified as having a high risk to reoffend, when in reality, he/she has a low risk.

The next two questions related to respondents' frequency of consulting with third parties about FREs (*1 = never to 5 = always*). We defined consulting in the survey as "seeking advice about the evaluation." We also asked respondents to specify with whom they consulted from a list of options (*colleagues, supervisor, other treatment provider(s) of the evaluatee, prison/jail staff, evaluatee's family members, probation/parole officer(s), or other*). Respondents could select multiple options, and if *other* was selected, we asked them to specify in free-text.

Use of Specific SRAIs and Usefulness Ratings

In the subsequent section, respondents were asked which SRAIs they were required by their employer or jurisdiction to use, and which SRAIs they chose to use. We created a list of commonly-used SRAIs in the Netherlands (see Table 2.1 for complete list); respondents could select multiple SRAIs and provide free-text responses. For the SRAIs that respondents indicated they were required or chose to use, they rated the usefulness of each separately on a 5-point Likert scale (*1 = not at all useful to 5 = extremely useful*).

Cognitive Bias Concerns in FREs

We asked respondents about their views regarding cognitive bias in forensic risk evaluations (FREs). The extent to which respondents thought cognitive bias is a problem in FREs was rated on a 5-point Likert scale (*1 = not at all, 2 = a little, 3 = a moderate amount, 4 = a lot, 5 = a great deal*). We defined cognitive bias in the survey as an error in reasoning, evaluating, remembering, or processing information. Respondents then indicated *yes* or *no* in response to questions about whether they had received any specific training about cognitive bias in FREs and whether they had ever been concerned about bias in an FRE conducted by (a) themselves or (b) someone else.

Ratings of Effectiveness of Potential Debiasing Strategies

In the next section, respondents were asked to rate on a 5-point Likert scale the efficacy of each of the 25 debiasing strategies (Neal & Brodsky, 2016) to reduce the potential that cognitive bias will influence an evaluator's judgment about future violence risk (1 = *not effective at all* to 5 = *extremely effective*). In contrast to Neal and Brodsky, we did not give participants an option to provide a rating of *unsure*, but rather we used *moderately effective* as the midpoint of our scale. We were interested in respondents' *beliefs* about how effective the strategies are, not whether the strategies are effective in reality. A response of *unsure* would have made it impossible to distinguish between a respondent who was unsure about whether the debiasing strategy was effective and a respondent who was unsure about their own opinion.

Potential Sources of Bias in FREs

Finally, respondents indicated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*) the extent to which they believed that various situations have the potential to bias an evaluator when conducting an FRE. Seven items were derived from Zapf et al.'s (2018) survey. The statements included situations such as whether an evaluator's prior beliefs affect how they analyze a case or their ultimate opinion about the case, whether making a conscious effort to set aside prior beliefs can reduce the likelihood the evaluator will be influenced by them, and whether evaluators know in advance what conclusion they are expected to reach in a case and whether that affects their conclusion (see Table 2.3 for the complete list).

One difference that should be noted between the current study and Zapf et al.'s study is that we eliminated questions about the effects of irrelevant contextual information (Zapf et al., 2018, p. 5, Table 2.2, #5-7). Despite the fact that research suggests irrelevant contextual information can result in bias and errors in other areas of forensic science (Dror, 2012, 2018; Dror et al., 2006, 2015; Kukucka & Kassin, 2014; Nakhaeizadeh et al., 2014), there is debate about what type of information should be considered irrelevant contextual information in these fields (Curley et al., 2020a, 2020b; Dror & Murrie, 2018; Gardner et al., 2019; Thompson, 2020). Further, we were unable to locate any agreed-upon definition of exactly what constitutes irrelevant contextual information specifically for FREs. Therefore, we opted not to include items based on this concept.

Data Analysis

All analyses were conducted using IBM SPSS software version 25.

Results

Experience Conducting FREs and Consulting about FREs

In the two years preceding the survey, the majority of respondents (59%) conducted 10 or more forensic risk evaluations ($Mdn = 12$; range of 0–250), seven (6.4%) of whom had conducted 100 or more. Forty-six respondents (41.8%) had performed 10 or more presentencing evaluations, and six (5.5%) had conducted 100 or more. Twenty-two respondents (20.0%) had performed at least one inpatient TBS evaluation in the past two years, 11 (10.0%) of whom had performed 10 or more. Fifteen (13.6%) respondents had conducted TBS extension evaluations within the previous two years, eight (7.3%) of whom had done 10 or more of these evaluations.

Nearly all respondents ($n = 104$; 94.5%) said they consulted with others about FREs at least some of the time. Over half (57.3%) said they consulted with others *always* or *most of the time*. More than two-thirds of respondents (69.1%) indicated they consult with colleagues and a substantial minority (41.8%) said they consulted with another treatment provider of the examinee. More than one-third consult with either the examinee's probation/parole officer or prison staff (29.1% and 6.4%, respectively), while less than 10% consult with the examinee's family members (9.1%).

Frequency of SRAI Use and Required and Optional Use of Specific SRAIs

Almost all respondents indicated they used an SRAI for risk evaluations *always* or *most of the time* ($n = 107$; 97.3%). Only one respondent indicated that they never used an SRAI when conducting a risk evaluation, one said they did about half the time, and one said they did sometimes. The respondents also mostly agreed that SRAIs are useful in conducting risk evaluations, with 14 (12.7%) rating them as extremely useful and 69 (62.7%) rating them as very useful. However, 19 respondents (17.3%) rated SRAIs as moderately useful, and eight (7.3%) rated them as only slightly useful. Most of our respondents were required to use a specific SRAI, with only about 8% indicating they were not required to do so. Five respondents chose not to use any SRAIs; however, it is unknown whether they chose not to use any specific SRAIs because they used only those that were required.

Many respondents indicated they were both required and choose to use the same SRAIs. In other words, these respondents were in the required and optional users group and provided duplicate ratings for the same SRAI. We therefore created two conceptually distinct groups: *required users* and *optional users*. If respondents indicated they were required, or both required and chose to use the SRAI, we considered them required users. If respondents indicated only that they chose to use the SRAI, we considered them optional users (see Tables 2.1 and 2.2).

The Historical Clinical Risk Management-20 (HCR-20V2; Webster et al., 1997; HCR-20V3; Douglas, Hart, et al., 2014) and the Historische, Klinische, Toekomstige-30 (HKT-30; Werkgroep Pilotstudy Risicotaxie, 2002) or its successor, the HKT-R (Spreen et al., 2014), were the two most commonly reported required SRAIs (48.2% and 47.3%, respectively). The HCR-20 is an SPJ tool designed to assess psychosocial functioning as it relates to violence risk among adults and includes 10 historical, five clinical, and five risk management items (Douglas, Hart, et al., 2014). The HKT-30 is a Dutch-language SPJ tool that was created in the Netherlands for the evaluation of violence risk in forensic psychiatric settings. The HKT-30 has a structure similar to the HCR-20, and consists of 11 historical, 13 clinical and dynamic, and six risk management items rated on a 5-point scale (Werkgroep Pilotstudy Risicotaxie, 2002).

Similarly, nearly half (47.3%) of our respondents were required to use the Structured Assessment of PROtective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robb  , 2009; 2012). The SAPROF is designed to be used in conjunction with an SPJ SRAI and comprises 17 factors considered to be protective against violent behavior (de Vries Robb   et al., 2013). Each protective factor is rated on a three-point scale (0 = *clearly absent*, 1 = *somewhat present*, 2 = *clearly present*). An Integrative Final Risk Judgment is obtained by integrating information about the protective factors assessed with the SAPROF and the risk factors measured with an SPJ risk assessment tool (de Vries Robb   et al., 2013).

Nearly half of respondents (44.5%) were required to use the Psychopathy Checklist-Revised (PCL-R; Hart, 2003) or the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995). The PCL-R and PCL:SV are psychometric tools to assess an individual's level of psychopathy. Although the PCL-R and PCL:SV are not violence risk assessment instruments *per se*, psychopathic traits are linked with an increased risk of violent recidivism (Barbaree et al., 2001; Hawes et al., 2013; Lanterman et al., 2014). Therefore, the PCL-R or PCL:SV is commonly used in combination with SRAIs to assess the risk of violent or sexual reoffending.

To estimate sexual reoffending risk, more than half of respondents (53.6%) were required to use the Static-99 (Hanson & Thornton, 1999) or the Static-99R (Helmus et al., 2012). The Static-99R is an actuarial tool comprised of 10 static, historical risk factors. Although not specifically listed as a choice, two respondents reported they were required to use an SPJ tool, the *Sexual Violence Risk-20* (SVR-20; Boer, Hart, Kropp, & Webster, 1997) to assess sexual reoffending risk. The SVR-20 consists of 20 items related to three domains: psychosocial adjustment, sexual offenses, and future plans (Rettenberger et al., 2017).

A number of participants indicated by free-text response that they used a combination of the Static-99R, the STABLE-2007 (Hanson et al., 2007), and the ACUTE-2007 (Hanson et

al., 2007) to evaluate the risk of sexual reoffending. Some were required to use the STABLE (Hanson et al., 2007) or the ACUTE (Hanson et al., 2007). The STABLE-2007 is used in sexual offending recidivism risk assessment and contains 13 risk factors related to the offender's ability to regulate his sexual behavior (Hanson et al., 2015). The ACUTE-2007 is a measure of dynamic risk factors and contains seven items related to sexual reoffending risk that can change over the short-term (Hanson et al., 2007).

Table 2.1 Risk Assessment Instruments: Required and Optional Users

Risk assessment instrument	Instrument type	Prescribed use	Required users n (%)	Optional users n (%)
<i>Static-99</i> (Hanson & Thornton, 1999) or <i>Static-99R</i> (Helmus et al., 2012)	Actuarial	Sexual reoffending risk	59 (53.6%)	10 (9.1%)
<i>Historical Clinical Risk Management-20</i> (HCR-20V2; Webster et al., 1997; HCR-20V3; Douglas, Hart, et al., 2014)	SPJ	Violence risk	53 (48.2%)	14 (12.7%)
<i>Historische, Klinische, Toekomstige-30</i> (HKT-30; Werkgroep Pilotstudy Risicotaxie, 2002) or HKT-R (Spreen et al., 2014)	SPJ	Violence risk	52 (47.3%)	7 (6.4%)
<i>Structured Assessment of PROtective Factors</i> (SAPROF; de Vogel et al., 2009, 2012)	SPJ	Protective factors	52 (47.3%)	22 (20.0%)
<i>Psychopathy Checklist Revised</i> (PCL-R; Hare, 2003) or the <i>Psychopathy Checklist: Screening Version</i> (PCL:SV; Hart et al., 1995)	Actuarial	Psychopathic traits	49 (44.5%)	18 (16.4%)
OTHER (Free text entry)	N/A	N/A	39 (35.5%)	43 (39.1%)
<i>Structured Assessment of Violence Risk in Youth</i> (SAVRY; Borum et al., 2002)	SPJ	Youth (12-18 years) violence risk	18 (16.4%)	5 (4.5%)
<i>Short-Term Assessment of Risk and Treatability</i> (START; Webster et al., 2004)	SPJ	Violence risk	10 (9.1%)	2 (1.8%)
None	N/A	N/A	9 (8.2%)	5 (4.5%)

Risk assessment instrument	Instrument type	Prescribed use	Required users n (%)	Optional users n (%)
<i>Brief Spousal Assault Form for the Evaluation of Risk</i> (B-SAFER; Kropp et al., 2005)	Actuarial	Spousal assault risk	8 (7.3%)	2 (1.8%)
<i>Violence Risk Appraisal Guide</i> (VRAG; G. T. Harris et al., 1993)	Actuarial	Violence risk	1 (0.9%)	1 (0.9%)
<i>Recidive Inschattingsschalen [Recidivism Assessment Scales]</i> (RISC; Adviesbureau Van Montfoort & Reklassering Nederland, 2004)	Actuarial	General risk of antisocial behavior, criminogenic needs	1 (0.9%)	1 (0.9%)
<i>Violence Risk Screening-10</i> (V-RISK-10; Hartvig et al., 2007)	Actuarial	Violence risk screening	0 (0.0%)	0 (0.0%)
<i>Violence Risk Scale</i> (VRS; Wong & Gordon, 2006)	Actuarial	Violence risk	0 (0.0%)	1 (0.9%)
<i>Classification of Violence Risk</i> (COVR; Monahan et al., 2006)	Actuarial	Violence risk in community	0 (0.0%)	0 (0.0%)
<i>Forensic Operationalized Therapy/Risk Evaluation System</i> (FOTRES; Urbaniok, 2007)	SPJ	Recidivism risk and treatment progress	0 (0.0%)	0 (0.0%)
<i>Sex Offender Risk Appraisal Guide</i> (SORAG; Quinsey et al., 1998a)	Actuarial	Sexual reoffending risk	0 (0.0%)	0 (0.0%)

Note. SRAIs are arranged in descending order of number of required users.

Perceived Usefulness of Most Frequently Used SRAIs

An independent samples *t*-test revealed that the only significant difference between required and optional users with respect to their usefulness ratings of the five most commonly used SRAIs (i.e., HCR-20, HKT-30/R, SAPROF, PCL-R/SV, and Static-99/99R) were found for the Static-99/99R. The optional users rated the Static-99/99R as significantly less useful ($M = 3.90$, $SD = 0.32$) than required users ($M = 4.31$, $SD = 0.75$), $t(30.02) = 2.90$, $p = .007$, Hedges' $g = 0.58$ (medium effect), 95% CI [-0.10, 1.26]. Mean usefulness ratings of the five most commonly used SRAIs for required and optional users are presented in Table 2.2, as are the results of the independent samples *t*-tests.

Table 2.2 Group Mean and Differences in Usefulness Ratings of Most Frequently Used SRAIs

Risk assessment instrument	Required users M (SD)	Optional users M (SD)	t	p	Hedges' g [95% CI]
HCR-20	3.98 (0.72)	3.93 (0.83)	$t(65) = 0.24$.815	0.07 [-0.52, 0.66]
HKT-30/R	3.92 (0.88)	3.57 (0.98)	$t(57) = 0.98$.332	0.39 [-0.40, 1.18]
SAPROF	3.71 (0.85)	3.91 (0.75)	$t(72) = -0.95$.347	0.24 [-0.26, 0.74]
PCL-R/SV	3.88 (1.03)	3.89 (0.83)	$t(65) = -0.04$.967	0.01 [-0.64, 0.66]
Static-99/99R ^a	4.31 (0.75)	3.90 (0.32)	$t(30.02) = 2.90$.007	0.57 [-0.10, 1.26]

Note. SRAIs were rated on a Likert scale from 1 = *not at all useful* to 5 = *extremely useful*.

^a Welch's *t*-test (unequal variance between groups).

Concerns about Cognitive Bias and Errors in FREs

Overall, respondents rated cognitive bias as a *moderate* problem in FREs in general ($M = 3.32$, $SD = 0.75$). A McNemar's test revealed no significant difference in the proportion of respondents who were concerned about bias in an FRE conducted by someone else compared to those concerned about bias in an FRE conducted by themselves, $p = .388$. A paired samples *t*-test revealed that respondents were significantly more concerned about the possibility of a false negative ($M = 3.07$, $SD = 1.00$) than a false positive ($M = 2.85$, $SD = 0.81$) in a risk evaluation, $t(109) = -2.38$, $p = .019$, Hedges' $g_{avg} = 0.31$ (small effect), 95% CI [0.05, 0.57].

Potential Sources of Bias in FREs

Similar to the approach of Zapf and colleagues (2017), we asked respondents about situations that may have the potential to bias evaluators when they are conducting risk evaluations. Mean ratings, response frequencies, and modal responses among our respondents for each level of agreement regarding potential sources of bias in forensic evaluations are presented in Table 2.3.

A substantial majority (83.6%) of our respondents did not agree that cognitive bias is *less* of a problem in forensic psychology than in other forensic sciences (e.g., fingerprint analysis, hair matching). They generally agreed (57.2%) that evaluators sometimes know what conclusions they are expected to reach and that this affects their conclusions (61.8%). Most of our respondents agreed that an evaluator's prior beliefs and expectations can affect how they analyze a case (90.9%) or their ultimate opinion on the case (90.0%). Yet, a substantial majority (71.8%) agreed that if an evaluator makes a conscious effort to set aside their prior beliefs and expectations, they are less likely to be influenced by them.

Table 2.3 Mean Ratings and Frequencies of Beliefs about Potential Sources of Cognitive Bias in Forensic Risk Evaluations

Item	<i>M</i> (<i>SD</i>)	Strongly	Somewhat	Neither	Somewhat	Strongly
		disagree <i>n</i> (%)	disagree <i>n</i> (%)	agree nor disagree <i>n</i> (%)	agree <i>n</i> (%)	agree <i>n</i> (%)
An evaluator's prior beliefs and expectations can affect how s/he goes about analyzing a forensic case	4.15 (0.74)	1 (0.9%)	4 (3.6%)	5 (4.5%)	67 (60.9%)	33 (30.0%)
An evaluator's prior beliefs and expectations can affect his/her ultimate opinion about a forensic case	4.12 (0.71)	1 (0.9%)	3 (2.7%)	7 (6.4%)	70 (63.6%)	29 (26.4%)
An evaluator who makes a conscious effort to set aside his or her prior beliefs and expectations is less likely to be influenced by them	3.75 (0.85)	4 (3.6%)	3 (2.7%)	24 (21.8%)	65 (59.1%)	14 (12.7%)
When evaluators know what they are expected to find, it affects the conclusions they reach	3.53 (1.01)	6 (5.5%)	11 (10.0%)	25 (22.7%)	55 (50.0%)	13 (11.8%)
Evaluators sometimes know what conclusion they are expected to reach	3.49 (1.05)	7 (6.4%)	10 (9.1%)	30 (27.3%)	48 (43.6%)	15 (13.6%)
An experienced evaluator is less likely than a new evaluator to be influenced by prior beliefs/expectations	2.38 (1.09)	27 (24.5%)	37 (33.6%)	25 (22.7%)	19 (17.3%)	2 (1.8%)
Cognitive bias is generally less of a problem in forensic psychology than in other forensic sciences (e.g., fingerprint analysis; ballistics; hair matching, etc.)	1.69 (0.87)	57 (51.8%)	35 (31.8%)	14 (12.7%)	3 (2.7%)	1 (0.9%)

Note. Items are presented according to mean ratings in descending order. Modal responses are shown in bold.

Perceived Effectiveness of Debiasing Strategies

We asked our respondents to rate how effective they thought the 25 strategies identified in Neal and Brodsky (2016) are in reducing the potential for cognitive bias in a risk evaluation. The mean effectiveness ratings for all debiasing strategies are presented in Table 2.4.

Table 2.4 Means and Frequencies of Effectiveness Ratings for Debiasing Strategies

Strategy	<i>M</i> (<i>SD</i>)	Effectiveness Ratings <i>n</i> (%)				
		Not at all	Slightly	Moderately	Very	Extremely
<i>Using structured evaluation methods</i>	4.00 (0.89)	1 (0.9%)	5 (4.5%)	22 (20.0%)	47 (42.7%)	35 (31.8%)
Taking personal responsibility to continue learning after completing formal training and education	3.82 (0.79)	0 (0.0%)	6 (5.5%)	28 (25.5%)	56 (50.9%)	20 (18.2%)
Investigating all relevant data before forming an opinion ^a	3.75 (0.82)	1 (0.9%)	6 (5.5%)	30 (27.3%)	56 (50.9%)	17 (15.5%)
Basing conclusions and opinions on sound data ^b	3.74 (0.82)	0 (0.0%)	9 (8.2%)	28 (25.5%)	56 (50.9%)	17 (15.5%)
<i>Consulting with colleagues about issues of potential bias</i>	3.73 (0.80)	1 (0.9%)	6 (5.5%)	30 (27.3%)	58 (52.7%)	15 (13.6%)
Avoiding advocacy	3.50 (0.97)	1 (0.9%)	18 (16.4%)	33 (30.0%)	41 (37.3%)	17 (15.5%)
Fostering a continuing commitment to objectivity	3.45 (0.88)	1 (0.9%)	15 (13.6%)	39 (35.5%)	44 (40.0%)	11 (10.0%)
Being an active consumer of scientific knowledge	3.45 (0.96)	4 (3.6%)	15 (13.6%)	29 (26.4%)	52 (47.3%)	10 (9.1%)
<i>Critically examining conclusions (e.g., considering alternative hypotheses)</i>	3.38 (0.82)	2 (1.8%)	11 (10.0%)	47 (42.7%)	43 (39.1%)	7 (6.4%)
Limiting the scope of the inquiry and report to the referral question	3.36 (0.98)	4 (3.6%)	16 (14.5%)	38 (34.5%)	40 (36.4%)	12 (10.9%)
<i>Receiving explicit didactic training about objectivity</i>	3.32 (0.97)	3 (2.7%)	20 (18.2%)	36 (32.7%)	41 (37.3%)	10 (9.1%)
<i>Taking time to think about evaluation information rather than immediately writing the report</i>	3.29 (0.94)	5 (4.5%)	15 (13.6%)	40 (36.4%)	43 (39.1%)	7 (6.4%)
Resisting allegiance effects	3.27 (.99)	4 (3.6%)	18 (16.4%)	44 (40.0%)	32 (29.1%)	12 (10.9%)

Strategy	<i>M</i> (<i>SD</i>)	Effectiveness Ratings n (%)				
		Not at all	Slightly	Moderately	Very	Extremely
Continuous introspection about personal biases	3.26 (0.81)	1 (0.9%)	17 (15.5%)	49 (44.5%)	38 (34.5%)	5 (4.5%)
Attending to wording choice in reports to edit out value-laden language	3.19 (0.92)	4 (3.6%)	19 (17.3%)	46 (41.8%)	34 (30.9%)	7 (6.4%)
<i>Exposure to the importance of objectivity through reading professional literature</i>	3.12 (0.99)	6 (5.5%)	23 (20.9%)	40 (36.4%)	34 (30.9%)	7 (6.4%)
Taking careful notes during an evaluation ^c	3.11 (1.04)	8 (7.3%)	21 (19.1%)	40 (36.4%)	33 (30.0%)	8 (7.3%)
Disengaging emotionally from cases	3.07 (0.92)	5 (4.5%)	23 (20.9%)	45 (40.9%)	33 (30.0%)	4 (3.6%)
Examining patterns of personal decision-making (e.g., agreement with referral party preferences) ^d	3.05 (0.96)	8 (7.3%)	18 (16.4%)	49 (44.5%)	30 (27.3%)	5 (4.5%)
Restricting conclusions and opinions to scientific foundations	2.89 (1.02)	11 (10.0%)	24 (21.8%)	47 (42.7%)	22 (20.0%)	6 (5.5%)
Intentionally controlling existing bias	2.83 (0.92)	9 (8.2%)	26 (23.6%)	54 (49.1%)	17 (15.5%)	4 (3.6%)
Observing others who manage their personal biases successfully	2.60 (1.02)	18 (16.4%)	32 (29.1%)	38 (34.5%)	20 (18.2%)	2 (1.8%)
Developing a sense of pride in one's professional identity	2.45 (1.20)	31 (28.2%)	26 (23.6%)	31 (28.2%)	16 (14.5%)	6 (5.5%)
Accepting referrals only for cases in which bias is unlikely	2.14 (0.97)	35 (31.8%)	35 (31.8%)	30 (27.3%)	10 (9.1%)	0 (0.0%)
Limiting empathy and rapport in forensic cases	1.87 (1.06)	56 (50.9%)	23 (20.9%)	22 (20.0%)	7 (6.4%)	2 (1.8%)

Note. *N* = 110. Items are presented according to mean ratings arranged in descending order. Modal effectiveness ratings are shown in bold. Debiasing strategies identified by Neal and Brodsky (2016) as appearing as suggestions in scientific literature are in italics.

^{a, b, c, d} Additional strategies we identified as suggested in the literature about FREs.

We calculated the sample mean rating across all six strategies that have been suggested in the literature as potentially effective in mitigating bias, and the sample mean rating for the 19 strategies identified by Neal and Brodsky (2016) as ineffective or not specifically suggested previously. Only introspection has been shown to be ineffective, because people are often unaware of the existence of a particular biasing stimulus, that their response is influenced by the stimulus, or both (Nisbett & Wilson, 1977a; Pronin & Kugler, 2007; Wilson & Brekke, 1994).

Strategies previously suggested in scientific literature include the following: critically examining conclusions, consulting with colleagues, taking time to think about the evaluation information rather than immediately writing the report, receiving explicit didactic training about objectivity, exposure to the importance of objectivity through reading professional literature, and using structured evaluation methods. A paired samples *t*-test revealed that respondents gave significantly higher ratings to strategies that have been suggested in the literature ($M = 3.47$, $SD = 0.57$) than to those identified by Neal and Brodsky as not having been specifically suggested previously ($M = 3.09$, $SD = 0.59$), $t(109) = 10.51$, 95% CI [0.31, 0.45], $p < .001$, Hedges' $g_{\text{avg}} = 0.65$ (medium effect), 95% CI [0.50, 0.80].

Effects of Cognitive Bias Training

A minority of respondents (27%) indicated that they had received training on cognitive bias. An independent samples *t*-test revealed no significant difference between respondents who had received bias training ($M = 3.37$, $SD = 0.77$) and those who had not ($M = 3.30$, $SD = 0.75$) with regard to the extent to which they believed cognitive bias is a problem in FREs, $t(108) = -0.41$, Hedges' $g = 0.09$, 95% CI [-0.33, 0.51]. Overall, this finding indicates respondents think cognitive bias is a moderate problem in FREs, regardless of whether they have had cognitive bias training or not.

We also examined whether training about cognitive bias in forensic evaluations affected ratings of the debiasing strategies. A Welch's *t*-test revealed that practitioners who had some training related to bias gave significantly higher effectiveness ratings ($M = 4.10$, $SD = 0.61$) to "taking personal responsibility to continue learning after completing formal training and education" as a debiasing strategy than those who had not received such training ($M = 3.71$, $SD = 0.83$), $t(71.02) = -2.68$, $p = .009$, Hedges' $g = 0.50$ (medium effect), 95% CI [0.08, 0.93]. We are unaware of research that tests whether this strategy is effective in counteracting cognitive bias.

Discussion

Risk evaluation practices and use of SRAIs may vary by country, by type of legal system (i.e., adversarial vs. inquisitorial), and other system factors, such as recommendations by professional organizations, organizational or statutory requirements to use (specific) SRAIs, and regulations governing recognition as an expert in the legal system (McLaughlin & Kan, 2014). In addition, awareness and education about the potential for cognitive bias in forensic risk evaluations and potentially useful debiasing strategies may also vary between countries. Therefore, the purpose of this survey was threefold: (a) to identify which SRAIs are commonly used by forensic mental health evaluators who conduct FREs in the Netherlands and their perceptions of the usefulness of those SRAIs; (b) to gain insight into evaluators' views about potential sources of bias when conducting FREs; and (c) to examine evaluators' concerns about cognitive bias and their views of potential strategies to diminish bias.

Frequency of Use and Usefulness Ratings of SRAIs

Similar to previous surveys (Hurducas et al., 2014; Neal & Brodsky, 2016; Zapf et al., 2018), respondents in the current study were quite experienced in their respective fields and averaged approximately 18 years of experience in mental health settings. The majority of our respondents had conducted more than 10 risk assessments in the two years preceding our survey, with a small proportion having conducted 100 or more.

Our respondents reported a very high rate of SRAI use, as nearly all of them indicated that they use an SRAI always or most of the time. Further, the vast majority of respondents were required to use a specific SRAI. These high rates of SRAI use likely reflect professional recommendations in the Netherlands that urge the use of SRAIs in forensic assessments (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). Our results are in line with findings in other recent international surveys suggesting the increasing use of SRAIs (Neal & Grisso, 2014b; Singh et al., 2014). For example, Neal and Grisso (2014b) reported that 96.9% of their respondents reported using an SRAI for sexual offender risk assessment and 89.0% for violence risk assessment. Similarly, among an international sample of mental health professionals from Europe, Singh and colleagues (2014) reported that over the 12 months preceding the survey, approximately 63% of respondents used SRAIs to conduct risk assessments.

In addition, in the Netherlands, a strong emphasis is placed on using an SPJ approach to FREs (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). Our findings indicate that SPJ tools—specifically, the HCR-20 and the HKT-30/R—are the SRAIs most commonly used among Dutch forensic evaluators. In fact, most of our respondents reported

that they are required to use one of these two tools. Regardless of whether they were required or optional users, evaluators rated the HCR-20 and the HKT-30/R as moderately to highly useful. About two-thirds of our respondents also used the SAPROF and rated it as moderately to very useful.

It is common practice for forensic evaluators to use a psychopathy measure when performing a (sexual) violence risk assessment. Psychopathy is generally considered relevant in psycholegal contexts because of its relationship with risk for general criminal, violent, and sexual recidivism (DeMatteo et al., 2020). Nearly half of our respondents reported that they were required to use a psychopathy measure (PCL-R or PCL:SV) and a substantial majority of these required users rated the measure as very or extremely useful. Similarly, in their survey of forensic psychologists, Viljoen and colleagues (2010) reported that nearly 65% of respondents who conducted any type of adult risk assessment used the PCL-R or PCL:SV.

Interestingly, the majority of our respondents indicated they were required or chose to use an actuarial (rather than SPJ) SRAI—the Static-99/99R—for evaluating the risk of sexual reoffending. This is not entirely unexpected as the Static-99/R is one of the most commonly used SRAIs for estimating the likelihood of sexual recidivism (Archer et al., 2006; Chevalier et al., 2015; Kelley et al., 2020; Neal & Grisso, 2014b). Furthermore, previous surveys have reported that actuarial SRAIs are used more often than SPJ SRAIs for evaluating sexual recidivism risk (Kelley et al., 2020; Neal & Grisso, 2014b).

Despite its widespread use and popularity, recent field studies indicate that the Static-99R can result in significant overestimates of sexual recidivism risk (Boccaccini, Rice, et al., 2017). The Static-99R has also been criticized because it considers only historical risk factors that cannot be changed with treatment (Cauley, 2007; Craig et al., 2005). The Risk-Needs-Responsivity (RNR; Bonta & Andrews, 2007) model of rehabilitation for people who have committed criminal offenses indicates that addressing treatment needs requires the assessment of dynamic factors that can be targets for treatment (Bonta & Andrews, 2007; Mann et al., 2010). In fact, some respondents in our survey wrote a free-text response that said they used the Static-99/99R in combination with the ACUTE-2007/STABLE-2007 to measure static and dynamic factors and treatment needs. Some respondents also wrote that they use the Sexual Violence Risk-20 (SVR-20; Boer et al., 1997), which is the most commonly used SPJ instrument for evaluating sexual recidivism risk.

Concerns about Bias and Errors in Forensic Risk Evaluations

Respondents to our survey were significantly more concerned about the possibility of a false negative than a false positive outcome in their risk evaluations. This finding suggests that Dutch forensic evaluators, like evaluators elsewhere (Bonta & Motiuk, 1990), may tend to err

on the side of caution (i.e., being more averse to the potential outcome of improperly classifying someone as low risk who then reoffends than improperly classifying someone as high risk who would not have reoffended). Furthermore, erring on the side of caution can increase the likelihood of false positives, the consequence of which is that people are unjustly deprived of their freedom and limited mental health and correctional resources are needlessly wasted (Bonta & Motiuk, 1990; P. M. Harris, 2006). Of course, false positives and false negatives both carry harmful consequences and neither error is desirable. Yet it is much easier to identify a false negative because the crime the person commits is likely to come to the attention of law enforcement, and potentially, the media. On the other hand, a false positive is unlikely to be identified because the absence of reoffending is likely to be attributed to incapacitation and treatment, not a misclassification of the individual's risk.

We were also interested in how often and with whom our respondents consult regarding FREs. We defined consulting for our respondents as “seeking advice about the evaluation,” thereby leaving open the possibility of accounting for collateral interviews. Previous studies indicate that risk judgments can be more accurate than individual evaluator ratings by using a consensus method (de Vogel & de Ruiter, 2006; Huss & Zeiss, 2004; McNeil et al., 2000), and this method is commonly used in the Netherlands when conducting FREs (Harte & Breukink, 2010). Our findings confirm that it is common practice for Dutch evaluators to seek advice from others about an FRE *always or most of the time*. In addition, over 40% of our respondents indicated that they consult with other treatment providers of the evaluatee, who is the subject of the FRE.

Beyond consultations with colleagues, we also note that our respondents indicated they “consult” with other parties about FREs, including probation and parole officers, the evaluatee’s family, and prison staff. Because of how we defined “consult” in our survey, the responses elicited from our respondents likely include inter-professional consultations and collateral interviews. In fact, collateral sources of information appear fairly common in FREs. Neal and Grisso (2014b) reported that their international respondents ($N = 434$) conducted collateral interviews, both with other professionals (54.5% for FREs for violence and 25.0% for FREs for sexual violence) and with non-professionals (27.0% and 35.2%, respectively).

Guidelines from The Netherlands Institute of Forensic Psychiatry and Psychology indicate forensic evaluators should include relevant information in their reports, but the decision about relevant and irrelevant information is left to the subjective opinion of the evaluator (de Ruiter & Kaser-Boyd, 2015). On the one hand, collateral interviews may provide evaluators with important information needed to conduct a complete FRE. On the other hand, there is a concomitant risk that the evaluator will be exposed to potentially biasing, irrelevant contextual information (Zapf & Dror, 2017). There is a growing body of empirical studies

providing evidence that exposure to task-irrelevant contextual information can bias forensic evaluators (Dror, 2012, 2018; Dror et al., 2006, 2015; Kukucka & Kassin, 2014; Nakhaeizadeh et al., 2014) and should be avoided. In fact, avoiding potentially biasing, irrelevant information may be one of the most effective strategies to prevent bias before it occurs (Dror et al., 2015; Gardner et al., 2019; National Research Council, 2009; Wilson & Brekke, 1994).

Potential Sources of Bias in Forensic Evaluations

Similar to Zapf et al.'s (2018) findings, our respondents were inclined to agree that an evaluator's prior beliefs and expectations can affect how they analyze a forensic case and formulate their ultimate opinion, suggesting some awareness of confirmation bias among our respondents (Kassin et al., 2013). Nevertheless, less than half of our respondents viewed bias as a problem in FREs. This contrasts with nearly 86% of forensic evaluators in Zapf et al.'s survey who expressed concern about cognitive bias in forensic evaluations as a whole, and about 79% of whom said they were concerned about cognitive bias in their specific domain of forensic evaluations.

Furthermore, a substantial majority of our respondents agreed that a conscious effort to set aside prior beliefs or expectations makes it less likely an evaluator will be influenced by them. Yet, conscious efforts to set aside prior beliefs and expectations are unlikely to be effective in eliminating bias, because bias operates outside of awareness (Wilson & Brekke, 1994). Furthermore, even when an evaluator is aware of the potential for preexisting motivations and emotions to affect their evaluation, efforts to counteract them are not necessarily effective (Kassin et al., 2013).

Perhaps our respondents were not as concerned about bias in FREs because they commonly used SRAIs (de Ruiter, 2016; de Ruiter & Hildebrand, 2007). As applied to FREs, SRAIs may help reduce the effects of bias, although a growing body of evidence suggests this is not always the case (Chappell et al., 2013; Gowensmith & McCallum, 2019; Guay & Parent, 2018; Guy et al., 2014; J. Miller & Maloney, 2013; Murrie et al., 2008, 2009; Murrie & Balusek, 2008; Schmidt et al., 2016; Shepherd & Sullivan, 2017; Storey et al., 2012; Wormith et al., 2012). Therefore, evaluators should still remain aware of the potential for bias if they are to take steps to effectively minimize its effects on FREs (Dror, 2018; Lilienfeld et al., 2009; Wilson & Brekke, 1994).

Perceived Effectiveness of Debiasing Strategies

Neal and Brodsky (2016) identified six debiasing strategies that have previously been suggested as potentially effective debiasing techniques for use in forensic evaluations. These strategies

are: using structured methods to gather and analyze data (Croskerry et al., 2013; Graber et al., 2012; Neal & Grisso, 2014b; Zapf & Dror, 2017), consulting with colleagues (Croskerry et al., 2013; Graber et al., 2012), critically examining conclusions (e.g., considering alternative hypotheses; Galinsky et al., 2000; Galinsky & Moskowitz, 2000; Grisso, 2010; Lord et al., 1985; Mumma & Wilson, 1995; Soll et al., 2015; Zapf & Dror, 2017), receiving explicit didactic training about objectivity (Bridge & Marić, 2019; Croskerry et al., 2013; Graber et al., 2012; Soll et al., 2015), taking time to think about the evaluation before writing the report (Croskerry et al., 2013; Lilienfeld et al., 2009), and reading professional literature about the importance of objectivity (Croskerry et al., 2013). Our respondents gave significantly higher effectiveness ratings to the six debiasing strategies with some empirical support than they did to the remaining 19 strategies. However, we note that the empirical support for several of these ‘effective’ debiasing strategies is derived from medical research related to reducing errors and improving accuracy in clinical and diagnostic decision-making. Therefore, we cannot say with certainty that the debiasing strategies suggested in other fields will be effective in mitigating bias in FREs (Fischhoff, 1982; Soll et al., 2015).

In addition to the six strategies identified by Neal and Brodsky (2016) as having been suggested in scientific literature, we think there are at least four others (from the 25 listed) that have been suggested as potentially effective in countering bias specific to forensic evaluations. These four strategies are: investigating all relevant data before forming an opinion (Grisso, 2010; Zapf & Dror, 2017), basing conclusions and opinions on sound data (Grisso, 2010; Zapf & Dror, 2017), taking careful notes during an evaluation (Arkes, 1981; Borum et al., 1993; Mumma & Wilson, 1995), and examining patterns of personal decision-making (e.g., agreement with referral party preferences; comparing one’s decisions over time to base rates; Brodsky, 2013; DeClue & Rice, 2016; Gowensmith & McCallum, 2019; Murrie & Balusek, 2008; Murrie & Warren, 2005; Parker, 2016). However, we note that very few strategies to mitigate bias have been empirically tested. Therefore, field and task-specific research regarding the effectiveness of potential strategies to mitigate bias in FREs would be of significant benefit for the discipline, forensic examinees, and legal decision-makers.

Limitations and Conclusions

We note there are a number of limitations to our findings that relate to our sample. First, is the relatively small sample size of our study ($N = 110$), although the sample size is in line with previous similar surveys conducted in other countries (e.g., Archer et al., 2006; Hill & Demetriooff, 2019; Viljoen et al., 2010). Second, practitioners who chose to participate in the survey may have done so because they possess more awareness and concern about the potential

for bias than evaluators who chose not to participate. Third, several of our respondents had not attained a graduate degree, which may have limited the types of risk assessment tools they were qualified or trained to use. Finally, we are aware of one previous international survey by Singh and colleagues (2014) that differentiated between forensic psychologists' and forensic psychiatrists' use of violence SRAIs. Their findings indicate that forensic psychologists may use violence SRAIs more frequently than forensic psychiatrists. However, we did not ask our respondents to specify their discipline, which may further limit our findings. Future researchers may consider differentiating between practitioners in these two disciplines regarding the frequency of use and perceived utility of SRAIs. For these reasons, our sample may not be representative of all forensic mental health practitioners who conduct FREs in the Netherlands.

In addition, the actual response rate to our survey is unknown, and any conclusions about the broader population of forensic mental health practitioners in the Netherlands are therefore tentative. Furthermore, we do not know the context in which our respondents are performing FREs: for example, are they conducting FREs for the court, for the purpose of treatment planning and management, for decisions related to patient restrictions? The context in which an FRE is undertaken may have an effect on SRAI use and how concerns about potential bias are managed. Therefore, our findings may not generalize across all forensic contexts and purposes for which FREs are conducted.

Our survey results are also limited with respect to the use of sexual recidivism SRAIs. Unfortunately, we failed to include SPJ instruments for sexual violence, such as the SVR-20, in our survey list of SRAI options. Still, a number of respondents included the SVR-20 in free text, but we cannot exclude the possibility that more respondents would have selected the SVR-20 if it had been listed as an option.

Finally, we cannot rule out the possibility of socially desirable responding. Although the survey was completed anonymously, and participants were assured of anonymity, it is possible that some people were concerned about negative portrayals of their profession, their work, or their employer. Therefore, it is possible that there is a disparity between what our respondents say they do in practice and what they actually do.

The findings from this study also point to potential areas for future research. For example, we did not ask participants about "irrelevant contextual information" as in Zapf et al.'s (2017) study, because of the lack of consensus about what constitutes irrelevant contextual information in risk evaluations (Dror & Murrie, 2018). A recent survey of forensic scientists ($N = 189$) from several forensic disciplines (biology, pattern evidence, chemistry, and crime scene investigation) is illustrative of the challenges in obtaining agreement regarding what constitutes irrelevant contextual information in their tasks (Gardner et al., 2019). Gardner

and colleagues reported that among crime scene investigators, a substantial majority agreed that a description of the evidence, how it was collected, the type of offense, and a synopsis of the case were essential to their tasks. Conversely, 25% said that a suspect's statement or confession was essential, whereas greater than half said they would review that information if it was available (58.3% and 66.7%, respectively). A small minority agreed that the suspect's statement or confession was irrelevant (16.7% and 8.3%, respectively).

Similar to crime scene investigation, the process of a forensic risk evaluation involves gathering information and evidence. Information that is deemed to be relevant in that process is likely to vary considerably between cases and individual experts. Given the potential for irrelevant contextual information to introduce bias into FREs (Neal & Saks, 2016; Zapf & Dror, 2017), surveying forensic mental health professionals about their views of what information is irrelevant to the FRE task seems to be an important avenue for future research. Only within the past decade or so have researchers begun to investigate the role that cognitive bias may play in FREs (Camilleri et al., 2019; Charman, 2013). Furthermore, many potentially effective debiasing techniques have not been empirically tested with respect to forensic evaluations, or specifically FREs. This point is worth noting because not all debiasing strategies are appropriate and/or effective across domains or tasks (Soll et al., 2015). It is not necessarily the case that a debiasing strategy that is effective in countering one type of bias (e.g., confirmation bias) will be effective in countering a different type of bias (Fischhoff, 1982). Therefore, direct investigations on the effect structured evaluation methods on countering bias in FREs may shed light on the debiasing potential of these methods.

Chapter 3

Negative Pretrial Publicity, Evaluator Attitudes toward Offenders, and Risk Assessment with the HCR-20^{V3}

This chapter is based on the following paper:

Kamorowski, J., de Ruiter, C., Schreuder, M., Jelicic, M., & Ask, K. (2020). Negative pretrial publicity, evaluator attitudes toward offenders, and risk assessment with the HCR-20^{V3}.

Manuscript in preparation.

Abstract

Structured risk assessment instruments (SRAIs), such as the Historical Clinical Risk Management (HCR-20^{v3}) are increasingly used to inform criminal justice decision-making, highlighting the significance of examining the potential for bias when using these measures and effective strategies to mitigate it. In this experimental study, we examined the possible biasing effects of (1) negative pretrial publicity (PTP) about a person who committed a double homicide and (2) evaluators' attitudes toward offenders, on scale scores and final risk judgments of the HCR-20^{v3}. Participants ($N = 54$) included graduate students, clinicians, and researchers who had all been trained to use the HCR-20. Contrary to expectation, negative PTP showed no effects on the HCR-20 total scores, subscale scores, or final risk judgments. In line with our hypothesis, more positive attitudes toward offenders were predictive of lower HCR-20 total scores and lower ratings on the Clinical and Risk Management subscales and final risk judgment of imminent violence. These findings add to a growing body of research indicating forensic risk evaluations conducted using SRAIs are not immune to the effects of some types of bias.

Introduction

The risk of violent reoffending is a primary concern for decision-makers in the criminal justice system. Structured risk assessment instruments (SRAIs) are increasingly used to inform an array of criminal justice practices, including decisions related to detention vs. noncustodial sanctions, and release from forensic institutions in numerous countries (Desmarais & Singh, 2013; Fazel et al., 2012; Singh et al., 2011). For example, the risk of future violence is a key factor in parole decisions in the United States (Guy, Kusaj, et al., 2015; Rhine et al., 2018), and most states that retain parole rely on SRAIs to evaluate the risk of violent reoffending (Ruhland et al., 2016).

SRAIs have been demonstrated to be more transparent and accurate than risk assessments based on unstructured clinical judgment (Ægisdóttir et al., 2006; de Vogel et al., 2004; Hanson & Morton-Bourgon, 2009; Hilton et al., 2006). Unstructured clinical judgment is a method of risk assessment in which a mental health professional does not utilize an SRAI to evaluate an individual's risk of future violence (Murray & Thomson, 2010), but relies upon their own clinical experience and/or other measures that are not specifically designed to evaluate violence risk (e.g., personality measures, cognitive ability testing). Research suggests that clinicians who use unstructured clinical judgment in violence risk assessment vary widely in terms of the factors they consider important in making a judgment of future violence risk (e.g., B. Brown & Rakow, 2016; Huls et al., 2018). Moreover, unstructured clinical judgment has been criticized as being vulnerable to evaluator bias (e.g., Borum et al., 1993; Elbogen, 2002, 2016; Quinsey et al., 1998a) and exhibiting poor predictive validity (Hart et al., 2007; Levinson & Ramsay, 1979; Monahan, 1981; Mulvey & Lidz, 1985).

Although unstructured clinical judgment continues to be used in practice (Singh et al., 2014) and accepted in U.S. courts (Guy, Douglas, et al., 2015), forensic risk evaluators often employ SRAIs when they conduct this type of assessment (Desmarais & Singh, 2013; Neal & Grisso, 2014b; Singh et al., 2014). *Actuarial* SRAIs are based on numerical scoring of weighted risk factors that have been shown to be associated with recidivism risk (Dawes et al., 1989; Doyle & Dolan, 2007; Quinsey et al., 1998b). *Structured professional judgment* (SPJ) SRAIs allow an evaluator to combine professional judgment with risk factors that are empirically and/or clinically related to recidivism risk and to make decisions about the relevance and management of those risk factors with respect to the examinee (Campbell et al., 2009; Douglas & Kropp, 2002; Douglas & Skeem, 2005).

The Historical-Clinical-Risk-20, currently in its third iteration (HCR-20^{V3}: Douglas, Hart, et al., 2014), is the most commonly used SPJ instrument for violence risk assessment worldwide (Archer et al., 2006; Neal & Grisso, 2014b; Singh et al., 2014; Viljoen et al., 2010). The HCR-20^{V3} exhibits a moderate degree of predictive validity (de Vogel et al., 2014; Doyle et

al., 2014; Persson et al., 2017; K. J. Smith et al., 2020; Strub et al., 2014) and good to excellent inter-rater reliability (Douglas, Shaffer, et al., 2014; Douglas & Belfrage, 2014; cf. Tully, 2017). Hence, the HCR-20 is widely used to inform forensic treatment and legal decisions about institutional confinement and release from an institutional setting to the community (e.g., Cox et al., 2018; de Ruiter, 2016; Vitacco et al., 2012).

External and Internal Factors Can Bias Forensic Risk Evaluations

As an ideal, forensic risk evaluation should be conducted with “untainted and unbiased” judgment (Greenberg & Shuman, 2007, p. 131), although it is doubtful this ideal can ever be fully achieved. However, it is important to acknowledge and investigate how various types of cognitive bias can threaten the objectivity of forensic risk evaluations (Neal et al., 2019; Zapf & Dror, 2017) given the role of these evaluations in decision-making about evaluatees’ civil liberties.. SRAIs are believed to mitigate the potential influence of cognitive bias on a risk assessment of violent or sexual recidivism (Carroll, 2007; Kamorowski, de Ruiter, et al., 2021; Shepherd & Sullivan, 2017), specifically by limiting evaluator discretion when considering factors relevant to recidivism risk (Bonta & Andrews, 2007; Hannah-Moffat et al., 2009; P. M. Harris, 2006; Krauss & Sales, 2001; Zapf & Dror, 2017).

Yet, some types of SRAIs, particularly those based on structured professional judgment, permit more flexibility and evaluator discretion in their use, which likely also increases the potential for bias to influence risk judgments (Childs et al., 2014; Grove & Meehl, 1996; Shepherd & Sullivan, 2017). In fact, there is limited research that examines how forensic evaluators use information that is collected during a risk evaluation process to estimate risk when employing an SPJ risk assessment instrument (Schwalbe, 2008). Thus, there is a pressing need for research that examines if certain types of information could bias evaluators’ judgments of risk when employing SPJ instruments.

Potential sources of bias can be external or internal to the evaluator (Gowensmith & McCallum, 2019). An example of *external bias* is the *adversarial allegiance* effect, which is evaluators’ tendency to reach conclusions supportive of the retaining party’s interests (Boccaccini, Chevalier, et al., 2017; Chevalier et al., 2015; Murrie et al., 2009, 2013). This phenomenon has been shown to affect selection of the relevant group norms (e.g., routine sample versus high risk/high needs sample) and forensic reporting practices (Chevalier et al., 2015), as well as SRAI scoring and interpretation (Boccaccini, Chevalier, et al., 2017; Chevalier et al., 2015; Edens et al., 2015; Murrie et al., 2008, 2009, 2013; Murrie & Boccaccini, 2015).

Internal bias includes evaluators’ personality traits (A. K. Miller et al., 2011), experience level (Boccaccini, Rufino, et al., 2017), feelings toward the examinee (de Vogel & de Ruiter,

2004; Vera et al., 2019), emotional responses to case information (Blumenthal et al., 2010; Huls et al., 2018), and preexisting attitudes or beliefs (Deitchman et al., 1991; Everson & Sandoval, 2011; Homant & Kennedy, 1986, 1987b, 1987a; Neal, 2016). For example, Miller and colleagues (2011) examined whether personality characteristics of forensic graduate students and clinical psychology faculty members influenced their ratings of an examinee's level of psychopathy using the Psychopathy Checklist-Revised (PCL-R; Hare, 2003). The PCL-R—although not a risk assessment instrument *per se*—is a measure for evaluating the level of psychopathic traits, which is associated with an increased risk of violent and sexual recidivism (Barbaree et al., 2001; Hawes et al., 2013; Lanterman et al., 2014).

The findings from the Miller et al. (2011) study indicated that evaluators who exhibited more agreeableness in their own personality tended to rate offenders as less psychopathic, whereas evaluators who were more conscientious tended to rate offenders as more psychopathic. Furthermore, other aspects of evaluator personality traits also showed significant correlations with particular facets of the PCL-R (for example, evaluators whose level of extroversion and thrill-seeking was high tended to rate offenders as lower on the Impulsive Lifestyle facet). The Miller study was an exploratory study with a small sample ($N = 22$), and we have not been able to retrieve any subsequent published research examined how specific evaluator characteristics may influence forensic risk evaluations.

Potentially Biasing Effects of Negative Media Exposure on Perceptions of Risk

External sources of bias may present in the form of *irrelevant contextual information*, which includes collateral information that is not relevant to a forensic evaluation of violent recidivism risk (Zapf & Dror, 2017). Zapf and colleagues (2018) note that “[f]orensic psychology has yet to really take on and acknowledge the impact of irrelevant contextual information on bias” (p. 8). One source of irrelevant contextual information is pretrial publicity (PTP), which can take the form of extensive and mostly negative portrayals of the character of a crime suspect (Bakhshay & Haney, 2018; Grabe et al., 2001; Neal & Grisso, 2014a; Yanich, 2004). In fact, media coverage of high-profile crimes, particularly homicides, tends to be overwhelmingly negative toward the suspect or defendant (Bakhshay & Haney, 2018; Grabe et al., 2001; Steblay et al., 1999; Studebaker & Penrod, 1997; Yanich, 2004). Media outlets also commonly employ sensationalized depictions of the crime that use extreme, shocking, or emotional language (e.g., “bloodbath,” “execution style murder”; Bakhshay & Haney, 2018).

The killing of one or both biological parents, so-called parricide, is an exceptionally rare form of homicide (Heide, 2013). However, when these cases occur, they tend to draw widespread and extensive media coverage (Boots & Heide, 2006). Furthermore, the media

tends to sensationalize certain types of parricide cases (e.g., multiple victims, a juvenile perpetrator) that are likely to evoke shock and horror among their audience (Boots & Heide, 2006). However, the risk of violent recidivism among people who commit parricide tends to be very low because the killing tends to be related to intra-familial strife (Bourget et al., 2007; Dantas et al., 2014). Nevertheless, the shocking nature of the crime and sensationalized accounts of it, may evoke negative emotional responses and thereby increase perceptions of the offender's risk of violent reoffending.

Extensive research about the effects of PTP on juror decision-making indicates that negative PTP can increase jurors' perceptions of the defendant's guilt (for reviews, see Ruva, 2018; Steblay et al., 1999). We wanted to examine whether negative PTP about a double-parricide might similarly bias evaluators' perceptions of future violence risk of the perpetrator (Neal & Grisso, 2014a). Specifically, we were interested in whether exposure to negative media about the examinee has the potential to affect forensic evaluators' ratings of factors associated with the risk of violent recidivism.

Evaluator Attitudes Toward Offenders

Previous research suggests that evaluator feelings toward an examinee can affect risk ratings using the HCR-20 v.2. De Vogel and de Ruiter (2004) conducted a study in which a researcher, a treatment supervisor, and a group leader (i.e., a clinician involved with daily activities/programming for the patients) in a Dutch forensic psychiatric hospital coded the HCR-20 v.2 (Webster et al., 1997) for 60 patients. Group leaders' negative feelings (e.g., feeling controlled or manipulated by the patient) significantly correlated with higher total HCR-20 scores, whereas more positive feelings correlated significantly with lower risk scores.

As indicated by de Vogel and de Ruiter's (2004) study, specific feelings toward an examinee can influence an evaluator's ratings of the examinee's risk of violent reoffending. We aimed to examine how evaluator attitudes toward offenders more generally influence risk ratings. Attitudes, which are often outside of conscious awareness, imply a predisposition in perception that creates a tendency to evaluate the target of the attitude more negatively or more positively (Eagly & Chaiken, 2007). Evaluator attitudes toward offenders may thereby serve as an internal source of bias with the potential to influence risk ratings. We are unaware of published research that examines how evaluator attitudes toward offenders may bias their evaluations of violent reoffending risk.

The Current Study

Previous studies have mostly focused on either internal (e.g., Boccaccini, Murrie, Rufino, et al., 2014; A. K. Miller et al., 2011) or external (e.g., Murrie & Boccaccini, 2015) sources of cognitive bias for their effects on SRAI coding. We included both a (potential) external (i.e., negative media exposure) and an internal (i.e., evaluator attitudes toward offenders) source of bias, to examine their impact on the coding of the HCR-20^{V3}.

Participants in the current study were randomly assigned to review a case file only (control group) or to watch a video clip about the crime and the examinee before reviewing the same case file (experimental condition). We hypothesized that evaluators who viewed the video clip would rate the examinee as at significantly higher risk for violent recidivism than evaluators who reviewed only the case file (H1). Furthermore, we planned to examine whether preexisting attitudes toward offenders (i.e., an internal source of bias) are associated with evaluators' coding of the HCR-20^{V3}. We expected that evaluators who held more negative attitudes toward offenders would rate the examinee as at significantly higher risk for violent recidivism than evaluators who held more positive attitudes toward offenders (H2). Finally, we examined whether the effect of negative media exposure was moderated by the evaluators' attitudes toward offenders. We expected that negative media exposure would have a larger effect on evaluators with more negative attitudes toward offenders as compared to evaluators with more positive attitudes toward offenders (H3).

3

Method

Participant Recruitment

Participation in this study was limited to participants who had been trained to complete the HCR-20 (any version) and who were proficient in English. Recruitment e-mails were sent to mental health professionals who were publicly listed on the *Nederlands Register Gerechtelijk Deskundigen* (NRGD; Dutch Register of Court Experts) website as psychologists/psychiatrists working in the area of criminal law. The NRGD was created in 2010 and is the national register of forensic science experts in the Netherlands. Experts must apply and may be approved for the register only if they meet certain field-specific requirements (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). The NRGD advertised the study in their membership newsletter via e-mail in early 2019. The International Association of Forensic Mental Health Services (IAFMHS) also advertised this study to their membership in late 2018 via e-mail.

E-mail invitations were also sent to former and current Forensic Psychology Masters' students at a Dutch university, and to current Forensic Psychology Masters' students at a

university in Scotland. The study was also advertised in July 2019 on KNAPP (<https://www.knapp-efp.nl>), a website for professional practitioners working in forensic psychiatric care in the Netherlands.

Participants

Out of 113 respondents who began the study, we excluded 53 who did not complete it. An additional six respondents were excluded because they indicated they had not been trained in any version of the HCR-20. Therefore, our remaining sample size was 54 participants.

About two-thirds of participants were women ($n = 35$; 64.8%), one-third were men ($n = 18$; 33.3%), and one respondent chose not to provide information on gender ($n = 1$; 1.9%). Most of our participants resided in the United Kingdom ($n = 20$; 37.0%), the Netherlands ($n = 16$; 29.6%), and the United States ($n = 8$; 14.8%). Three participants (5.6%) resided in Germany. There was one participant from each of the following countries: Australia, Belgium, Canada, New Zealand, Norway, Sweden, and Switzerland. Participants had a lot of experience working in mental health settings ($M = 11.13$ years; $SD = 11.29$; $Mdn = 7.0$, range = 0–41 years). Participants had conducted between 0 and 225 risk evaluations with the HCR-20^{V3} ($M = 13.98$, $SD = 35.52$, $Mdn = 4$). Additional demographics are presented in Table 3.1.

Table 3.1 Demographic Characteristics of Participants ($N = 54$)

	n	%
Age group		
18-24	5	9.3
25-34	23	42.6
35-44	12	22.2
45-54	6	11.1
55-64	8	14.8
65+	0	0.0
Highest educational level		
Bachelor	18	33.3
Master	23	42.6
Post-graduate	13	24.1
Work environment		
Hospital	13	24.1
Secure forensic psychiatric facility	11	20.4
Probation office	3	5.6
Mental health/forensic clinic	12	22.2
Private practice	6	11.1
University	9	16.7
Occupational group		
Student	14	25.9
Practitioner	32	59.3
Researcher	8	14.8

Power Analysis

A sensitivity analysis was conducted using G*Power 3.1 (Faul et al., 2009) for an *F*-test of linear regression with three predictor variables, $N = 54$, $\alpha = .05$, and power of 0.8. The analysis showed that our study was acceptably powered to detect an effect size of $f^2 = 0.22$ (a medium to large effect).

3

Procedure and Materials

The study was conducted online via Qualtrics. Data collection took place between mid-June 2018 and February of 2020. Participants were compensated for their time with a €10 (or local equivalent) online shopping gift card. This study was approved by the Ethics Review Committee of the Faculty of Psychology and Neuroscience at Maastricht University [reference number 189_07_03_2018].

Data Collection

Participants read general information about the study, which contained a cover story that the goal of the research was to examine the inter-rater reliability of the HCR-20^{V3}. After giving informed consent, participants were asked to confirm training on the HCR-20^{V3} or a previous version. We collected demographic information including age group, gender, education level, and country of residence. We also collected information about years of work experience in mental health settings, current work setting, primary occupation, as well as numerical data about the number of various types of risk evaluations (e.g., *violence, presentencing, sexual offense*) conducted over the preceding two years.

Attitudes Toward Offenders. The Attitudes Toward Prisoners Scale (ATP; Melvin et al., 1985) contains 36 statements that are rated on a 5-point Likert scale (1 = *disagree strongly* to 5 = *agree strongly*). Examples of the statements include: “Most prisoners can be rehabilitated,” “Prisoners respect only brute force,” and “Most prisoners are victims of circumstances and deserve to be helped.” Nineteen of the statements are reverse-scored, the scores are summed, and a constant of 36 is subtracted, which produces a score range of 0 to 144. Lower scores indicate more negative attitudes toward prisoners and higher scores indicate more positive attitudes toward prisoners. Split-half reliability was reported by Melvin et al. for several groups using the Spearman-Brown formula, including undergraduate students ($r = .90$), correctional officers ($r = .84$), and people who were involved in prison reform or prisoner rehabilitation work ($r = .86$).

We used a slightly modified version of the ATP Scale. First, we changed the word “prisoners” in the ATP scale to the word “offenders,” thus, we will refer to the Attitudes Toward Offenders Scale (ATO) for the purposes of this study. Second, we used a 6-point Likert scale

(1 = *strongly disagree* to 6 = *strongly agree*) and eliminated a midpoint rating of *undecided* that was in the original ATP scale. Scoring was carried out as described for the ATP scale, with the exception that the range of possible scores on the ATO is 0 to 180. We expected a negative correlation between ATO scores and HCR-20 scores, such that more negative attitudes toward offenders (lower ATO scores) would correlate with higher HCR-20 scores. Cronbach's α for the ATO in our sample was .93.

Case File and Media Exposure. The case we used was an actual criminal case that occurred in February 2014 in the United States: a young man (Mr. S.) killed both of his parents (for the case file information, see Appendix A). Participants were randomly assigned to either the experimental condition (media exposure plus case file) or the control condition (case file only). Participants were approximately evenly distributed between the experimental ($n = 28$) and control conditions ($n = 26$). They were told to imagine that they were to perform a violence risk evaluation of Mr. S. as if he was about to be released to the community without further supervision or intervention after having served his prison sentence.

In the experimental condition, participants viewed a 6.5-minute video, which was compiled from actual online media reporting about the case. The video portrayed aspects of sensationalism (e.g., “killing spree,” “the shooter stood just feet from his victims,” “he came ready to do some serious damage,” “devastating firepower”), negative comments about Mr. S.’s character (e.g., “he had a sadistic, warped sense of humor,” “he kind of terrorized people”), Mr. S.’s alleged confession to the crime, and detailed descriptions of both incriminating physical evidence and circumstantial evidence (Bakhshay & Haney, 2018). There were also aspects of neutral information in the media clip, including details of the arrest, unsympathetic portrayal of the victims, community reactions to the crime, statements regarding motive, details of the crime, information about Mr. S. and his family background, and statements by law enforcement (Bakhshay & Haney, 2018).

In the control condition, participants read a factual description of the offense committed by Mr. S. and a brief biographical and clinical history about him. Participants in the experimental condition read the exact same background information, in addition to viewing the video. The biographical information contained details about the family’s life that suggested Mr. S.’s history of emotional problems (e.g., suicide attempt, feeling neglected by his father) and discord within the family (e.g., strained parental relationship, divorce). For example, the family history contained information suggesting dysfunctional coping strategies to deal with stress and trauma (e.g., failure to seek professional help in dealing with the death of a child and a suicide attempt by Mr. S.).

HCR-20^{v3}. The HCR-20^{v3} contains a total of 20 risk factors that are divided into three scales: Historical, Clinical, and Risk Management (Douglas, Hart, et al., 2014). There

are 10 historical factors (e.g., history of problems with: (1) violence, (2) relationships, (3) employment, (4) personality disorder, five clinical factors (e.g., recent problems with: (1) insight, (2) symptoms of major mental disorder, (3) instability), and five risk management factors (e.g., future problems with: living situation, personal support, stress or coping; Douglas, Hart, et al., 2014). In addition, there are three final risk judgments: Future Violence/Case Prioritization, Serious Physical Harm, and Imminent Violence, each of which must be categorized as *low*, *moderate*, or *high*.

Participants were provided with access to a refresher on how to complete the HCR-20^{V3} and were reminded to use the *omit* category sparingly (see Appendix B). The “refresher” was a two-page document adapted for the purposes of this study from the administration overview provided in the user manual for the HCR-20^{V3} (Douglas et al., 2013). The HCR-20^{V3} rating sheet was replicated in Qualtries and participants marked their selections for each item on *presence* (0 = *omit*, 0 = *not present*, 1 = *possibly/partially present*, 2 = *definitely present*) and *relevance* (0 = *omit*, 1 = *low*, 2 = *medium*, 3 = *high*). Similar to previous research (Yang et al., 2010), we calculated overall risk ratings by summing the values for the presence ratings, resulting in a total score ranging from 0 to 40, with higher scores indicating higher risk of violent recidivism. We note that while numerical scores are often used for statistical analysis purposes, this is not recommended for clinical practice (Douglas, Hart, et al., 2014). Participants also provided final risk judgment ratings for Future Violence/Case Prioritization, Serious Physical Harm, and Imminent Violence (1 = *low*, 2 = *moderate*, 3 = *high*).

Sentencing Recommendations. Participants were asked, “If you were the one to decide, what length of prison sentence do you think would be appropriate for Mr. S.”? Ranges of sentence lengths were presented and participants selected one (0 = *no prison*, 1 = *less than 1 year*, 2 = *1-5 years*, 3 = *6-10 years*, 4 = *11-15 years*, 5 = *16-20 years*, 6 = *21-25 years*, 7 = *26-30 years*, 8 = *life imprisonment*).

Debiasing Strategies. Participants were presented with 25 potential debiasing strategies (Neal & Brodsky, 2016) and asked to provide ratings on a 5-point Likert scale (1 = *not at all effective* to 5 = *extremely effective*) to indicate the extent to which they believed the strategies are effective in mitigating the effects of cognitive bias. However, because debiasing strategies were not central to the research questions in the current study, these data will not be reported.

Prior Case Knowledge. We asked participants about their level of familiarity with Mr. S.’s case before the study (1 = *not familiar at all*, 2 = *somewhat familiar*, 3 = *very familiar*) to verify that most participants were not familiar with the case.

Finally, participants were provided with debriefing information in which we disclosed the true purpose of the study.

Data Analysis

All analyses were conducted using SPSS version 25. We used an alpha level of .05 for all statistical tests. Assessments of effect sizes as small, medium, or large rest on the conventions proposed by Cohen (1988).

Results

Preliminary Analyses

Most participants ($n = 51$; 94.4%) indicated they were not at all familiar with this case before participating in the study, and three (5.6%) indicated they were somewhat familiar with it. Attitudes toward offenders (ATO) scores ranged from 61 to 160 ($M = 120.17$, $SD = 20.77$, $Mdn = 123.00$) and were somewhat negatively skewed (-.849). A Welch's t -test revealed no significant difference in total ATO scores between the media ($M = 121.82$, $SD = 15.41$) and the control group ($M = 118.38$, $SD = 25.52$), $t(40.50) = -0.594$, $p = .556$, Hedges' $g = 0.16$, 95% CI [-0.37, 0.70].

Confirmatory Analyses

The mean HCR-20 subscale and total scores for the media and control groups are reported in Table 3.2. The HCR-20 total score refers to the summation of the scores of the three subscales. Failing to support H1, an independent samples t -test revealed no significant differences between the media and control group with respect to scores on the H, C, or R subscales, nor the HCR-20 total score.

Table 3.2 Means and Standard Deviations for HCR-20 Subscales and HCR-20 Total Score

Scale	Manipulation condition	Mean	SD	t(52)	p	Hedges'g [95% CI]
Historical	Media	9.75	4.33	-0.32	.751	0.09 [-0.45, 0.62]
	Control	9.38	4.06			
Clinical	Media	4.25	3.05	0.03	.979	0.01 [-0.53, 0.54]
	Control	4.27	2.20			
Risk Management	Media	4.57	2.80	0.98	.331	0.26 [-0.27, 0.80]
	Control	5.27	2.39			
HCR-20 total	Media	18.57	8.91	0.16	.875	0.04 [-0.49, 0.58]
	Control	18.92	7.34			

Note. The t -tests pertain to the comparison between the media and control conditions for each of the scales and the total score.

The frequencies of low, moderate, and high final risk judgments for the media and control conditions are reported in Table 3.3. We report the results of Fisher's Exact Tests as some cell counts < 5. Contrary to our hypothesis, we did not find significant differences between the media group and the control group regarding their final risk judgments.

Table 3.3 Frequencies and Cross Tabulations of Final Risk Judgments by Media and Control Conditions

	Condition	Low	Moderate	High	p	Cramer's V
Future violence	Media	4	20	4	.081	.32
	Control	2	13	11		
Serious physical harm	Media	6	15	7	0.54	.34
	Control	0	17	9		
Imminent violence	Media	10	17	1	.277	.21
	Control	12	11	3		

Note. Fisher's Exact Test reported due to expected cell counts < 5.

Thus, we found no support for the hypothesis (H1) that evaluators in the media exposure group would rate the examinee as higher risk for violent recidivism as compared to the control group on HCR-20 subscales or total scores, nor on final risk judgments. We used Spearman correlational analyses to examine the associations between ATO scores, HCR-20 scores, and the three final risk judgments (see Table 3.4).

Table 3.4 Spearman p Correlations of ATO Scores, HCR-20 Subscale Scores, and Final Risk Judgments

Variable	1	2	3	4	5	6	7
1. ATO scores							
2. Historical	-.25 [-.49, .02]						
3. Clinical	-.33* [-.55, -.07]	.66** [.48, .79]					
4. Risk Management	-.36** [-.57, -.10]	.56** [.35, .72]	.53** [.31, .70]				
5. Total Score	-.32* [-.54, -.06]	.91** [.85, .95]	.83** [.72, .90]	.78** [.64, .86]			
6. Future Violence	-.22 [-.46, .05]	.42** [.17, .62]	.43** [.18, .62]	.33* [.07, .55]	.45** [.21, .64]		
7. Serious Physical Harm	-.13 [-.39, .14]	.13 [-.14, .39]	.01 [-.26, .27]	.16 [-.12, .41]	.15 [-.12, .40]	.25 [-.02, .49]	
8. Imminent Violence	-.48** [-.66, -.25]	.49* [.25, .67]	.34* [.08, .56]	.31* [.041, .53]	.43** [.18, .62]	.33* [.07, .55]	.13 [-.15, .38]

Note. Values in square brackets indicate the 95% confidence interval for each correlation.

*p < .05; **p < .01.

ATO scores displayed moderate negative correlations with the Clinical and Risk Management subscale scores, and a strong negative correlation with the final risk judgments related to Imminent Violence. Thus, more negative attitudes toward offenders were significantly associated with higher scores on the Clinical and Risk Management scales and with a higher risk estimate for Imminent Violence, in support of H2.

Finally, we conducted a hierarchical multiple linear regression analysis with the HCR-20 total score as the outcome variable. In Step 1, the group condition was entered as the predictor. In Step 2, group condition and ATO scores were entered as predictors. In Step 3, we included an interaction term of mean-centered ATO scores and group condition to examine whether attitudes toward offenders moderated the effects of the media on HCR-20 total scores (H3). Results are presented in Table 3.5.

In line with our correlational analysis, the regression analysis demonstrated support for H2 in that ATO scores were significantly negatively associated with HCR-20 total scores. This result confirms that more negative attitudes toward offenders predicted higher HCR-20 total scores. However, we found no support for H3 as the interaction term between ATO scores and media did not significantly add to the prediction of HCR-20 total scores, $\Delta R^2 = .019$, $F(3, 50) = 1.19$, $p = .281$.

Table 3.5 Results of Hierarchical Multiple Linear Regression Analysis Predicting HCR-20 Total Scores ($N = 54$)

Predictor	b	95% CI	SE _b	β	t	p
Step 1 ($\Delta R^2 = .000$)						
Media ^a	-0.352	[-4.831, 4.127]	2.23	-0.022	-0.158	.875
Step 2 ($\Delta R^2 = .160$)						
Media	0.188	[-3.973, 4.349]	2.07	0.012	0.091	.928
ATO ^b	-0.157	[-0.258, -0.056]	0.050	-0.402	-3.12	.003
Step 3 ($\Delta R^2 = .019$)						
Media	0.271	[-3.887, 4.429]	2.07	0.017	0.131	.896
ATO	-0.123	[-0.242, -0.004]	0.059	-0.314	-2.07	.044
Media ' ATO	-0.122	[-0.346, 0.103]	0.112	-0.165	-1.09	.281

Note. CI = confidence interval. SE = Standard Error. ATO = Attitudes Toward Offenders scale. Step 3 model $F(3, 50) = 3.66$, $p = .018$, $R^2_{\text{Adj.}} = .131$.

^aMedia exposure condition = 1, control condition = 0.

^bAttitudes represented by mean-centered total score on Attitudes Toward Offenders scale.

Exploratory Analyses

We conducted an exploratory analysis to examine the extent to which participants' recommended sentence length was associated with HCR-20 total scores and ATO total scores. Spearman's rho correlation analysis revealed a significant moderate relationship

between HCR-20 total scores and recommended sentence length, $r_s = .32$, 95% CI [.04, .54], $p = .020$, $N = 54$. Furthermore, participants' attitudes toward offenders were significantly negatively correlated with recommended sentence lengths, $r_s = -.36$, 95% CI [-.58, -.09], $p = .008$, $N = 54$, indicating that more positive attitudes toward offenders were associated with recommendations for shorter sentence length.

3

Discussion

The current study is one of the few that has examined whether an SRAI can mitigate the effects of bias on the part of a forensic evaluator. In addition, to our knowledge, it is the only study to date that has attempted to examine whether an SRAI can mitigate the effects of bias that derive from a source both internal and external to a forensic evaluator. Based on previous research related to the effects of negative PTP, we hypothesized that exposure to negative media about the examinee would increase evaluators' ratings of the examinee's violence risk using the HCR-20^{V3}. We made an effort to obtain participants who were not familiar with the actual case upon which the materials were based (e.g., we did not advertise the study to U.S.-based professional organizations) to permit a "pure" test of the media effect. Upon checking, nearly all of our participants were unfamiliar with this case before participating in the study.

Contrary to our hypothesis, we found no significant differences between the media exposure and control groups on dimensional HCR-20 scales. In line with this result, we did not find significant differences between the groups in their judgments of Mr. S.'s categorical judgment of violent recidivism risk. One possible explanation for the null findings is that use of the HCR-20 was effective in mitigating any negative bias created by the media exposure (Gottfredson & Moriarty, 2006; Hoge, 2002). We also cannot rule out the possibility that, because of the small sample size, the study was underpowered to detect an actual effect.

Another possible explanation for the null findings is that the video clip did not prompt a negative bias toward Mr. S. In fact, about 50 years of research has produced conflicting results about whether PTP results in bias toward a criminal defendant (for a review, see Bruschke et al., 2016). Some studies have shown evidence for the development of a negative bias toward the defendant that increases the likelihood of a guilty finding (e.g., Hope et al., 2004; Ruva et al., 2007, 2011), whereas others have found a positive bias in favor of the defendant, in that jurors are less likely to convict after exposure to PTP (e.g., Bruschke et al., 2016; Kovera, 2002; Ruva & Hudak, 2013). In other words, the effects of PTP on juror bias toward the defendant have been less than consistent, raising some doubts about the robustness of a PTP effect (Bruschke et al., 2016). However, we also note that jurors' decisions regarding guilt (or sentencing) in a criminal trial context and decisions forensic evaluators' assessments pertaining to future

violence risk are fundamentally different. In addition, the effects of training, knowledge of forensic psychology, and experience may influence whether negative media about a criminal defendant affects forensic evaluators as compared to jurors from the general public. Therefore, we suggest that future researchers attempt to further examine the effects of PTP on forensic evaluations as distinct from juror decision-making.

Our second hypothesis was that evaluators who held more negative attitudes toward offenders would rate the examinee as at higher risk of violent recidivism. This hypothesis was supported; we found that more negative attitudes toward offenders were associated with significantly higher HCR-20^{v3} total scores. This finding concurs to some degree with previous findings that clinicians with more negative feelings toward their forensic patients gave significantly higher HCR-20 scores (de Vogel & de Ruiter, 2004; Dernevik et al., 2001). Specifically, we found that more negative attitudes toward offenders were significantly related to higher ratings on the Clinical and Risk Management subscales, as well as the final risk judgment for risk of imminent violence. Taken together, these findings suggest that evaluator attitudes toward offenders may serve as an internal source of bias that affects an evaluator's ratings of an examinee's risk of violent recidivism using the HCR-20.

As a final hypothesis, we predicted that evaluator attitudes toward offenders would moderate the effects of the negative media exposure on HCR-20 total scores. We found no support for this hypothesis, possibly because the expected effect of negative media exposure was absent. There is limited research that examines how preexisting attitudes may moderate the effect of PTP (Daftary-Kapur et al., 2014), mostly concerned with how attitudes toward authoritarian approaches to crime or attitudes toward crime in general moderate the effects of PTP on juror verdicts (e.g., Butler, 2007; Daftary-Kapur et al., 2014). Those studies indicate that jurors who hold strong views about an issue (e.g., police racism, death penalty) show relatively less sensitivity to the effects of PTP than do jurors who hold more moderate views. In our study, evaluator attitudes toward offenders were negatively skewed (i.e., skewed toward positive attitudes toward offenders), which may have made participants relatively less sensitive to the effects of negative media exposure. If replicated, it would obviously be an encouraging finding that forensic evaluators with more positive attitudes toward criminal offenders are less prone to bias due to negative media exposure about the defendant.

Results of our exploratory analysis into the association between evaluators' recommended sentence lengths, HCR-20 total scores and attitudes toward offenders indicate an interesting avenue for future research. Although forensic evaluators are not directly involved in sentencing decisions, the sentencing recommendation we solicited in this study potentially taps into reasoning related to evaluators' formulation of risk management plans. It is important to

consider evaluators' reasoning for both their risk judgments and recommendations because findings from several studies indicate judges tend to make decisions that align with forensic evaluators' opinions (e.g., Acklin et al., 2015; Gowensmith et al., 2012; Zapf et al., 2004). Therefore, a forensic evaluator's assessment of recidivism risk may exert a strong influence over judicial decisions about an offender's disposition (e.g., conditional release), particularly when the evaluator's recommendation favors detention (Blais, 2015; Elwood, 2019; Gowensmith et al., 2017). In the present study, we cannot disentangle the relationship between attitudes toward offenders, risk judgments and evaluators' recommended sentences. However, the moderate associations between these three variables suggest future research could examine their interaction, as well as their relation to decision making by triers-of-fact.

Limitations

One limitation of this study is the small sample size, which limited our ability to detect anything but large-sized effects. We note that a meta-analysis by Steblay and colleagues (1999) examining the effects of PTP on juror decision-making found consistently small effect sizes. Therefore, it would be premature to conclude that PTP has no effect at all on evaluators' risk judgments based on our inability in the present study to detect a small effect size. Obtaining a large sample of trained forensic practitioners for participation in an experimental study proved to be challenging. Although 113 people began the study, 53 did not complete it. It is possible that the participation in our study required considerable time and effort, especially in relation to the compensation that was offered to participants. Future studies on this topic would likely benefit from a collaborative effort across different research labs, to obtain a larger sample size.

A second limitation is that participants in the media exposure condition viewed only a 6.5-minute video about the case. Exposure to one type of media at one discrete time has been shown in previous studies to limit the strength of media manipulations (Ruva et al., 2014). In other words, stronger effects of media exposure have been found when various forms of media were presented multiple times and over a longer time period, similar to what is likely to occur in real-life circumstances. Moreover, accessing social media for the purposes of conducting a forensic risk evaluation appears to be increasingly common (Coffey et al., 2018; Pirelli et al., 2016, 2018). It would be worthwhile to conduct future research examining combined effects of more "traditional" forms of media and social media on forensic risk evaluations. In addition, previous juror decision making studies suggest that when there is an insufficient length of time between exposure to pretrial publicity and the relevant judgment, media appears to have an insignificant effect (Steblay et al., 1999). Since participants were asked to render their

ratings within a matter of minutes after viewing the video, it is possible that this substantially diminished the impact of the media exposure.

A third possibility that may explain the null findings related to the media manipulation is that participants may have guessed the true purpose of the study even though we included a cover story that the research question was related to the inter-rater reliability of the HCR-20^{v3}. Specifically, participants completed the Attitudes Toward Offenders scale prior to reading the case file or viewing the media, which could have indicated to many participants that the study was about bias. If participants guessed the true purpose of the study, those who viewed the media may have made conscious efforts to disregard the media (Goldyne, 2007).

Although the effects of bias tend to occur outside of conscious awareness (Nisbett & Wilson, 1977a; Wilson & Brekke, 1994), prior research indicates that when individuals become aware of potential factors that may have biased their judgments, they may seek to correct what they perceive are the effects of the bias (see, e.g., Wegener et al., 2000; Wegener & Petty, 1995). Hence, if our participants thought that the video had a potentially negative influence on their perceptions of Mr. S., they may have attempted to make corrections for this effect in their risk judgments (Sommers & Kassin, 2001). Unfortunately, we did not measure the evaluators' perceptions of Mr. S. after the media exposure, as a manipulation check, thereby making it difficult to know what impression was created by the media as compared to the control condition. A separate manipulation check would have helped to assess whether the media had a biasing effect or not, independent of whether the biasing effect of media exposure would be demonstrated on the HCR-20 coding.

Finally, the null findings related to the influence of PTP in the current study may be attributable to weaknesses in the case materials presented to participants. We utilized a real case in this study and therefore we wanted to avoid introducing any information that was not present in the actual reporting. We did not want to risk introducing fictional information into the real case narrative, thereby creating a false impression of the real people involved. However, the effect of maintaining fidelity to the real case likely posed serious restrictions on the available information with which participants could reasonably evaluate the risk for future violence. Relatedly, the observed recidivism rate among people who commit parricide tends to be exceptionally low (Bourget et al., 2007; Dantas et al., 2014), a fact our participants may have known, which may have further restricted the range of risk judgments.

Conclusion

Concerns about the potential for bias in forensic evaluations are appearing with increased frequency in the scientific literature (Zapf et al., 2018; Zapf & Dror, 2017; Zappala et al., 2017).

The structured framework of SRAIs can help clinicians “see the wood for the trees”, by keeping opinions and information unrelated to violence risk from influencing forensic risk evaluations (Carroll, 2007, p. 303). In fact, forensic clinicians tend to endorse SRAIs as an effective strategy to mitigate the influence of bias in forensic risk evaluations (e.g., Kamorowski, de Ruiter, et al., 2021; Neal & Brodsky, 2016). Yet, the limited empirical research on whether SRAIs effectively mitigate external or internal sources of bias cautions against concluding that SRAIs *do* mitigate the effects of bias. Future research regarding other types of potentially biasing, risk-irrelevant information (e.g., victim impact statements) could shed light on the question whether forensic risk evaluators should avoid exposure to this type of information.

There is a growing body of research, including the present study, that indicates that external and internal sources of bias can have an effect on SRAI completion (Boccaccini, Chevalier, et al., 2017; Chevalier et al., 2015; de Vogel & de Ruiter, 2004; Edens et al., 2015; Ismail & Looman, 2018; A. K. Miller et al., 2011; Murray et al., 2014; Murrie et al., 2008, 2009, 2013; Murrie & Boccaccini, 2015), indicating the need for further investigations of how biases may affect forensic risk evaluators’ collection of data, coding, interpretation, and decision making. Given the high-stakes nature of forensic risk evaluations, research identifying potential sources of bias that can lead to errors in recidivism risk assessment could increase the accuracy of these assessments. Furthermore, the increasing reliance on SRAIs as an aid to decision making in criminal justice systems makes it imperative that when purportedly objective risk measures are used, their results truly reflect the most accurate evaluation of risk factors.

Chapter 4

“He Seems Odd”: The Effects of Risk-Irrelevant Information and Actuarial Risk Estimates on Mock Jurors’ Perceptions of Sexual Recidivism Risk

This chapter is based on the following paper:

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. (2021). “He seems odd”: The effects of risk-irrelevant information and actuarial risk estimates on mock jurors’ perceptions of sexual recidivism risk. *Psychology, Crime and Law*. Advance online publication. <https://doi.org/10.1080/1068316X.2021.1909016>

Abstract

Previous research has shown that mock and actual jurors give little weight to actuarial sexual offending recidivism risk estimates when making decisions regarding civil commitment for so-called sexually violent predators (SVPs). We hypothesized that non-risk related factors, such as irrelevant contextual information and jurors' information-processing style, would influence mock jurors' perceptions of sexual recidivism risk. This preregistered experimental study examined the effects of mock jurors' ($N = 427$) need for cognition (NFC), irrelevant contextual information in the form of the offender's social attractiveness, and an actuarial risk estimate on mock jurors' estimates of sexual recidivism risk related to a simulated SVP case vignette. Mock jurors exposed to negative risk-irrelevant characteristics of the offender estimated sexual recidivism risk as higher than mock jurors exposed to positive information about the offender. However, this effect was no longer significant after mock jurors had reviewed Static-99R actuarial risk estimate information. We found no support for the hypothesis that the level of NFC moderates the relationship between risk-irrelevant contextual information and risk estimates. Future research could explore additional individual characteristics or attitudes among mock jurors that may influence perceptions of sexual recidivism risk and insensitivity to actuarial risk estimates.

Introduction

An estimated 6,500 people in the United States are believed to present a risk of sexual reoffending that is sufficiently high to warrant indefinite confinement in a detention facility (Koeppel, 2019). These individuals are civilly committed pursuant to Sexually Violent Predator (SVP) or Sexually Dangerous Person (SDP) statutes, which have been enacted by the federal government and 20 states (Knighton et al., 2014). These laws permit confinement of an individual who has been charged with,¹ or convicted of a sexual offense for which he or she has served their prison sentence (Knighton et al., 2014).

Most often, jurors decide whether the individual against whom an SVP petition is brought—also known as the *respondent*—meets the requisite legal criteria for SVP civil commitment (Krauss & Scurich, 2014; Lieberman et al., 2007; Zolfo, 2018). These legal criteria are: (a) the respondent has a history of sexual offending; (b) the respondent has a mental abnormality or personality disorder; (c) the respondent has some volitional impairment (i.e., difficulty controlling his or her sexual behavior); and (d) as a result of these factors, the respondent is *likely to engage in future sexual offending* (emphasis added; *Kansas v. Hendricks*, 1997). In most cases, at least one forensic mental health professional—called an *SVP evaluator*—offers expert testimony about whether the respondent meets each of the legal criteria (Jackson et al., 2004; Jackson & Hess, 2007). Ideally, jurors should make decisions based on the expert's science-based estimate of the risk that the respondent will engage in future sexual offending (Boccaccini, Murrie, & Turner, 2014; Guy & Edens, 2003; Janus & Prentky, 2008; Krauss et al., 2012; H. A. Miller et al., 2005; Scurich & Krauss, 2013; Sreenivasan et al., 2003).

Actuarial Risk Estimates and Juror Decision Making in SVP Cases

SVP evaluators most often use actuarial risk assessment instruments (ARAIs; Janus & Prentky, 2003) for the purpose of estimating the likelihood of sexual reoffending (Barbaree et al., 2001; Janus & Prentky, 2003; Murrie et al., 2009). In fact, for several decades, researchers have consistently found that ARAIs provide more accurate estimates of violent or sexual recidivism than predictions based on unstructured clinical judgment alone (Ægisdóttir et al., 2006; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009). SVP evaluators commonly use the

¹Most SVP/SDP statutes also permit SVP commitment proceedings to be initiated based on a crime(s) for which the respondent was found not guilty by reason of insanity or incompetent to stand trial (Fanniff et al., 2010; Phenix & Jackson, 2015). In addition, most jurisdictions also permit the application of SVP statutes to crimes not statutorily defined as sexual offenses, but alleged to have been “sexually motivated” (National District Attorneys Association, 2012).

Static-99R (Helmus et al., 2012) or its predecessor, the Static-99 (Hanson & Thornton, 1999), to assess the likelihood of sexual reoffending (Krauss et al., 2018). The Static-99R differs from the original version only in different scores associated with the offender's age at release (Boccaccini, Rice, et al., 2017).

Accurate risk-relevant information derived from ARAIs is particularly important in SVP hearings because the general public tends to grossly overestimate the risk of sexual reoffending (Ellman & Ellman, 2015; Katz-Schiavone et al., 2008; Levenson et al., 2007). Levenson and colleagues (2007) reported that among a sample of laypeople, the average estimate was that 74% of persons convicted of a sexual offense would reoffend. In contrast, observed recidivism rates among persons who have been convicted of a sexual offense range between 5-15% (Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Helmus, et al., 2016). Therefore, objective risk estimates derived from ARAIs can ostensibly correct jurors' apparent overestimates of sexual recidivism risk.

SVP evaluators often discuss ARAI results in their testimony at SVP hearings. Despite the finding that jurors in real SVP cases appear to *believe* forensic evaluators who use actuarial instruments make more accurate estimates of reoffending risk than evaluators who use only unstructured clinical judgment (Boccaccini, Murrie, & Turner, 2014), both mock jurors and actual jurors in SVP civil commitment hearings tend to *give little weight* to expert testimony based on ARAIs (Boccaccini et al., 2013; Guy & Edens, 2003; Krauss & Sales, 2001). Some studies suggest jurors are influenced by factors unrelated to recidivism risk, including the likability or credibility of the expert (Batastini, Hoeffner, et al., 2019), and give more weight to these factors than to the expert testimony about actuarial risk estimates (Boccaccini et al., 2013; Krauss et al., 2012; Krivacska, 2011; S. Turner et al., 2015).

In other words, most previous studies that examine the effects of ARAI risk information on jurors' perceptions of sexual recidivism risk cannot differentiate between the effects of the ARAI risk information itself and juror perceptions of expert or attorney likability/credibility on perceptions of the SVP respondent's risk of sexual recidivism, because the two information sources are confounded. For example, participants in Krauss et al. (2012) watched a 1-hour video simulation of an SVP trial that included testimony from a psychological expert witness and opening statements from both the petitioner's (the state) and the SVP respondent's attorney. The expert in the Krauss et al. study testified that he had used two ARAIs, but the risk estimates derived therefrom were not mentioned by the expert—only that the ARAIs had been completed as part of the risk evaluation.

In fact, we are aware of only one previous experimental study that does *not* employ some form of expert testimony or attorneys' arguments (either in video or written format) in

SVP civil commitment studies. Varela et al. (2014) examined how varying the communication format of Static-99R risk information affects jurors' sexual recidivism risk estimates in an SVP civil commitment context. Specifically, they varied the format of a Static-99R risk estimate as a categorical estimate (i.e., low or high risk), a relative risk ratio (i.e., the offender would be expected to have a [three-fourths recidivism rate] or [2.91 times higher than the recidivism rate] of the "typical sexual offender"), or a recidivism rate (i.e., [9.4%] or [31.2%] were re-arrested for a sexual offense within five years). Varela and colleagues then measured how the communication format affected 211 prospective jurors' ratings of an SVP respondent's sexual recidivism risk. They found no significant main effect of communication format on jurors' perceptions of recidivism risk based on Likert scale ratings of the likelihood of sexual reoffending within five years.

However, there was a significant interaction effect between the risk communication format and the risk level indicated by the Static-99R information. When participants read recidivism risk information derived from the Static-99R (score of 6) that was presented in a categorical format (i.e., "high risk"), they rated the risk of recidivism as significantly higher than those who read about a recidivism rate (i.e., of men who scored a 6 on the Static-99R, 31.2% were re-arrested for a sexual offense within five years; $d = .68$, medium to large effect). There was no significant interaction effect when the Static-99R indicated a score of 1 (i.e., "low risk"). In other words, when the Static-99R results indicated higher than average risk of sexual recidivism, participants in the Varela et al. study rated the risk as significantly lower when they were provided with a recidivism rate rather than a categorical risk estimate.

Still, it remains unclear *why* jurors do not appear to make effective use of actuarial risk estimates, particularly as communicated by forensic evaluators in SVP civil commitment hearings (S. L. Miller & Brodsky, 2011; Pennycook et al., 2014; D. B. Turner et al., 2015; Walters et al., 2014). Perhaps they are influenced by characteristics of the expert witness, or perhaps certain risk communication formats are more effective than others. At present, there is no experimental research that can help disentangle these various effects. Furthermore, we are unaware of any previous experimental study that directly examines whether risk-irrelevant factors have an independent effect on jurors' perceptions of an SVP respondent's sexual recidivism risk and whether that influences their apparent insensitivity to a risk estimate derived from an ARAI. Yet, the Varela et al. (2014) findings suggest that, in some circumstances, providing jurors with a risk estimate—particularly a recidivism rate—that is independent of expert effects may be helpful in promoting jurors' estimates of sexual recidivism risk that are more in line with the results of an ARAI.

Juror Neglect of Actuarial Risk Estimates

While some researchers have suggested that a lack of knowledge about science (Batastini, Hoeffner, et al., 2019) or a lack of numeracy (Krauss et al., 2004; Krauss & Lee, 2003; Krauss & Sales, 2001) may account for jurors' tendency to discount expert testimony based on ARAI results, others have argued that individual differences may contribute to how information is perceived and integrated in the context of legal decision-making (Gunnell & Ceci, 2010; Krauss et al., 2004; Lieberman, 2002; Lieberman et al., 2007). Research examining how potentially biasing factors, such as risk-irrelevant information, affect jurors' risk perceptions may help clarify jurors' apparent insensitivity to actuarial risk estimates (Krauss & Sales, 2001; Lieberman et al., 2007).

Risk-Irrelevant Contextual Information and Juror Decision Making in SVP Cases

We define risk-irrelevant contextual information as those factors that, on their own, have not been proven to be empirically associated with sexual recidivism. Examples are: type of employment, number of previous marriages, volunteerism, public outrage about the crime, and the victim's opinion about the sufficiency of a criminal sentence. Yet, risk-irrelevant contextual information (such as emotionally-evocative victim statements) has been found to bias psychologists, thereby causing these experts to increase their estimate of an SVP respondent's risk of future sexual offending (Jackson et al., 2004). Furthermore, risk-irrelevant contextual information, such as the respondent's *social attractiveness* or likability (Alicke & Zell, 2009; Landy & Aronson, 1969; Michelini & Snodgrass, 1980; Richardson & Campbell, 1982), media stories and public reactions to an offense, as well as the victim's view of the fairness of the punishment for the crime (P. H. Robinson et al., 2012), has been shown to affect attributions of guilt and sentencing decisions in criminal cases (Abel & Watters, 2005; Alicke & Zell, 2009; Michelini & Snodgrass, 1980).

The psychological phenomenon known as the *halo effect* could perhaps account for this biasing effect. The halo effect refers to the human tendency to make unconscious judgments about the specific attributes of an individual based on one's global assessment of the individual (Nisbett & Wilson, 1977b). For example, Nisbett and Wilson (1977b) found that whether a teacher was rated as "warm" or "cold" by students had a significant and pronounced effect on ratings of the teacher's physical attractiveness, mannerisms and accent. When the teacher was perceived as warm, participants provided significantly higher ratings of the teacher's physical attractiveness and rated his mannerisms and accent as significantly more appealing, whereas participants who perceived the teacher as cold rated him as less physically attractive and his mannerism and accent as more irritating. Therefore, we suspect bias created by risk-

irrelevant contextual information may generate a global impression of a ‘good’ versus a ‘bad’ person, which then impacts the juror’s estimate of recidivism risk (Gunnell & Ceci, 2010; Hilton et al., 2005, 2015). Yet, there is limited research on whether and how risk-irrelevant contextual factors may influence jurors’ case-specific estimates of reoffending risk or SVP civil commitment decisions (Boccaccini et al., 2013).

Need for Cognition and Decision Making

Individuals differ in the extent to which they enjoy expending cognitive effort to understand a situation, a characteristic known as *need for cognition* (NFC; Cacioppo & Petty, 1982). The Need for Cognition Scale (Cacioppo & Petty, 1982) is a measure of NFC (Cacioppo et al., 1984). People with low NFC scores tend to process information experientially, which means they engage in rapid decision-making that is largely unconscious, emotion-driven and leads to conclusions based on generalizations (Epstein, 1993). They also tend to focus on peripheral cues, such as the attractiveness or likability of the source of information (Cacioppo et al., 1996; Haugvedt et al., 1992), and pay little attention to the quality of the information (Cacioppo et al., 1996; Shestowsky & Horowitz, 2004).

Conversely, people with higher NFC scores appear more likely to use a rational approach by exerting conscious effort to seek evidence and apply logic in their thinking (Epstein, 1993). They are more likely to evaluate the quality of arguments contained in a message than those who are lower in NFC (Cacioppo et al., 1983, 1996). Based on cognitive-experiential self-theory (CEST), people use both rational and experiential modes of information processing, but individual differences and situation-specific factors determine which mode is predominant in guiding behavior (Denes-Raj & Epstein, 1994).

Previous research has revealed that when people are encouraged to think analytically, they tend to be more influenced by testimony based on actuarial risk estimates than testimony based on unstructured clinical judgments in SVP cases (cf., Krauss et al., 2012; Lieberman et al., 2007; Lieberman & Krauss, 2009). Therefore, we hypothesize that jurors who are higher in NFC are more likely to render an estimate of the likelihood of sexual reoffending by the SVP respondent that aligns with that indicated by an ARAI than jurors who are lower in NFC.

Although a number of studies about juror decision-making suggest that jurors who score lower on NFC may be more persuaded by clinical testimony about sexual recidivism risk than the same testimony based on an actuarial risk estimate (Gunnell & Ceci, 2010; Krauss et al., 2004; Lieberman et al., 2007), some studies indicate that mock jurors may be more influenced by testimony based on actuarial risk estimates when they are encouraged to think rationally and analytically (Lieberman et al., 2007; Lieberman & Krauss, 2009). Yet it is unclear whether individuals’ natural inclination to rationally and analytically weigh evidence

predicts their ability to adequately utilize actuarial risk information when also confronted with potentially biasing risk-irrelevant information. Such research may provide insights into how individual differences among jurors affect how they integrate both types of information in their judgment of risk.

The Current Study

Given that most SVP cases are decided by jurors (Lieberman et al., 2007; Zolfo, 2018), the aim of this preregistered experiment is to gauge whether and how potentially biasing contextual information may affect jurors' estimates of sexual reoffending risk in the context of an SVP case when they are provided with both risk-irrelevant and risk-relevant information about the respondent. Therefore, we examined whether bias affects mock jurors' perceptions of sexual reoffending risk, whether more objective risk estimates derived from a risk assessment instrument can counteract such bias among mock jurors, and whether NFC as an individual trait variable moderates the amount of bias in mock jurors' risk estimates.

First, we examined whether mock jurors' perceptions of the likelihood of sexual reoffending are affected by either positive or negative irrelevant contextual factors in a simulated sexual offender civil commitment case. We selected factors that are not empirically associated with sexual offender recidivism, hence our use of the term *risk-irrelevant*. For example, in the positive contextual information condition, the respondent is portrayed as more "socially attractive" in that he had a long record of employment as a bank manager, he volunteered at a homeless shelter, and played the organ at church. In contrast, in the negative information condition, the respondent is portrayed as less socially attractive because he had been employed for a brief time in his job as a janitor, a neighbor described him as "odd," and he played online games in his free time. We hypothesized that participants who receive negative contextual information would provide higher estimates of the likelihood of sexual reoffending before receiving any information on actuarial recidivism risk (Time 1; T1) than would participants who received positive contextual information (H1).

Second, we measured the extent to which mock jurors adjusted their perceptions of the likelihood of sexual reoffending after reviewing the respondent's Static-99R information indicating moderate risk (Time 2; T2). Based on previous research that suggests jurors give relatively little weight to actuarial risk estimates (Boccaccini et al., 2013; Krauss & Sales, 2001; D. B. Turner et al., 2015), we expected that mock jurors who have previously received risk-irrelevant negative contextual information about the respondent would rate him as significantly more likely to reoffend at T2 than mock jurors who received risk-irrelevant positive information, despite the presentation of the same risk-relevant Static-99R information to both groups (H2).

Third, we evaluated the extent to which mock jurors' level of NFC moderated the effect of Static-99R information on their T2 estimates of the likelihood of the respondent sexually reoffending. Accordingly, we expected that the effects of risk-irrelevant information on reoffending risk estimates at T2 would vary based on the mock jurors' level of NFC, such that the effects would be larger among mock jurors with lower NFC scores (H3).

Most previous studies that have examined the effects of ARAI information on jurors' estimates of the likelihood that an SVP respondent will reoffend are based on *expert testimony* about ARAI results (Boccaccini et al., 2013; Guy & Edens, 2003; Krauss et al., 2012; McCabe et al., 2010; Scurich & Krauss, 2014; D. B. Turner et al., 2015). The findings from these studies suggest that jurors are not strongly influenced by such expert testimony (Boccaccini et al., 2013; Krauss et al., 2012; D. B. Turner et al., 2015). Yet, it is difficult to separate the effects of expert testimony from the content of the testimony as the effect of the information itself may be obscured by factors such as the expert's credibility (Edens et al., 2012) or likability (Batastini, Hoeffner, et al., 2019). Therefore, we presented the Static-99R results to jurors as they might be conveyed in an expert's written report, rather than through the filter of an expert's testimony.

The preregistration for this study can be found at https://osf.io/wr48v/?view_only=fdd52d991eo491c8f2b3d3901boa4b7.

Method

Participants

Participants were recruited through Amazon Mechanical Turk (MTurk) and required to be at least 18 years of age, proficient in English, and reside in the United States. We limited our sample to U.S. residents for two reasons. First, we expected that Americans may have different views regarding how people who have committed a sexual offense should be managed, compared to people in other countries. Second, the SVP civil commitment system in the U.S. is different than other countries, particularly with respect to the fact that jurors often decide whether a person convicted of a sexual offense should be civilly committed (Janus, 2013; Witt & DeMatteo, 2019). We also required that participants have at least a 99% approval rating on MTurk and must not have participated in the pretesting phase of the study in which we evaluated the strength of the manipulation.

We planned to exclude participants who did not meet the inclusion criteria, who did not complete the study, who failed one or more attention checks, or indicated that their data should not be used (Meade & Craig, 2012). We also planned to exclude participants whose

total time spent on the study was more than three times the median absolute deviation below the sample median (Leys et al., 2013), but none met this criterion.

In addition, we set an exclusion criterion on a per-analysis basis such that we would exclude participant responses whose T1 or T2 reoffending risk estimates were more than three times the median absolute deviation above or below the corresponding sample median (Leys et al., 2013). No participants met this exclusion criterion based on their T1 risk estimates. However, we encountered 100 participants (23.4%) whose T2 risk estimates were more than three times the median absolute deviation above the sample median. This rate of exclusion is far above what would be considered as representative of true outliers. Therefore, this group of participants was not excluded from the main analyses. Instead, the T2 risk estimates were log-transformed to address the skewness in the data (see Results).

Of the 427 participants whose data were analyzed, 240 identified as male (56.2%), 181 identified as female (42.4%), two identified as transgender male (0.5%), one identified as gender variant/non-conforming (0.2%), one identified as an option not listed (0.2%), and two declined to provide gender identity information. Participants ranged in age from 19 to 75 years ($M = 37.7$, $SD = 11.47$). The highest education level varied from less than a high school diploma ($n = 2$; 0.5%) to participants who had a professional ($n = 4$; 0.9%) or doctoral ($n = 4$; 0.9%) degree. About 43.3% held a bachelor's degree ($n = 185$), 11.5% had an associate's degree ($n = 49$), and 21.1% had some college education ($n = 90$). Participants who had a high school diploma as their highest level of educational attainment comprise 10.3% of the sample ($n = 44$). Nearly every U.S. state ($n = 46$) was represented, of which the highest number of participants resided in California ($n = 50$; 11.7%). Three participants did not provide their state of residence. Seventy-two (16.9%) participants had previously served as a juror in a legal case.

Procedure and Materials

Participants self-selected to participate in the study as advertised on MTurk. The study was conducted online via Qualtrics. Participants were paid \$1 to participate and a bonus of \$2 if they correctly answered both attention check questions. This study was approved by the ethical review committee of the Faculty of Psychology and Neuroscience at Maastricht University (reference number 207_13_04_2019).

Power Analysis

An *a priori* power analysis was conducted using G*Power 3.1 (Faul et al., 2009) for a hierarchical multiple regression analysis to test for the increase in explained variance (ΔR^2)

associated with the interaction term (as predicted in H3). The parameters were set as $R^2 = .022$ to detect a small effect, $\alpha = .05$, power = .80, and total number of predictors = 4. The minimum number of participants required was 351. However, because we expected the need to exclude participants for various reasons (e.g., failed attention checks), we planned to collect data from at least 400 participants to achieve sufficient statistical power.

Data were collected from 690 participants. Fifty-eight people did not complete the study, and 171 failed at least one attention check. We also excluded data from an additional 34 participants based on geo-coordinates falling outside the range for the United States, geo-coordinates that did not match the state indicated in the demographic data response, or if the geo-coordinate metadata were absent. Our remaining sample size was 427.

Procedure

Informed consent was obtained from participants, and they were asked to confirm that they met the eligibility criteria of being at least 18 years of age, proficient in reading and understanding English, and reside in the United States. Participants read a brief introduction about the study and received contact information of the first author in the event of any issues with the study. We used several controls for quality assurance. First, participants were not allowed to go back to a previous page once they moved on to the next one. Second, a timer was used to track how much time they spent on each page (participants were not aware that they were being timed). Finally, the “next” button to move forward was also delayed on most pages (participants were informed of this) to discourage inattentive or random responding.

Before participants read anything about the case, they provided an estimate of how likely they think it is that a person who has been convicted for a sexual offense will commit another sexual offense (hereafter, general risk estimate), using a graphic sliding scale ranging from 0–100% in 1% increments. We chose to have participants complete the Need for Cognition Scale (NFCS) immediately thereafter, as a ‘distractor task’, to reduce the likelihood that they would conflate their own beliefs about sexual offending recidivism risk and the information contained in the case vignette.

Need for Cognition Scale. We used a slightly modified version of the short form Need for Cognition Scale (NFCS), which contains 18 items and is intended to assess an “individual’s tendency to engage in and enjoy effortful cognitive endeavors” (Cacioppo et al., 1984, p. 306). The NFCS produces a need for cognition (NFC) score based on the respondent’s agreement with statements such as “I would prefer complex to simple problems” and “I like tasks that require little thought once I’ve learned them.” In the original version of the scale, item ratings are based on an 8-point Likert scale (-4 to +4) and nine of the items are reverse-scored (Cacioppo et al., 1984). Higher NFC scores are associated with a higher level of NFC,

lower scores with a lower level of NFC. The NFCS demonstrates strong internal consistency ($\alpha = .90$; Cacioppo et al., 1984), good test-retest reliability (Bertrams & Dickhäuser, 2010; Sadowski & Gulgoz, 1992; Verplanken, 1991), and additional studies have found support for divergent, discriminant, and predictive validity (see Cacioppo et al., 1996 for a review; Osberg, 1987; Stark et al., 1991).

In the current study, we asked participants to indicate the extent to which each NFCS item is typical of their own behavior on a 5-point Likert scale (1 = *does not describe me*, 2 = *describes me slightly well*, 3 = *describes me moderately well*, 4 = *describes me very well*, 5 = *describes me extremely well*), rather than an 8-point scale. For our analyses, we calculated a total score by adding all ratings together and subtracting a constant of 18, yielding a lowest possible score of 0 and a highest possible score of 72. Cronbach's α in our sample was .94.

Case Vignette. In this study, we defined risk-irrelevant contextual information as information that has not been shown to be empirically related to sexual reoffending risk (e.g., type of job, length of employment, others' perceptions about his sociability, preferred leisure activities, and marital history). Participants were randomly assigned to read a case vignette containing positive ($n = 213$) or negative ($n = 214$) risk-irrelevant contextual information about an SVP respondent named John Smith (see Appendix C). The two case vignettes varied with respect to the previously mentioned risk-irrelevant contextual information. The effectiveness of the manipulation had been confirmed through participant ratings of Mr. Smith's likability in a prior pilot study.

We did not provide any offense-related information before participants in the current study rated the likability of Mr. Smith because we expected the manipulation would have been less effective if participants were informed of the nature of his crime (i.e., a sexual offense) and prior non-sexual assault conviction. Therefore, after reading only the initial risk-irrelevant background information, participants were asked to rate Mr. Smith's likability using three items: (a) likability, (b) friendliness, and (c) warmth. These three items were reproduced from the Reysen Likability Scale (Reysen, 2005). The Reysen Likability Scale is comprised of 11 items rated on a 7-point Likert scale (1 = *very strongly disagree* to 7 = *very strongly agree*). Lower ratings indicate the target is less likable, and higher ratings indicate the target is more likable. For this study, we used a 6-point Likert scale (1 = *completely disagree*, 2 = *mostly disagree*, 3 = *somewhat disagree*, 4 = *somewhat agree*, 5 = *mostly agree*, and 6 = *completely agree*) and removed the *neutral* option. Cronbach's α for the original scale was .91 (Reysen, 2005). Cronbach's α in this sample for the three items measured was .98. We calculated the average rating of the three likability factors to create a composite variable for statistical analysis of the manipulation's effectiveness.

Participants then read the remainder of the case vignette describing the offense for which Mr. Smith was currently incarcerated (i.e., rape of a female acquaintance), his criminal history (one previous conviction for assault related to a bar fight when he was 18 years old), public opinion about the length of his prison sentence (5 years) for the sexual assault, and the victim's opinion about the fairness of Mr. Smith's prison sentence. The two criminal convictions are relevant to sexual recidivism risk and are accounted for by the Static-99R. Furthermore, we did not vary criminal history or Static-99R information between the positive and negative case vignette. However, the additional risk-irrelevant factors of the public's and victim's opinion of the length of the prison sentence differed between the positive and negative vignette conditions.

Participants were then instructed to imagine that they had been selected to be a juror in a civil commitment hearing. We provided background information about sex offender civil commitment, including the nature and purpose of civil commitment and the process generally followed to determine whether an offender should be subjected to civil commitment. To ensure attention and comprehension of the explanation of civil commitment, we used a multiple-choice attention check question of: "Confinement in a secure facility under SVP laws most often occurs," with response options of (a) *while a person is serving their prison sentence*; (b) *before a person is sentenced to prison*; or (c) *after a person has served their prison sentence*. Participants were also provided with information about the legal criteria that must be met for SVP civil commitment (see Appendix D).

Subsequently, participants were asked to provide an initial risk estimate (T1; 0–100% using a graphic sliding scale) and an initial disposition recommendation for Mr. Smith (1 = *released to the community with no conditions*, 2 = *released to the community under supervision only*, 3 = *release to the community under supervision and mandated sex offender treatment*, 4 = *civil commitment*) based on all the information they considered relevant. They were also informed that they would later receive additional information, after which they would be asked to provide a risk estimate and disposition recommendation once again.

Static-99R Information. We operationalized risk-relevant information as Mr. Smith's Static-99R (Helmus et al., 2012) results. The Static-99R comprises 10 risk factors that are empirically related to the likelihood of sexual offense recidivism (Phenix, Fernandez, et al., 2016). An evaluator codes each of the 10 factors and the resulting total score is associated with an estimated likelihood of reoffending based on previous recidivism rates of people convicted of a sexual offense who were released into the community (Knighton et al., 2014).

Similar to the approach taken by Varela and colleagues (2014) in their study about risk communication formats, we informed participants about risk assessment tools and their use in estimating recidivism risk, provided information about the development of the Static-99R

and how a risk score is calculated using the Static-99R (see Appendix D). We also provided participants with a list of the items contained in the Static-99R (see Table 4.1). In contrast to Varela et al., we provided all participants with identical Static-99R results for Mr. Smith including his absolute score (2), risk category (medium risk), and the 5-year reoffending rate ranges associated with that risk category based on a routine sample (between 7.0% and 8.8%; Phenix, Fernandez, et al., 2016; see Appendix E). The format of the risk score information was provided exactly as outlined in the template recommended in the Static-99R and Static-2002R Evaluator's Workbook (Phenix, Helmus, et al., 2016, p. 30). An attention check question followed in which participants typed in Mr. Smith's Static-99R score.

Table 4.1 Risk Factors of the Static-99R

Items of the Static-99R
1. Age at release from index offense
2. Ever lived with a lover (at least 2 years)
3. Index non-sexual violence
4. Prior non-sexual violence
5. Prior sex offenses
6. Four or more prior sentencing dates (excluding index)
7. Any convictions for non-contact sex offenses
8. Any unrelated victims
9. Any stranger victims
10. Any male victims

Note. Adapted from (Phenix, Helmus, et al., 2016). Used with permission.

Participants' T2 risk estimates (0-100% on graphic sliding scale) and disposition recommendations (same options as presented at T1) were obtained after they had reviewed the Static-99R information. In addition, participants rated the importance of several factors to their estimate of the likelihood of the respondent's sexual reoffending, including his: (a) relationships with others, (b) employment history, (c) criminal history, (d) impact of the crime on the victim, (e) length of prison sentence, and (f) Static-99R score, using a 5-point Likert scale (*1 = not at all important, 2 = slightly important, 3 = moderately important, 4 = very important, 5 = extremely important*). Using the same scale, participants rated the importance of each of the aforementioned factors to their disposition recommendation, as well as one additional factor: the participant's own estimate of the likelihood of the respondent committing another sexual offense. Participants' disposition recommendations and importance ratings of the factors were analyzed as exploratory analyses and are reported in Supplemental Materials (see Appendix G).

Demographic Information. Participants were then asked if they had ever served on a jury in a legal case and to provide their age, gender, level of education, and state of residence. Finally, we asked participants to indicate whether we should use their data in our analyses (*yes/no*) and we provided them with debriefing information about the nature and purpose of the study and the research questions we hoped to answer.

Data Analysis

We conducted confirmatory analyses in accordance with our preregistration. Where data transformations were necessary, we have noted these in the respective sections. All analyses were conducted using SPSS version 26. We used an alpha level of .05 for all statistical tests. Assessments of effect sizes as small, medium, or large rest on the conventions proposed by Cohen (1988).

4

Results

Manipulation Check

We conducted an independent samples Welch's *t*-test to compare the likability ratings between the positive and negative contextual information conditions. On average, participants in the positive condition rated the respondent as significantly more likable ($M = 5.06$, $SD = 0.68$) than did participants in the negative condition ($M = 2.22$, $SD = 1.02$), $t(370.85) = -33.96$, $p < .001$, Hedges' $g = 3.28$, 95% CI [2.99, 3.57]. This finding confirms that before reading about the crime information, there was a significant difference between the groups in the perceived likability of the SVP respondent, suggesting that the contextual information manipulation was effective.

The Effect of Contextual Information on T1 Risk Estimates

Overall, participants' general risk estimates ($M = 66.3\%$, $SD = 20.0$) were much higher than observed sexual recidivism rates of 5-15% (Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Helmus, et al., 2016). An independent samples *t*-test confirmed that there was no significant difference between conditions with respect to participants' general sexual recidivism risk estimates, $t(425) = -1.43$, $p = .154$, Hedges' $g = 0.14$, 95% CI [-0.05, 0.33]. Nevertheless, to isolate the effects of the contextual information, we controlled for participants' general risk estimates in the analysis of T1 risk estimates.

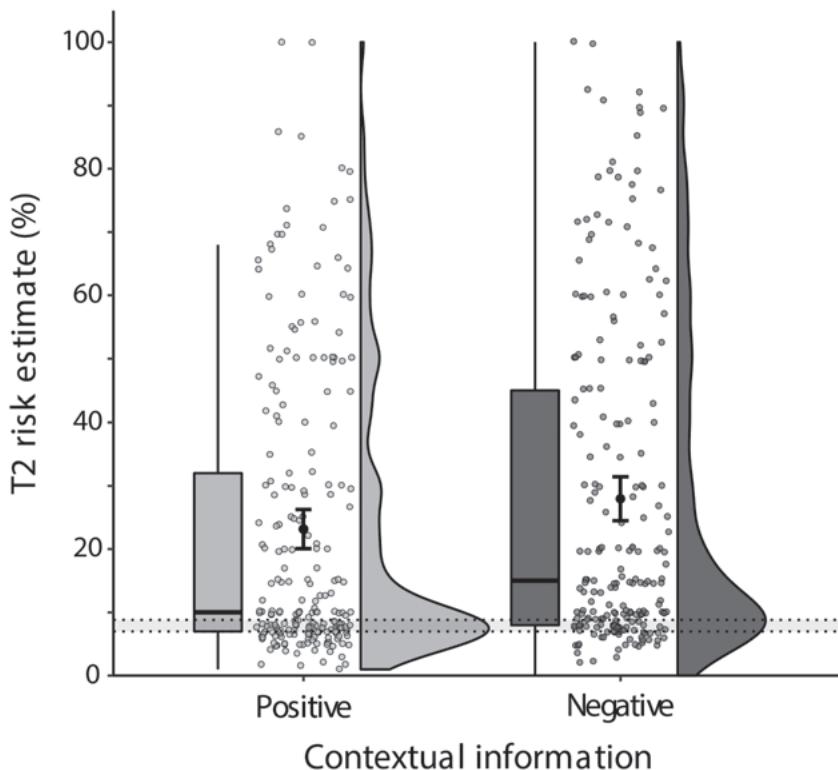
A one-way ANCOVA was conducted on participants' ratings of the respondent's recidivism risk at T1, with contextual information (positive vs. negative) as the independent variable and participants' general estimates of sexual recidivism risk as covariate.² There was a significant, medium-sized effect of contextual information, $F(1, 424) = 39.42, p < .001, \eta^2_p = .085, 90\% \text{ CI } [.047, .130]$. In support of H1, participants in the positive contextual information condition rated the recidivism risk as significantly lower ($M = 40.9, SD = 25.3$) than did participants in the negative contextual information condition ($M = 55.1, SD = 22.6$). Moreover, the analysis indicated that participants' general recidivism risk estimates accounted for a large and significant portion of the variance in T1 risk estimates, $F(1, 424) = 161.1, p < .001, \eta^2_p = .275, 90\% \text{ CI } [.217, .330]$. As expected, participants who reported higher general risk estimates reported higher case-specific risk estimates at T1, bivariate $r = .52, p < .001$.

The Effects of Contextual Information and NFC on T2 Risk Estimates

At T2, participants' mean risk estimate ($M_{\text{raw}} = 25.6\%, SD = 24.4$) was substantially decreased from T1, but nevertheless exceeded that indicated by the Static-99R (i.e., 7.0–8.8%). Additional exploratory analyses of the T2 ratings are presented in Supplemental Materials (see Appendix G). To test the effects of the positive and negative contextual information, and whether NFC moderates the relationship between the type of risk-irrelevant contextual information (positive or negative) and T2 risk estimates, we conducted a hierarchical multiple linear regression analysis with participants' T2 risk estimates as the outcome variable. A visual inspection of the data revealed that participants' risk estimates were severely positively skewed (see Figure 4.1). Moreover, a regression analysis using the raw values resulted in non-normally distributed residuals and considerable heteroscedasticity. Hence, we added a constant of one to participants' T2 risk estimates (because one participant estimated T2 risk as 0%) and performed a log transformation before conducting the regression analysis.³

² To address concerns related to the exclusion of participants who failed one or both of the attention checks, we reran the analyses for H1 for the entire sample that completed the study and were not excluded for geocoordinate issues and we found no significant change. Results are reported in Supplemental Materials (see Appendix F).

³ The heavy skew in the data was unanticipated and the log-transformation was not part of the preregistered analysis plan. Moreover, we are aware of the problems that can be introduced into hypothesis testing when adding a constant to the original values and/or using a log-transformation of the data (Feng et al., 2014). Therefore, to exclude the possibility that adding a constant of one and log-transforming the T2 risk estimates substantially altered our conclusions based on the regression analysis, we ran a moderation analysis on the raw data to compute bootstrapped confidence intervals using the PROCESS macro in SPSS (Hayes, 2018). This analysis did not change the nature of the results. Full details of the analysis can be found in the Supplemental Materials in Table S1 (see Appendix G).

Figure 4.1 Raincloud Plot and Boxplot of T2 Risk Estimate Distribution

Note. Raincloud plot and boxplot showing the positively skewed distribution of participants' risk estimates at T2. The horizontal shaded band represents the 5-year sexual reoffending rate indicated by the Static-99R information (7.0%–8.8%). The black dots with error bars represent group means and 95% confidence intervals. The boxplot whiskers span 1.5 times the interquartile range (trimmed at 0% and 100%).

In Step 1, contextual information ($0 = \text{positive}$, $1 = \text{negative}$), general risk estimates, and participants' NFC total scores were entered as predictor variables.⁴ In Step 2, the interaction term between contextual information and NFC scores was entered. Participants' NFC scores were mean-centered for the regression analysis to avoid multicollinearity and to ease the interpretation of the interaction regression coefficient (Jaccard & Turrisi, 2003). Regression statistics are reported in Table 4.2.

The predictors entered at Step 1 contributed significantly to the regression model, $F(3, 422) = 18.75, p < .001$, and explained 11.7% of the variation in T2 risk estimates. We did

⁴ In our preregistered analysis plan, we erroneously stated that the T1 risk estimates would be controlled for in the analysis. However, because the T1 estimates were measured after (and were influenced by) the experimental manipulation, it is not appropriate as a covariate. Instead, to control for preexisting differences in risk perception, we entered the general risk estimate as covariate.

not find support for H2, as contextual information was not a significant predictor of T2 risk estimates, although participants in the negative contextual information condition ($M_{raw} = 27.9, SD = 25.8$) reported higher risk estimates than did participants in the positive contextual information condition ($M_{raw} = 23.2, SD = 22.9$). This result indicates that participants were not significantly influenced by the contextual information after reviewing the risk-relevant Static-99R information. Moreover, although participants' general risk estimates were significantly associated with their T2 risk estimates, NFC scores were not. Adding the interaction term of contextual information and NFC at Step 2 did not significantly improve the model, $F(4, 421) = 0.002, p = .969$. Thus, the analysis failed to support H3.

Table 4.2 Predictors of T2 Risk Estimates

Predictor	b	95% CI	SE _b	β	p
Step 1 ($R^2 = .118^{***}$)					
General risk estimate	0.015	[0.011, 0.019]	0.002	.324	< .001
NFC	-0.002	[-0.007, 0.003]	0.003	-.041	.372
Contextual information	0.157	[-0.006, 0.319]	0.083	.087	.059
Step 2 ($\Delta R^2 = .000$)					
General risk estimate	0.015	[0.011, 0.019]	0.002	.324	< .001
NFC	-0.003	[-0.010, 0.005]	0.004	-.045	.485
Contextual information	0.157	[-0.006, 0.319]	0.083	.087	.059
NFC × Contextual Information	0.000	[-0.009, 0.010]	0.005	.006	.923

Note. $N = 427$. Contextual information condition was dummy-coded (0 = positive, 1 = negative). NFC = Need for Cognition scale. Final model: $F(4, 422) = 14.03, p < .001, R^2_{Adj} = .109$.

Exploratory Analysis

Our preregistered exclusion criterion, according to which we excluded participants who failed one or both attention checks, resulted in the exclusion of 171 participants (24.8% of the 690 participants who began the study). We suspected that this exclusion criterion may have inadvertently led to participants who were lower in NFC being excluded from analysis. This speculation was confirmed through a Welch's *t*-test, which showed that excluded participants had a significantly lower NFC total score ($M = 40.3, SD = 12.9$) than non-excluded participants ($M = 44.7, SD = 16.2$), $t(389.95) = 3.53, p < .001$, Hedges' $g = 0.29$, 95% CI [0.11, 0.47]. Therefore, we repeated the hypothesis testing for H2 and H3 to assess whether including participants with a failed attention check would affect the results.

H2 and H3 Analysis Including Participants Who Failed One or Both Attention Checks

Like the main analysis, at T₂, participants' mean risk estimate ($M_{\text{raw}} = 31.7\%$, $SD = 28.0$) was substantially decreased from T₁, and exceeded that indicated by the Static-99R (i.e., 7.0–8.8%). To test the effects of the positive and negative contextual information, and whether NFC moderates the relationship between the type of risk-irrelevant contextual information (positive or negative) and T₂ risk estimates, we conducted a hierarchical multiple linear regression analysis with all participants' T₂ risk estimates as the outcome variable (adding a constant of one and log-transforming the T₂ risk estimates to correct for skewed data and non-normally distributed residuals and heteroscedasticity in the raw values before conducting the regression analysis).⁵

In Step 1, contextual information (0 = *positive*, 1 = *negative*), general risk estimates, and participants' NFC total scores were entered as predictor variables. In Step 2, the interaction term between contextual information and NFC scores was entered. Participants' NFC scores were mean-centered for Step 2 of the regression analysis to avoid multicollinearity and to ease the interpretation of the interaction regression coefficient (Jaccard & Turrisi, 2003). Regression statistics are reported in Table 4.3.

Table 4.3 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' T₂ Risk Estimates (No Exclusions for Failed Attention Checks)

Predictor	<i>b</i>	95% CI	<i>SE_b</i>	<i>β</i>	<i>p</i>
Step 1 ($R^2 = .118^{***}$)					
General risk estimate	0.014	[0.010, 0.017]	0.002	.290	< .001
NFC	-0.007	[-0.011, -0.002]	0.002	-.109	.005
Contextual information	0.128	[-0.017, 0.274]	0.074	.067	.083
Step 2 ($\Delta R^2 = .000$)					
General risk estimate	0.014	[0.010, 0.017]	0.002	.290	< .001
NFC	-0.007	[-0.014, 0.000]	0.003	-.113	.041
Contextual information	0.128	[-0.017, 0.274]	0.074	.067	.084
NFC × Contextual Information	0.001	[-0.009, 0.010]	0.005	.006	.916

Note. $N = 598$. Contextual information condition was dummy-coded (0 = *positive*, 1 = *negative*). NFC = Need for Cognition scale. Final model: $F(4, 593) = 17.01$, $p < .001$, $R^2_{\text{Adj}} = .103$.

⁵ Related to the concerns about log-transforming the data expressed in Footnote 3, we again used the PROCESS v. 3.5 macro (Hayes, 2018) to obtain the bootstrapped confidence intervals for the exploratory analysis for H₂ and H₃ that we conducted on the entire sample that completed the study and were not excluded due to geocoordinate issues. Results are reported in Table S2 of the Supplemental Materials (see Appendix G).

The predictors entered at Step 1 contributed significantly to the regression model, $F(3, 593) = 17.011, p < .001$, and explained 10.3% of the variation in T2 risk estimates. Again, we did not find support for H2, as contextual information was not a significant predictor of T2 risk estimates, although participants in the negative contextual information condition ($M_{raw} = 33.5, SD = 28.4$) reported higher risk estimates than did participants in the positive contextual information condition ($M_{raw} = 29.9, SD = 27.4$). This result indicates that participants were not significantly influenced by the contextual information after reviewing the risk-relevant Static-99R information. Like the findings in the previous analysis, participants' general risk estimates were significantly associated with their T2 risk estimates. However, in contrast to the prior analysis, participants' NFC scores did have a significant main effect on their T2 risk estimates. The negative coefficient indicates NFC scores were negatively associated with T2 risk estimates, such that as NFC scores decreased, risk estimates increased. Nevertheless, like the previous analysis, adding the interaction term of contextual information and NFC at Step 2 did not significantly improve the model, $F(4, 594) = 0.011, p = .916$. Thus, reanalysis of the data including participants who failed one or more attention checks again failed to support H2 and H3.

Discussion

Our primary aims in this study were to examine whether there was a biasing effect of risk-irrelevant contextual information on jurors' perceptions of SVP respondent sexual recidivism risk, and if so, whether presenting risk-relevant information (i.e., Static-99R actuarial risk estimate) could eliminate the bias, and whether mock jurors' level of NFC moderated the effects of risk-relevant information on biasing risk-irrelevant contextual information.

Effects of Contextual Information on T1 Risk Estimates

We first hypothesized that positive or negative risk-irrelevant contextual information would have a significant effect on mock jurors' estimates of the likelihood that the respondent would sexually reoffend. The results supported our hypothesis that participants who read negative risk-irrelevant contextual information about the respondent rated his likelihood of reoffending at T1 significantly higher than participants in the positive condition. This finding demonstrates that bias created by the perceived social (un)attractiveness of the respondent potentially affects jurors' perceptions of reoffending risk in an SVP context. This finding lends support to a suspected halo effect in operation in jurors' estimates of sexual recidivism risk.

However, we also note that participants' general sexual recidivism risk estimates accounted for a significant portion of the T1 risk estimates, lending support to previous findings

that indicate that public misconceptions about sexual recidivism rates among sex offenders contribute to SVP decision-making (Scurich & Krauss, 2014). In fact, the mean general sexual recidivism risk estimate among our participants far exceeded observed sexual recidivism rates (e.g., Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005; Helmus et al., 2012; Langan et al., 2003). For example, Langan et al. (2003) reported a sexual recidivism rate of 5.3% three years post-release based on arrest data for 9,691 male sex offenders released from 15 state prisons in 1994. In a retrospective study nine years post-release, Alper and Durose (2019) reported an observed sexual recidivism rate of 8% among a sample of 20,195 subjects in 30 states who had been released after serving a prison sentence for rape/sexual assault. Previous meta-analyses reported observed sexual recidivism rates ranging from 7% to 15% with an average follow-up period of five to six years (Hanson & Morton-Bourgon, 2005; Helmus et al., 2012). Furthermore, several studies indicate that longer periods of offense-free time in the community are associated with decreases in observed sexual recidivism rates (A. J. Harris et al., 2003), even among so-called high-risk sex offenders (Hanson et al., 2014).

The finding that general beliefs about the rate of sexual recidivism among people convicted of a sexual offense has a significant effect on perceptions of recidivism risk also extends to judges in both adversarial and inquisitorial settings. Judges appear to hold some of the same attitudes towards people convicted of a sexual offense as the general public (Nhan et al., 2012). In fact, in *McKune v. Lile* (2002), Justice Kennedy asserted, with paltry and questionable evidence (Ellman & Ellman, 2015), that the risk of sexual recidivism is “frightening and high” (p. 34), a phrase which has been cited in at least 100 subsequent lower court decisions (Liptak, 2017).

Mitigating Effects of Actuarial Risk Estimate on T2 Risk Estimates

We examined whether the biasing effect of risk-irrelevant contextual information predicted mock jurors' estimates of reoffending risk after they reviewed the risk-relevant Static-99R information (T2). The majority (63%) of our sample provided a case-specific sexual recidivism risk estimate that exceeded the score-wise risk estimate indicated by the Static-99R. This finding is in line with previous research showing that people tend to overestimate recidivism risk, even when presented with a likelihood percentage (Batastini, Hoeffner, et al., 2019; Varela et al., 2014).

Contrary to our expectations, the risk-irrelevant contextual information did not significantly predict mock jurors' estimates of sexual recidivism risk at T2. The hypothesis that the risk-irrelevant characteristics of the SVP respondent would influence mock jurors' estimates of sexual recidivism risk had an intuitive appeal based on research indicating that

certain characteristics such as physical (Abel & Watters, 2005) and social attractiveness (Alicke & Zell, 2009) affect attributions of guilt and punishment decisions in some criminal contexts. Moreover, previous research has suggested that jurors give relatively little weight to *expert testimony* based on structured risk assessment instruments (Boccaccini et al., 2013; Boccaccini, Murrie, & Turner, 2014; Guy & Edens, 2003; Krauss & Sales, 2001). In other words, we did not expect that the Static-99R report would successfully mitigate the biasing effects of the risk-irrelevant information. However, we are unaware of any previous study that has isolated the effects of risk-irrelevant information and Static-99R results in the manner we did (i.e., presenting only the Static-99R report) on mock SVP jurors' estimates of sexual recidivism risk or civil commitment decisions.

Not only did we eliminate potential effects of the likability or credibility of an expert, but we also communicated the risk estimate to mock jurors in several formats. In particular, previous violence risk assessment research suggests that when jurors are presented with recidivism rates (as compared to categorical estimates), they tend to render significantly lower risk estimates (e.g., Batastini, Hoeffner, et al., 2019; Batastini, Vitacco, et al., 2019), which may account for the substantial shift in recidivism risk estimates toward that indicated by the Static-99R in our study. Hence, although the finding of the effectiveness of the Static-99R information on mitigating bias should be considered tentative, we believe that the null effect of the risk-irrelevant information at T₂ on mock jurors' estimates of sexual recidivism risk suggest a potential debiasing effect of the actuarial risk estimate. This is a remarkable finding, because it runs counter to previous studies that suggest oral or written expert testimony regarding recidivism risk may increase jurors' perceptions of sexual recidivism risk (e.g., Krauss et al., 2012).

Also contrary to previous research, indicating that participants greatly overestimate sexual recidivism risk regardless of the communication format (Varela et al., 2014), our findings indicate a substantial reduction in overall recidivism risk estimates after reviewing the Static-99R information. However, despite the fact that there are some similarities between our study and that of Varela et al. (2014), the current study is distinguishable for several reasons. First, our participants provided a numerical risk estimate, whereas in the Varela et al. study, the participants rated recidivism risk based on Likert scale ratings of (1) the likelihood of re-offense; (2) dangerousness to the community; and (3) support for strict supervision strategies. Based on a composite rating of the first two factors, Varela et al. reported that 95% of their participants indicated the offender was likely to commit a new sexual offense within the next five years. Second, the individual presented in the Varela et al. case vignette had been convicted of two previous rapes, whereas the individual portrayed in our study had been convicted for one previous sexual assault. Moreover, participants in the Varela et al.

study were exposed to only one risk communication format, whereas our participants read a Static-99R report based on the template provided in the Static-99R Evaluator Workbook (Phenix, Helmus, et al., 2016). In fact, previous researchers have suggested that presenting risk estimates through combined methods (e.g., categorical and probabilistic) may improve decision-makers' understanding of a risk estimate (e.g., Hilton et al., 2015).

Relatedly, Scurich and Krauss (2014) conducted an experimental study in which their participants' sexual recidivism risk estimates averaged nearly 74%, substantially higher than T2 risk estimates of 25.6% in the current study. Scurich and Krauss sought to determine whether participants would vote for civil commitment regardless of whether all four legal criteria were met. In their study, participants who received expert testimony in the form of a categorical risk estimate (indicating moderate risk) based on the Static-99R did not provide recidivism risk estimates that differed significantly from participants who received no expert testimony about sexual recidivism risk. However, the difference in the risk estimates in the Scurich and Krauss study and the current study may be explained by several factors. First, Scurich and Krauss presented only a categorical risk estimate communicated through *expert testimony*. In addition, participants read expert testimony about recidivism risk only after they were exposed to other risk-relevant information, such as the SVP respondent's previous convictions and a clinical diagnosis of pedophilia. In combination, the presentation of a categorical risk estimate and other risk-relevant information in the Scurich and Krauss study may explain the large difference between the average risk estimate in their study and the average risk estimates at T2 in our study.

In conclusion, although the significant effect of the Static-99R on sexual recidivism risk estimates and the relatively low risk estimates at T2 in the current study appear to contradict previous research, we have discussed several possible explanations for these differences. These differences include how risk was communicated in most previous studies (i.e., *through expert testimony* and/or by presenting the risk estimate in a single format) and the presence of other risk-relevant factors that make it difficult to discern what factors were influencing jurors' perceptions of sexual recidivism risk. The current study design enabled us to directly examine (1) whether mock jurors' estimates of sexual recidivism risk are affected by risk-irrelevant information and (2) whether providing mock jurors with an actuarial risk estimate in various formats (i.e., categorical, relative risk, and probabilistic risk), free of the influence of potential confounding factors related to expert testimony, was effective in mitigating the influence of bias created by risk-irrelevant characteristics of the SVP respondent.

To our knowledge, this study is the first to directly examine whether an actuarial risk estimate can mitigate the effects of contextual information bias among legal decision-makers. Overall, we believe that failing to find support for our hypothesis that a Static-99R risk

estimate would be ineffective in mitigating the effects of contextual information bias (H2) is encouraging. Although future researchers should attempt to examine this question, we expect that these findings would also extend to judges. Previous research indicates that experts and judges are prone to the same cognitive biases as laypeople (e.g., Guthrie et al., 2001; Liu & Li, 2019; Rassin, 2020). Therefore, if an actuarial risk estimate is effective in mitigating the effects of bias among mock jurors, it is reasonable to expect the same would hold true for judges.

Mock Jurors' Level of Need for Cognition and Decision-Making

We did not find support for the hypothesis that mock jurors' NFC moderates the influence of risk-relevant and risk-irrelevant information on estimates of the SVP respondent's recidivism risk. This null finding suggests that varying levels of NFC may not explain mock jurors' ability or willingness to adjust their perceptions of the respondent's recidivism risk in light of actuarial information. It should be noted, however, that NFC and education level appear to be strongly correlated (McCabe et al., 2010). Our sample was fairly well-educated, with over 50% possessing at least a bachelor's degree. In line with this, the median NFC score among our sample was 47, where the midpoint of the scale was 36. Research suggests that people who have higher levels of NFC may be more inclined to adjust their initial judgments in light of new information (D'Agostino & Fincher-Kiefer, 1992; Gilbert et al., 1988; Martin et al., 1990). Therefore, we could not rule out the possibility that the Static-99R information had a substantial effect in our sample due to participants' relatively high level of NFC.

Because of the significant difference in NFC scores between participants who were excluded for a failed attention check and those who were not, we conducted an exploratory analysis to examine the possibility of a moderating effect of NFC. Although we again found no support for a moderating effect (H3), we did find a significant main effect of level of NFC on mock jurors' risk estimates at T2. While conclusions about this finding are tentative, we note that it is in line with previous research suggesting that those who are lower in NFC may be more prone to heuristic thinking (e.g., Lieberman, 2002; Lieberman et al., 2007) and less influenced by testimony based on ARAIs than on clinical judgment (Krauss et al., 2004). The significant main effect of NFC in the current study further suggests that individual information processing differences among jurors affect their motivation to effectively discriminate between information that is risk-irrelevant and that which is risk-relevant and thereby legally relevant. Additional research on the effect of juror levels of NFC on their capacity to render legally justifiable decisions could have significant implications for juror screening processes.

Limitations

We note several limitations to this study, specifically related to its experimental design and the realities in SVP hearings. First, we presented very limited background information about the SVP respondent. In actual cases, jurors are likely to learn much more about the respondent as they must consider not only the respondent's risk of sexual recidivism. For example, jurors are likely to hear expert testimony about mental health diagnoses, because the presence of a mental abnormality is one of the legal criteria that must be met to justify civil commitment. We did not present any diagnosis to our participants.

Second, the SVP respondent in the case vignette had been convicted of one sexual offense and one previous non-sexual offense. This type of criminal history may not be typical of the offending history of actual SVP respondents, although analysis of actual SVP cases indicates prior criminal history can vary widely (e.g., Lu et al., 2015). Third, although not measured by the Static-99R, some of the characteristics we labelled as "risk-irrelevant" in this study may be relevant in the context of a forensic sexual risk evaluation. For example, the type of employment a person has may be an indicator of their intelligence level and poor relationships with others may be indicative of psychopathy or a personality disorder, which are factors that may have clinical significance for sexual recidivism risk among certain types of offenders (see, for example, Hanson & Morton-Bourgon, 2005; Hawes et al., 2013; Hildebrand et al., 2004; Nijman et al., 2009; Porter et al., 2009) Yet, as presented in our case vignette and without additional context, concluding that there is evidence of low intelligence or a personality disorder would not be justified. In fact, in our study, we found that mock jurors' evaluations of an SVP respondent's sexual recidivism risk appear to be closely tied to their global evaluation of factors affecting the likability of the SVP respondent, such as would be expected by the operation of the halo effect (Nisbett & Wilson, 1977b). Nevertheless, future research might employ a "purer" test of factors that have no potential significance for jurors' sexual recidivism risk estimates (for example, physical attractiveness of the SVP respondent).

We also note that there are several limitations inherent to the mock jury paradigm. For example, decisions are rendered without jury deliberation, and deliberation has been shown to influence jury decision-making (Lynch & Haney, 2009). Hence, jury dynamics and the relative influence of risk-irrelevant contextual information and actuarial risk estimates might be a fruitful area of further study. Finally, we utilized MTurk workers as mock jurors in this study. There may be some differences between MTurk workers and a typical SVP juror. For example, there are a few criteria that may prevent someone from serving on a jury that are not relevant in the context of an online study (e.g., a criminal record). There may also be differences between MTurk workers and "typical" jurors in SVP cases in terms of education level, political values, and age (Paolacci & Chandler, 2014). Yet, MTurk respondents tend to be

diverse in terms of socioeconomic status and ethnicity (Casler et al., 2013), which is a benefit for studying psychological phenomena and individual differences. We also note that nearly 17% of our sample had previously served as a juror.

Conclusion

The current study has shown the potential for jurors to be biased by risk-irrelevant factors in their perceptions of an SVP respondent's sexual recidivism risk. Misjudgments of sexual recidivism risk are likely to undermine the intent of the legal statute that civil commitment be imposed only on offenders at highest risk of sexual reoffending (Carlsmith et al., 2007; Knighton et al., 2014; Krauss & Scurich, 2014; Scurich & Krauss, 2014; Sreenivasan et al., 2003). The risk estimate derived from an ARAI appears to be the most accurate (Ægisdóttir et al., 2006; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009) and most relied upon factor among evaluators who render an expert opinion in SVP cases (Chevalier et al., 2015). We have provided preliminary evidence suggesting that permitting jurors to review the actuarial risk estimate information directly in writing, may encourage them to rely more on scientifically-based information in their judgment of the respondent's sexual recidivism risk, thus limiting the influence of risk-irrelevant contextual information. Nevertheless, the finding from the reported exploratory analysis related to juror level of need for cognition also suggests that how jurors process information may have a significant effect on their capacity to integrate relevant information and render a legally justified verdict in an SVP civil commitment case.

Chapter 5

Asymmetrical Skepticism Toward Actuarial Risk Information and Mock Jurors' Estimates of Sexual Recidivism Risk

This chapter is based on the following paper:

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. (2020).
Asymmetrical Skepticism Toward Actuarial Risk Information and Mock Jurors'
Estimates of Sexual Recidivism Risk. *Manuscript in preparation.*

Abstract

One of the criteria that must be met to civilly commit a person under a so-called “Sexually Violent Predator” (SVP) statute in the United States is that the SVP respondent—the person whom the government seeks to commit—is more likely than not to engage in sexual reoffending. Jurors often make the decision about civil commitment based on expert testimony about the respondent’s sexual recidivism risk based on an actuarial risk estimate (for instance, a sexual recidivism risk estimate based on the Static-99R). We carried out an experiment with U.S. mock jurors ($N = 376$) to examine their ratings of the accuracy and reliability of the Static-99R report when it conflicted with their initial impressions of an SVP respondent. The tendency to selectively devalue evidence that conflicts with initial beliefs is known as *asymmetrical skepticism*. We did not find evidence of asymmetrical skepticism toward the actuarial risk estimate. In fact, presenting mock jurors with a written actuarial risk estimate mitigated the effects of bias toward the SVP respondent stemming from jurors’ perceptions of his likability. Our findings suggest that providing jurors with the written results of an actuarial risk estimate facilitates juror responsiveness to the estimate.

Introduction

In 1984, Leroy Hendricks was convicted for sexual misconduct with two teenaged boys and received an indeterminate prison sentence ranging from five to 20 years for the crimes (*Kansas v. Hendricks*, 1997). Ten years later, Hendricks was about to be released from prison when the State of Kansas sought to have him involuntarily committed as a Sexually Violent Predator (SVP). The Kansas SVP Act permitted involuntary and indefinite civil commitment of a person who: (1) had a history of sexual offending; (2) had a mental abnormality or personality disorder; and (3) due to the mental abnormality or disorder was *likely to engage in future sexual offending* (emphasis added; *Kansas v. Hendricks*, 1997). A jury determined that, beyond a reasonable doubt, Hendricks met the legal criteria within the SVP Act and Hendricks was remanded to Larned State Hospital for psychiatric treatment (Vanginderen, 1998). Hendricks was the first person to be civilly committed under the Kansas SVP Act (D. M. Smith, 2015).

Hendricks challenged the constitutionality of his further confinement after having served his prison sentence. The case made it to the U.S. Supreme Court, which decided that SVP statutes are constitutional as long as they are non-punitive and narrowly restricted to a class of people who are *more likely than not* to sexually reoffend (*Kansas v. Hendricks*, 1997). Hendricks had a long history of sexually abusing children and he admitted he suffered from pedophilia. In fact, Hendricks himself said that it was likely that “to die” was the only way he would stop sexually abusing children (*Kansas v. Hendricks*, 1997, p. 355).

At present, an estimated 6,500 people in the United States are indefinitely confined under SVP or Sexually Dangerous Person (SDP) statutes (Koeppel, 2019), which have been enacted by the federal government and 20 states (Knighton et al., 2014). A decision to civilly commit a person as an SVP has grave consequences, because most who are committed will never be released (B. Cooper & Rizzo, 2013; Gookin, 2007). Leroy Hendricks, who had suffered a stroke in 2007 and was since wheelchair-bound, died in Larned State Hospital in 2009 at the age of 75 (*Baker v. DesLauriers*, 2014; Crosbie, 2018).

However, people whom the government seeks to commit as SVPs often do not evince the type of disorder and offending history as exhibited by Hendricks (Elwood et al., 2010; Jackson & Richards, 2007). Statistically-based sexual recidivism risk estimates play an important role in SVP civil commitment cases to help differentiate sexual offenders who are likely to sexually recidivate from those who are not. In the current study, we examine whether mock jurors exhibit disparate levels of skepticism toward a sexual recidivism risk estimate that depends upon whether the risk estimate is consistent with their first impression of a simulated SVP respondent’s risk of sexual recidivism.

Expert Estimates of Sexual Recidivism Risk in SVP Cases

It is typically left to a jury to decide whether a person against whom an SVP civil commitment petition is brought—also known as the *respondent*—meets the requisite legal criteria for SVP civil commitment (Krauss & Scurich, 2014; Lieberman et al., 2007; Zolfo, 2018). Evidence of the respondent's likelihood of engaging in future sexual offending is typically provided through the testimony of a forensic mental health professional, or SVP evaluator (Jackson et al., 2004; Jackson & Hess, 2007; Scurich & Krauss, 2013). An expert's estimate of the SVP respondent's risk of engaging in sexual recidivism is nearly always based on a structured risk assessment instrument (SRAI; Doren, 2002; Jackson & Hess, 2007; Janus & Prentky, 2003).

In fact, risk estimates derived from an SRAI have been shown to be more accurate than estimates based solely on an evaluator's clinical experience (Bengtson & Långström, 2007; Hanson et al., 2004; Hanson & Bussière, 1998; Jackson et al., 2004; Janus & Meehl, 1997). Specifically, actuarial risk assessment instruments (ARAIs; Barbaree et al., 2001) are commonly used by evaluators to estimate sexual recidivism risk (Janus & Prentky, 2003; Murrie et al., 2009). ARAIs are based on numerical scoring of weighted risk factors that have been shown to be associated with sexual or violent recidivism (Dawes et al., 1989; Doyle & Dolan, 2007; Quinsey et al., 1998b). A total risk score is calculated by summing the individual item scores and is associated with a recidivism risk estimate, usually in terms of low-moderate-high. In the United States, the most commonly used ARAI for estimating sexual recidivism risk is the Static-99R (Helmus et al., 2012; Jackson & Hess, 2007; Krauss et al., 2018), which contains 10 risk factors that are empirically associated with sexual recidivism. Scoring each of these factors leads to a total score, for which the Static-99R developers provide a score-wise risk estimate based on observed recidivism rates among individuals with the same score (Helmus et al., 2012).

Ideally, jurors in SVP cases would make use of a sexual recidivism risk estimate based on an SRAI when deciding on the SVP respondent's sexual recidivism risk. Yet, numerous studies indicate that they tend to give little weight to *expert testimony* about sexual recidivism risk estimates and tend to overestimate the SVP respondent's risk of sexual recidivism (Boccaccini et al., 2013; S. L. Miller & Brodsky, 2011; Pennycook et al., 2014; Walters et al., 2014). However, nearly all studies that have examined the effects of actuarial risk estimates have done so using expert testimony about the risk estimate. Because factors such as the likability of an expert can affect jurors' perceptions of their credibility (e.g., Cramer et al., 2009), it is important to disentangle the qualities of the messenger from the content of the message in SVP civil commitment cases.

Cognitive Bias and Jurors Estimates of Sexual Recidivism Risk

Public perceptions of sexual recidivism risk are particularly relevant to SVP civil commitment proceedings because lay juries decide whether the legal criteria for commitment are met. The general public tends to grossly overestimate the risk of sexual reoffending (Ellman & Ellman, 2015; Katz-Schiavone et al., 2008; Levenson et al., 2007). For example, Levenson and colleagues (2007) reported that among a sample of laypeople, the average estimate was that 74% of persons convicted of a sexual offense would reoffend. In reality, observed sexual recidivism rates among persons convicted of a sexual offense range between 5 and 15% with follow-up periods ranging from three to 10 years or more (Alper & Durose, 2019; Hanson et al., 2014, 2016; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Langan et al., 2003). However, observed recidivism rates appear to vary considerably depending on the type of offender (e.g., sexual offending against children; incest; A. J. Harris & Hanson, 2004).

Moreover, media coverage of particularly heinous sexual crimes, such as those committed by Leroy Hendricks, influences beliefs held by the general public that people who have committed a sexual offense (PCSOs) are stereotypical “deviants” who are strangers to their victims, who repeatedly prey on the most vulnerable victims, and are not amenable to treatment (e.g., Borhart & Plumm, 2015; Katz-Schiavone et al., 2008; Mesler et al., 2016; Quinn et al., 2004; Vess & Skelton, 2010; Yung, 2010). Therefore, when members of the public, or jurors, hear a term such as “sexually violent predator” or “sex offender,” it can prompt heuristic processing of information based on a sex offender stereotype (Scurich et al., 2016).

Heuristic information processing relies upon one’s previous life experiences and expectations for decision-making, particularly when estimating the likelihood of an uncertain event (Tversky & Kahneman, 1974). This can lead to errors in logic or reasoning as a result of cognitive bias (Tversky & Kahneman, 1982). For example, the degree to which an individual is considered to be representative of a group (e.g., sex offenders)—the *representativeness heuristic*—can influence whether an individual is believed to be a member of the group (Tversky & Kahneman, 1974). Therefore, the extent to which an SVP respondent “fits” the sex offender stereotype based on characteristics that are unrelated to sexual recidivism risk may bias jurors’ perceptions of the respondent’s risk of sexual recidivism.

Risk-Irrelevant Contextual Information and Jurors’ Estimates of Sexual Recidivism Risk

Previous studies indicate that information that is irrelevant to risk can bias individuals’ perceptions of violent or sexual recidivism risk and subsequent decision making (Gunnell

& Ceci, 2010; Hilton et al., 2005, 2015; Kamorowski, Ask, et al., 2021; Korva et al., 2013). For example, in a mock juror study, Kamorowski and colleagues (2021; Chapter 4 of this dissertation) examined the effects of the Static-99R on case-specific sexual recidivism risk estimates in an SVP civil commitment context. They found that although a Static-99R risk estimate mitigated the influence of risk-irrelevant offender characteristics on jurors' estimates of the offender's sexual recidivism risk, the Static-99R estimate did not eliminate the effects completely.

In fact, cognitive bias can diminish the ability of people to exercise reasoning and judgment that is independent of irrelevant contextual information (e.g., Dror, 2012; Dror et al., 2006; Edmond et al., 2015; Nakhaeizadeh et al., 2014) and their preexisting beliefs. Furthermore, bias affects how subsequent information is evaluated and interpreted (Nickerson, 1998). For example, if SVP jurors' initial perceptions of the SVP respondent's sexual recidivism risk are biased, this may affect jurors' perceptions about the accuracy and reliability of an actuarial risk estimate that does not comport with their initial perceptions. Jurors' biased evaluation of actuarial risk assessment testimony may help clarify why they seem relatively insensitive to an expert's estimate of an SVP respondent's sexual recidivism risk. We examine whether asymmetrical skepticism toward the results of an SRAI may be an explanatory mechanism for jurors' apparent insensitivity to an actuarial risk estimate.

Asymmetrical Skepticism

Asymmetrical skepticism is the tendency to interpret preference-consistent information as accurate and preference-inconsistent information as faulty (Ditto et al., 2003; Ditto & Lopez, 1992). Asymmetrical skepticism may therefore act as a mechanism that determines how much weight jurors give to an actuarial estimate of sexual recidivism risk. People have a tendency to interpret information in a way that is consistent with their initial beliefs (Ditto et al., 1998; Ditto & Lopez, 1992; Kunda, 1990; Pyszczynski & Greenberg, 1987). In fact, when people encounter information that is not consistent with their preferences, they tend to engage in more critical analysis of the content of that information, whereas they tend to uncritically accept information that aligns with their preferences (Ditto & Lopez, 1992). Furthermore, when engaging in critical analysis of information that contradicts their preference, people tend to conclude that the information is flawed or lacking in accuracy (Ditto et al., 2003; Ditto & Lopez, 1992).

For example, in an experimental study, Ask and Granhag (2007) found clear evidence of asymmetrical skepticism when they presented police officers ($N = 49$) with an initial case vignette that was strongly suggestive of a suspect's guilt in a homicide, followed by a witness

statement that either supported the hypothesis of the suspect's guilt (incriminating) or cast doubt on the suspect's guilt (exonerating). The results showed that the officers rated the witness who provided incriminating evidence as significantly more credible and reliable than the witness who provided information that appeared to exonerate the suspect, even though they all received the same information related to the witness's background and memory recall ability. These findings indicate that trained investigators were biased in their evaluation of the evidence by their preexisting belief in the suspect's guilt.

The effects of asymmetrical skepticism on the interpretation of evidence has been replicated in several studies (e.g., Ask et al., 2008, 2011; Marksteiner et al., 2011), but to our knowledge, has not been examined with respect to jurors' estimates of sexual recidivism risk in an SVP case. In fact, relatively few studies related to SVP civil commitment and sexual recidivism risk perceptions have contemplated that jurors must process risk-relevant (e.g., testimony about an actuarial risk estimate) and risk-irrelevant information (e.g., the SVP respondent's likability), and that this process is likely to affect the "weight" both risk-relevant and risk-irrelevant factors have on jurors' perceptions of sexual recidivism risk (e.g., Kamorowski, Ask, et al., 2021). If jurors are biased by information that is irrelevant to sexual recidivism risk, this may affect how much weight they give an actuarial risk estimate and how that estimate is interpreted. Therefore, in this study, we examined whether there is evidence of asymmetrical skepticism toward a Static-99R risk estimate among mock jurors in an SVP case when it conflicts with their initial impressions of the SVP respondent.

The Current Study

Because research has yet to clearly identify why jurors in SVP cases seem relatively insensitive to actuarial risk estimates, we sought to examine whether mock jurors would exhibit asymmetrical skepticism toward a Static-99R risk estimate. To test this, we analyzed participants' accuracy and reliability ratings of the Static-99R risk estimate as function of the type of irrelevant contextual information they received about the SVP respondent (positive or negative) and the risk level indicated by the Static-99R (below average risk or above average risk). Therefore, if participants' accuracy and reliability ratings of the Static-99R risk estimate were lower when it contradicted previous information they had received about the SVP respondent, this would be evidence supportive of the operation of asymmetrical skepticism.

First, we wished to examine whether mock jurors' perceptions of the likelihood that an SVP respondent would sexually reoffend were affected by risk-irrelevant contextual information about the respondent. We hypothesized that participants who received negative risk-irrelevant contextual information about the respondent, as contrasted with positive

contextual information, would report higher estimates of reoffending risk both before (Time 1; H1a) and after (Time 2; H1b) reviewing the Static-99R risk estimate.

Second, we wanted to determine whether asymmetrical skepticism toward the Static-99R risk estimate explained variation in participants' estimates of the respondent's risk of sexual recidivism at Time 2. We predicted that participants would exhibit asymmetrical skepticism regarding the Static-99R risk estimate, as evidenced by participants' reliability and accuracy ratings of the Static-99R risk estimate as applied to the SVP respondent, such that participants in the negative contextual information condition would exhibit a preference (i.e., higher reliability/accuracy ratings) for the Static-99R risk estimate indicating high risk, and participants in the positive information condition would exhibit a preference for the Static-99R risk estimate indicating low risk (H2).

Finally, we sought to assess whether asymmetrical skepticism was related to participants' adjustment of their perceptions of the respondent's likelihood of sexual reoffending from Time 1 (T1) to Time 2 (T2). We expected that higher levels of skepticism toward the Static-99R risk estimate (i.e., lower reliability/accuracy ratings) would predict smaller adjustments in participants' T2 risk estimates in the direction of the risk estimate indicated by the Static-99R risk estimate (H3).

The preregistration for this study can be viewed at the following link: https://osf.io/tge2z/?view_only=92e3643d479648f481a55dcc1caa302d.

Method

Participant Recruitment

Participants were recruited through Amazon Mechanical Turk (MTurk) and were required to be at least 18 years of age, proficient in English, and reside in the United States. We limited participation to U.S. residents for two reasons. First, we expected that Americans may have different views regarding how people who have committed a sexual offense should be managed, compared to people in other countries. Second, the SVP civil commitment system in the United States is different than other countries, particularly with respect to the fact that jurors often decide whether a person convicted of a sexual offense should be civilly committed (Janus, 2013; Witt & DeMatteo, 2019). We also required that participants have at least a 99% approval rating on 500 or more tasks on MTurk and must not have participated in the pretesting phase of this study, in which we evaluated the strength of the manipulation, or in a previous study we conducted using a comparable design and stimulus materials.

We planned to exclude participants who did not meet the inclusion criteria, who did not complete the study, who failed one or more attention checks, or indicated that their data

should not be used (Meade & Craig, 2012). We also planned to exclude participants whose total time spent on the study was more than three times the median absolute deviation below the sample median (Leys et al., 2013), but none met this latter criterion.

We also planned to exclude data from participants whose T1 or T2 reoffending risk estimates were more than three times the median absolute deviation (3MAD) above or below the corresponding sample median (Leys et al., 2013). None of the participants met this exclusion criterion based on their T1 risk estimates. However, 3MAD calculations for T2 risk estimates indicated we should exclude data from participants whose T2 risk estimate was 48% or higher, which would have resulted in removing 80 participants (21.3% of the total sample). This rate of exclusion is far above what would be considered as representative of true outliers. Therefore, we did not exclude this group of participants from the main analyses. Instead, we used a rank transformation with the average for ties to correct for skewness in T2 risk estimates.

Power Analysis

An *a priori* power analysis was conducted using G*Power 3.1 (Faul et al., 2009) for an ANOVA to test for main effects and interactions. The parameters were set as $f = 0.25$ to detect a medium effect, $\alpha = .05$, power = .80, and total number of predictors = 4. The minimum number of participants required was 351. We planned to terminate data collection once we reached the sample size of 351 participants after excluding participants based on exclusion criteria.

A total of 556 began the study, but 20 of these were automatically screened out because they indicated they did not meet one or more of the inclusion criteria. Another 30 did not complete the study, yielding a completion rate of 94.4%. Of the participants who completed the study, 99 were excluded for failing one or both attention checks and one was excluded for nonresponsive answers to free-text questions. We excluded an additional 28 participants for metadata that indicated geo-coordinate locations outside the United States. Additionally, two participants indicated we should not use their data. Our remaining sample size was 376.

Participant Demographics

Of the participants whose data were analyzed, 190 identified as male (50.5%), 182 identified as female (48.4%), one identified as a gender variant/non-conforming (0.3%), and three declined to provide gender identity information (0.8%). Participants ranged in age from 21 to 71 years ($M = 38.5$, $SD = 11.3$). Two participants had less than a high school diploma (0.5%), 38 were high school graduates (10.1%), and 89 had some college education (23.7%). Forty-five participants held an associate's degree (12.0%), 150 had a bachelor's degree (39.9%), 40 had

a master's degree (10.6%), two held a doctoral degree (0.5%), and seven had a professional degree (e.g., MD, JD; 1.9%). Nearly every U.S. state ($n = 47$) was represented, of which the highest number of participants resided in Florida ($n = 35$; 9.3%). Eleven participants did not provide their state of residence. Forty-seven participants (12.5%) had previously served as a juror in a legal case.

Procedure and Materials

Participants self-selected to participate in the study as advertised on MTurk. The study was conducted online via Qualtrics. Participants were paid 1 USD to participate and a bonus of 2 USD if they correctly answered both attention check questions. This study was approved by the ethical review committee of the Faculty of Psychology and Neuroscience of Maastricht University, reference 207_14_04_2019.

Data Collection

Informed consent was obtained from participants, and they were asked to confirm that they met the eligibility criteria (*yes/no*). The study was automatically terminated for any participant who indicated they did not meet one of the criteria. Participants then read a brief introduction about the study and received contact information of the first author in the event of any issues with the study. Participants were randomly assigned to one of the four cells in a 2 (contextual information: positive vs. negative) \times 2 (Static-99R risk estimate: below average risk vs. above average risk) between-subjects design.

We used several controls for quality assurance. First, participants were not allowed to go back to a previous page once they moved on to the next one. Second, a timer was used to track how much time they spent on each page (participants were not aware that they were being timed). Finally, the “next” button to move forward was also delayed on most pages (participants were informed of this) in an effort to encourage participants to read the information presented and prevent “clicking through” the study.

General Risk Estimate. Before participants read the case vignette, they provided an estimate of how likely they thought it was that a person who has been convicted for a sexual offense would commit another sexual offense, using a graphic sliding scale ranging from 0–100% in 1% increments (hereafter, general risk estimate).

Case Vignette. Participants were randomly assigned to read a case vignette containing positive ($n = 175$) or negative ($n = 201$) risk-irrelevant contextual information about an SVP respondent named John Smith (see Appendix C). In this study, we defined risk-irrelevant contextual information as information that has not been shown to be empirically related to sexual reoffending risk (e.g., type of job, length of employment, perceptions of others

about the respondent's sociability, preferred leisure activities, and marital history). The two case vignettes varied with respect to the previously mentioned risk-irrelevant contextual information. The effectiveness of the manipulation had been confirmed through participant ratings of Mr. Smith's likability in pretesting.

Likability Ratings. To obtain an untainted measure of the effectiveness of the manipulation (i.e., participants' perceptions of Mr. Smith's likability), we initially provided only risk-irrelevant information about Mr. Smith and made no mention that he had been convicted of a sexual offense, nor that he was an SVP respondent in a civil commitment case. Likability ratings were based on participants' perceptions of Mr. Smith's: (a) likability, (b) friendliness, and (c) warmth, which were reproduced from the Reysen Likability Scale (Reysen, 2005).

The Reysen Likability Scale is comprised of 11 items rated on a 7-point Likert scale (1 = *very strongly disagree* to 7 = *very strongly agree*). Lower ratings indicate the target is less likable, and higher ratings indicate the target is more likable. For this study, we used a 6-point Likert scale (1 = *completely disagree*, 2 = *mostly disagree*, 3 = *somewhat disagree*, 4 = *somewhat agree*, 5 = *mostly agree*, and 6 = *completely agree*) to eliminate the *neutral* option. Cronbach's α for the original scale was .91 (Reysen, 2005). Cronbach's α in this sample for the three items measured was .98. We calculated the average rating of the three likability factors to create a composite variable for statistical analysis of the manipulation's effectiveness.

Case Background and Crime Information. Participants then read the remainder of the case vignette describing the offense for which Mr. Smith was currently incarcerated (i.e., rape of a female acquaintance), his criminal history (one previous conviction for assault related to a bar fight when he was 18 years old), public opinion about the length of his prison sentence (5 years) for the sexual assault, and the victim's opinion about the fairness of Mr. Smith's prison sentence. The two criminal convictions are relevant to sexual recidivism risk, yet the contribution of the criminal history to sexual recidivism risk is accounted for by the Static-99R and did not vary between the positive and negative case vignettes.

However, the additional risk-irrelevant factors of the public's and victim's opinion of the length of the prison sentence differed between the positive and negative vignette conditions. In the positive vignette, public opinion and that of the victim was that the prison sentence was a fair punishment. In the negative vignette, public opinion and that of the victim was that the prison sentence was not sufficient punishment.

SVP Civil Commitment Information. Participants were then instructed to imagine that they had been selected to be a juror in a civil commitment hearing. We provided background information about sex offender civil commitment, including the nature and purpose of civil commitment and the process generally followed to determine whether an

offender should be subjected to civil commitment. To ensure attention and comprehension of the explanation of civil commitment, we used a multiple-choice attention check question of: “Confinement in a secure facility under SVP laws most often occurs: (a) *while a person is serving their prison sentence*; (b) *before a person is sentenced to prison*; or (c) *after a person has served their prison sentence*.” Participants were also provided with an explanation of the legal criteria that must be met for SVP civil commitment (see Appendix D).

Time 1 (T1) Sexual Recidivism Risk Estimate and Disposition

Recommendation. Subsequently, participants were instructed to provide an initial risk estimate, based on the information they considered relevant, of how likely they thought it was that Mr. Smith would commit another sexual offense (T1 risk estimate; 0–100% using a graphic sliding scale). Next, participants provided an initial disposition recommendation (T1 disposition recommendation) for Mr. Smith based on all the information they considered relevant (1 = *released to the community with no conditions*, 2 = *released to the community under supervision only*, 3 = *release to the community under supervision and mandated sex offender treatment*, 4 = *civil commitment*). They were also informed that they would later receive additional information, after which they would be asked to make a second estimate of the likelihood that Mr. Smith would commit another sexual offense and subsequent disposition recommendation. Participants’ disposition recommendations at T1 are reported in Supplemental Materials (see Appendix H, Figure S1).

Static-99R Information. The Static-99R comprises 10 risk factors that are empirically related to the likelihood of sexual offense recidivism (Phenix, Fernandez, et al., 2016). An evaluator codes each of the 10 factors and the resulting total score is associated with an estimated likelihood of reoffending based on previous recidivism rates of people convicted of a sexual offense who were released into the community (Knighton et al., 2014). We operationalized risk-relevant information as Mr. Smith’s Static-99R (Helmus et al., 2012) risk estimate.

Similar to the approach taken by Varela and colleagues (2014) in their study about risk communication formats, we informed participants about risk assessment tools and their use in estimating recidivism risk and we provided information about the development of the Static-99R and how a risk score is calculated using the Static-99R (see Appendix F). We also provided participants with a list of the items contained in the Static-99R (see Table 5.1). The format of the risk score information was provided exactly as outlined in the template recommended in the Static-99R and Static-2002R Evaluator’s Workbook (Phenix, Helmus, et al., 2016, p. 30).

Table 5.1 Risk Factors of the Static-99R

1. Age at release from index offense
2. Ever lived with a lover (at least 2 years)
3. Index non-sexual violence
4. Prior non-sexual violence
5. Prior sex offenses
6. Four or more prior sentencing dates (excluding index)
7. Any convictions for non-contact sex offenses
8. Any unrelated victims
9. Any stranger victims
10. Any male victims

Note. Adapted from (Phenix, Helmus, et al., 2016). Used with permission from the authors.

In the “low risk” condition, participants received Static-99R results for Mr. Smith including his absolute score (0), risk category (below average risk), the 5-year reoffending rate ranges associated with that risk category based on a routine sample (3.3% to 4.7%), and the information in a frequency format presented both with the number of people with the same Static-99R score out of 100 expected to be charged or convicted for a new sexual offense within five years (between 3 and 5 expected and 95 to 97 not expected to incur a new sexual offense charge or conviction; Phenix, Fernandez, et al., 2016).

In the “high risk” condition, participants received the same information, except for a difference in the Static-99R score (4), risk category (above average risk), the 5-year reoffending rate ranges for a score of 4 (10%–12.1%), and a frequency format of 10 to 12 out of 100 people with the same score expected to be charged or convicted for a new sexual offense, and 88 to 90 people who were not expected to be charged or convicted for a new sexual offense. An attention check question followed in which participants had to type in Mr. Smith’s Static-99R score.

Time 2 (T2) Sexual Recidivism Risk Estimate and Disposition

Recommendation. Participants’ T2 risk estimates (0–100% on graphic sliding scale) and disposition recommendations (selecting from the same options as presented at T1) were obtained after they had reviewed the Static-99R risk estimate. In addition, participants rated how important several factors were to their estimate of the likelihood of the respondent’s sexual reoffending, including his: (a) relationships with others, (b) employment history, (c) criminal history, (d) impact of the crime on the victim, (e) length of prison sentence, and (f) Static-99R score, using a 5-point Likert scale (*1 = not at all important, 2 = slightly important, 3 = moderately important, 4 = very important, 5 = extremely important*). Using the same scale, participants rated the importance of each of the aforementioned factors to their disposition recommendation, as well as one additional factor: the participant’s own

estimate of the likelihood of the respondent committing another sexual offense. Participants' disposition recommendations at T2 are reported in Supplemental Materials (see Appendix H, Figure S2). Participants' importance ratings of the various factors to their recommendations for civil commitment are reported in the Supplemental Materials (see Appendix H, Table S1).

Accuracy and Reliability Ratings of the Static-99R. To obtain participants' perceptions of the reliability and accuracy of the Static-99R with respect to Mr. Smith, we asked them to rate the accuracy and reliability on a 6-point Likert scale (1 = *completely unreliable*, 6 = *completely reliable* and 1 = *completely inaccurate*, 6 = *completely accurate*). The definitions of reliability and accuracy were provided to participants as follows: "Reliability in this question means how much you think the results can be trusted" and "Accurate in this question means how correct you think the results are". Participants were also asked to provide an open-ended explanation for the reliability and accuracy ratings they gave.

Demographic Characteristics of the Sample. Participants were then asked if they had ever served on a jury in a legal case and to provide their age, gender, level of education, and state of residence.

Debriefing and Quality Checks. Participants read the following statement and responded with *yes* or *no*: "It is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise, years of effort (the researchers' and the time of other participants) could be wasted. You will receive credit for this study no matter what. In your honest opinion, should we use your data in our analyses in this study?" (Meade & Craig, 2012, pp. 441–442). Participants were then provided with debriefing information and we provided them with debriefing information about the nature and purpose of the study and the research questions we hoped to answer.

Data Analysis

We conducted confirmatory analyses in accordance with our preregistration. Where data transformations were necessary, we have noted these in the respective sections. Data analyses were conducted using SPSS version 26 and *jamovi* version 1.2. We used an alpha level of .05 for all statistical tests. Assessments of effect sizes as small, medium, or large rest on the conventions proposed by Cohen (1988).

Results

Manipulation Check

An independent samples *t*-test indicated that, participants in the positive condition rated the respondent as significantly more likable ($M = 5.23$, $SD = 0.72$) than did participants in the

negative condition ($M = 2.10$, $SD = 0.87$), $t(374) = 37.98$, $p < .001$, Hedges' $g = 3.91$ (large effect), 95% CI [3.57, 4.26]. This result confirms that before reading about the sexual assault, there was a large difference between the groups in the perceived likability of Mr. Smith, indicating that the contextual information manipulation was effective.

General Risk Estimates

Overall, participants' general risk estimates ($M = 65.4\%$; $SD = 21.0$) were much higher than actual observed sexual recidivism rates of 5–15% (Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Helmus, et al., 2016). An independent samples t -test confirmed that there was no significant difference in the general risk estimates of participants in the positive contextual information condition compared to those in the negative contextual information condition, $t(374) = -0.47$, $p = .637$, Hedges' $g = 0.05$, 95% CI [-0.15, 0.25]. Nevertheless, in the subsequent analysis of covariance (ANCOVA), we controlled for participants' general risk estimates in evaluating the effects of the contextual information on T1 risk estimates.

The Effect of Risk-Irrelevant Contextual Information on T1 Risk Estimates

We conducted a one-way ANCOVA to evaluate the difference between the positive and negative contextual information on T1 risk estimates while controlling for participants' general risk estimates (H1a). There was a significant, medium- to large-sized effect of contextual information on T1 risk estimates, $F(1, 372) = 47.50$, $p < .001$, $\eta^2_p = .113$, 90% CI [.067, .164]. In support of H1a, participants in the negative contextual information condition rated the sexual recidivism risk significantly higher ($M_{raw} = 54.6$, $SD = 24.1$) than did participants in the positive contextual information condition ($M_{raw} = 40.1$, $SD = 23.1$), $t(374) = -5.68$, $p < .001$, Hedges' $g = 0.59$, 95% CI [0.38, 0.79].⁶ The analysis also showed that participants' general sexual recidivism risk estimates accounted for a significant and large portion of the variance in T1 risk estimates, $F(1, 372) = 239.89$, $p < .001$, $\eta^2_p = .391$, 90% CI [.330, .446]. This finding suggests that without further information on which to estimate risk, participants relied strongly on their own general risk estimate to anchor the case-specific risk estimate.

The Effect of Risk-Irrelevant Contextual Information on T2 Risk Estimates

To test the effects of contextual information and Static-99R risk estimate on T2 risk estimates (H1b), we conducted a 2 (contextual information: positive vs. negative) \times 2 (Static-99R

⁶ Reported effect size does not account for the covariate.

risk: low vs. high) ANCOVA while controlling for participants' general risk estimates. The T2 risk estimates were positively skewed in a visual inspection of the data. Furthermore, a regression analysis using the raw values resulted in non-normally distributed residuals and considerable heteroscedasticity. Therefore, before carrying out the ANCOVA, we used a rank transformation and used the average for ties as specified in the preregistration for this study (see Table 5.2 for raw means).⁷

Table 5.2 Means (and Standard Deviations) for T2 Risk Estimates, Accuracy-Reliability Ratings, and Change in Risk Estimates

Variable	Positive contextual info		Negative contextual info	
	Low risk (n = 92)	High risk (n = 83)	Low risk (n = 106)	High risk (n = 95)
T1 risk estimate	38.51 (22.52)	43.10 (23.60)	54.93 (23.76)	54.18 (24.66)
T2 risk estimate	12.80 (16.01)	33.75 (28.14)	15.60 (18.20)	40.84 (30.26)
Accuracy-reliability composite rating	4.55 (0.88)	4.17 (0.87)	4.35 (0.89)	4.15 (1.00)
Change in risk estimate between T1 and T2	-25.71 (19.70)	-9.35 (24.70)	-39.33 (23.53)	-13.37 (24.67)

Note. Reported means for the T2 risk estimate are based on raw (untransformed) values.

Contrary to our hypothesis, there was no significant main effect of contextual information on T2 risk estimates, $F(1, 371) = 3.50, p = .062, \eta^2_p = .009, 90\% \text{ CI } [.000, .032]$. Thus, although the group means differed in the predicted direction, H1b was not supported. However, there was a significant, medium- to large-sized effect of the Static-99R risk estimate, $F(1, 371) = 153.33, p < .001, \eta^2_p = .292, 90\% \text{ CI } [.229, .348]$. Participants in the Static-99R high risk condition overall provided a T2 risk estimate ($M_{\text{raw}} = 37.53, SD = 29.43$) that was significantly higher than that of the participants in the low-risk condition ($M_{\text{raw}} = 14.30, SD = 17.23$; see Table 5.2). The interaction term between the contextual information and the Static-99R risk estimate was not significant, $F(1, 371) = 0.26, p = .612, \eta^2_p = .001, 90\% \text{ CI } [.000, .012]$. These results indicate that at T2 contextual information no longer had a significant effect on risk estimates, whereas the Static-99R risk estimate did.

⁷ Although rank-transformed data has been recommended for use in factorial ANOVA (Iman, 1974), others have expressed concern that it complicates the interpretation of the results in complex designs (Sawilowsky, 1990). To make sure that our findings are robust to variations in the analytic approach, we conducted parallel analyses using untransformed data, rank-transformed data, log-transformed data, as well as generalized linear model regressions with gamma distribution (log link) and negative binomial distribution (logit link). As reported in the Supplemental Materials, the method of analysis had little influence on the results and did not lead to substantively different interpretations.

We also note that participants' general risk estimates accounted for a significant portion of the variance in T2 risk estimates, $F(1, 371) = 56.64, p < .001, \eta^2_p = .132, 90\% \text{ CI } [.083, .184]$. Nevertheless, these findings suggest that the Static-99R risk estimate had a mitigating effect on participants' general risk estimates as these estimates accounted for substantially less variance in T2 risk estimates than did the Static-99R risk estimate.

Asymmetrical Skepticism Regarding Static-99R Information

We expected that participants in the negative information condition would assign higher reliability and accuracy ratings for the Static-99R estimate indicating high risk relative to participants in the positive information condition (H2). Conversely, we expected that participants in the positive information condition would exhibit a stronger relative preference for Static-99R risk estimate indicating low risk. A correlational analysis revealed that accuracy and reliability were strongly correlated with one another, $r = .847, p < .001$. Therefore, to create an accuracy–reliability composite variable, we averaged the two ratings together, which became the outcome variable in a 2 (contextual information: positive vs. negative) \times 2 (Static-99R risk: low vs. high) ANOVA.

The predicted interaction between the contextual information condition and the type of Static-99R risk estimate was not significant, $F(1, 372) = 0.79, p = .375, \eta^2_p = .002, 90\% \text{ CI } [.000, .017]$. Therefore, we did not find support for asymmetrical skepticism regarding the Static-99R risk estimate (H2). Furthermore, there was no significant main effect of the contextual information on accuracy and reliability ratings of the Static-99R risk estimate, $F(1, 372) = 1.39, p = .239, \eta^2_p = .004, 90\% \text{ CI } [.000, .021]$. However, there was a small, significant main effect of the Static-99R risk estimate, $F(1, 372) = 9.46, p = .002, \eta^2_p = .025, 90\% \text{ CI } [.005, .056]$. The Static-99R was rated as significantly more accurate and reliable when the result indicated "low risk" ($M_{\text{raw}} = 4.44, SD = 0.06$) compared to a result that indicated "high risk" ($M_{\text{raw}} = 4.16, SD = 0.07$; see Table 5.2).

Adjustments in Estimates of Sexual Recidivism Risk

Our third and final hypothesis (H3) was that when participants were more skeptical of the Static-99R risk estimate (i.e., provided lower accuracy and reliability ratings), there would be a relatively smaller difference between their T1 risk estimate and T2 risk estimate in the direction of the Static-99R risk estimate. We calculated the change in risk estimates by subtracting the T1 risk estimate from the T2 risk estimate. Hence, a negative number indicated a decrease in the risk estimate between T1 and T2, whereas a positive number indicated an increase in risk estimate between T1 and T2, which was the outcome variable for the analysis. The mean change in sexual recidivism risk estimates is reported by group in Table 5.2.

We conducted a multiple regression analysis that included the accuracy-reliability rating, the Static-99R risk estimate, and the interaction term. However, in a slight change from our preregistered analysis plan, we also included contextual information and an interaction term of the Static-99R Risk Estimate \times Contextual Information.⁸ The results are presented in Table 5.3.

Table 5.3 Accuracy/Reliability Ratings, Static-99R Risk Estimate, and Contextual Information as Predictors of Change in Risk Estimates between T1 and T2

Predictor	<i>b</i>	95% CI	<i>SE_b</i>	β	<i>p</i>
Accuracy/reliability rating	-0.82	[-4.52, 2.88]	1.88	-0.029	.663
Static-99R risk estimate	16.10	[9.10, 23.11]	3.56	0.309	< .001
Contextual information	-13.78	[-20.35, -7.22]	3.34	-0.265	< .001
Accuracy/reliability \times Static-99R	0.37	[-4.83, 5.58]	2.65	0.009	.888
Contextual info \times Static-99R	9.78	[0.28, 19.29]	4.84	0.164	.044

Note. *N* = 376. Static-99R information was dummy-coded (0 = low risk, 1 = high risk). Contextual information was dummy-coded (0 = positive, 1 = negative). Model $F(5, 370) = 19.71$, $p < .001$, $R^2_{\text{Adj.}} = .200$, 90% CI [0.133, 0.254].

Contrary to our hypothesis, the interaction between the accuracy and reliability factor and the Static-99R risk estimate was not a significant predictor of change in risk estimates between T1 and T2. Therefore, we found no support for H3. There was a significant but small interaction effect of the contextual information and the Static-99R risk estimate. This finding indicates that the type of contextual information combined with the Static-99R risk estimate predicted the extent to which participants adjusted their estimate of sexual recidivism risk between T1 and T2. As can be seen in Table 5.2, the largest reduction in risk estimates occurred when the negative contextual information condition was paired with a low Static-99R risk estimate.

Overall, when participants received a high-risk Static-99R estimate, they made significantly smaller reductions in their estimates of the respondent's sexual recidivism risk ($M_{\Delta\text{RiskEst}} = -11.48$, $SD = 24.70$) than did participants who received low risk Static-99R information ($M_{\Delta\text{RiskEst}} = -33.00$, $SD = 22.82$). Moreover, participants in the negative contextual information condition made significantly larger adjustments to their risk estimates ($M_{\Delta\text{RiskEst}} = -27.04$, $SD = 27.31$) than did participants in the positive condition ($M_{\Delta\text{RiskEst}} = -17.95$, $SD = 23.61$). Taken together, these findings suggest that the Static-99R risk estimate was effective in mitigating the biasing effect of the contextual information.

⁸ In the preregistered analysis plan, we did not include the contextual information or an interaction term between the contextual information and the Static-99R risk estimate as predictors in the regression analysis. The omission of those predictors was an error because the contextual information and the interaction term need to be included to test for the presence of asymmetrical skepticism.

Nevertheless, as indicated by the mean risk estimates provided at T2 (see Table 5.2), participants in all conditions on average estimated the SVP respondent's sexual recidivism risk as higher than that indicated by the Static-99R risk estimate. Specifically, the low-risk Static-99R estimate was 3.3%–4.7%, yet the raw mean estimate of the low Static-99R group was 14.3% ($SD = 17.23$). The high-risk Static-99R estimate was 10%–12.1%, yet the raw mean estimate of the high Static-99R group was 37.5% ($SD = 29.43$).

Discussion

Mock Jurors' Risk Estimates

In this study, we examined whether risk-irrelevant contextual information influenced mock jurors' perceptions of an SVP respondent's risk of sexual recidivism (hereafter, referred to as mock jurors' risk estimates) both before and after receiving a Static-99R report. We expected that mock jurors who received negative contextual information about the SVP respondent would estimate his sexual recidivism risk as significantly higher than participants who received positive contextual information. We found support for this prediction in mock jurors' risk estimates before receiving the Static-99R report.

This finding indicates that the extralegal and risk-irrelevant factors presented about the SVP respondent had a biasing effect on mock jurors' estimates of his risk of sexual recidivism. In the absence of risk-relevant information, the jurors may have relied upon factors that suggested a more or less "stereotypical sex offender," which reinforces the importance of providing SVP jurors with a scientifically-based sexual recidivism risk estimate.

However, we should also note that mock jurors' estimates of the SVP respondent's sexual recidivism risk after reading the case vignette were significantly lower than their own initial estimates about the general risk of sexual recidivism among PCSOs. One possible explanation is that because the case vignette depicted the sexual assault of a female acquaintance, it did not fit the "rape stereotype" of an adult male stranger who uses physical force to sexually assault a female victim (Chapleau & Oswald, 2013) and tends to be perceived as more dangerous compared to an adult male who sexually assaults an acquaintance (Borhart & Plumm, 2015; Levenson et al., 2007; Socia et al., 2019).

After mock jurors had been presented with the risk-relevant Static-99-R report, however, the effect of risk-irrelevant contextual information was no longer significant, a result that failed to support our hypothesis. Instead, mock jurors' risk estimates were strongly influenced by the risk level indicated by the Static-99R. This finding contrasts with previous studies, which showed actual and mock jurors to be relatively insensitive to actuarial risk estimates (e.g., Boccaccini et al., 2013; Krauss et al., 2012). The current finding is encouraging

and suggests that actuarial risk estimates, when presented in an accessible way, may serve to limit the influence of biasing contextual factors.

In the present study, and also in previous research (e.g., Kamorowski, Ask, et al., 2021; Katz-Schiavone et al., 2008; Levenson et al., 2007; Scott et al., 2010; Varela et al., 2014), laypersons' general estimates of recidivism among PCSOs tend to be substantially higher than the observed sexual recidivism rates of 5–15% (Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Helmus, et al., 2016). We further note that in the present study, mock jurors' general recidivism risk estimates among PCSOs accounted for a significant portion of the variance in their estimates of the SVP respondent's risk of sexual recidivism both before and after receiving a Static-99R report. One possible explanation for this could lie in an "anchoring effect." The *anchoring effect* is a form of cognitive bias in which an initial value has a disproportionate effect on subsequent probability estimates (Tversky & Kahneman, 1974). The anchoring effect has been extensively studied and found to have a significant effect on probability estimates and legal decisions (for a review, see Furnham & Boo, 2011). As previous research has indicated a relative insensitivity to actuarial risk estimates among both mock jurors and actual jurors (e.g., Boccaccini et al., 2013; Krauss et al., 2012), our findings suggest that mock jurors' own preexisting beliefs about the sexual reoffending rates among PCSOs may limit the influence of actuarial risk estimates.

Perceived Accuracy and Reliability of the Static-99R

We predicted that we would find evidence of asymmetrical skepticism regarding Static-99R risk estimates, as indicated by mock jurors' reliability and accuracy ratings of the Static-99R risk estimate. Specifically, we predicted that mock jurors who had received negative contextual information condition would exhibit a preference (i.e., higher reliability/accuracy ratings) for a Static-99R risk estimate indicating high risk, and mock jurors who had received positive contextual information would exhibit a preference for a Static-99R risk estimate indicating low risk. However, our results did not support this hypothesis.

Instead, we found a small main effect of the Static-99R risk estimate, with mock jurors rating the low (vs. high) risk Static-99R estimate, as more reliable and accurate. This finding was somewhat unexpected and contrasts with a number of previous studies that found that judges and jurors tend to prefer risk estimates that indicate high risk rather than low risk (Kwartner et al., 2006; A. L. Robinson & Howarth, 2012; Seurich & Krauss, 2013; Varela et al., 2014).

A possible explanation for this result is that mock jurors may have believed the SVP respondent's recidivism risk was low because they did not perceive the crime as a "real

rape" (Borhart & Plumm, 2015; Socia et al., 2019). The stranger rape stereotype tends to be perpetuated by media coverage of these types of crimes (e.g., Norton & Grant, 2008) and is commonly endorsed by members of the public (Krahé et al., 2007). The reality is that most rape victims know the person who sexually offends against them (Levenson et al., 2007; Truman & Rand, 2009).

Adjustments in Mock Jurors' Estimates of Sexual Recidivism Risk

Our final hypothesis was that the direction and degree to which mock jurors adjusted their estimates of the SVP respondent's sexual recidivism risk would be moderated by their ratings of the accuracy and reliability of the Static-99R risk estimate. Specifically, we expected lower accuracy and reliability ratings to yield significantly smaller changes in risk estimates. However, we did not find support for our hypothesis. Instead, we found that the amount of adjustment in mock jurors' risk estimates differed as a joint function of contextual information and the risk level indicated by the Static-99R. While participants in all conditions lowered their risk estimates after having received the Static-99R, the greatest reduction was observed for participants who had received negative contextual information and a Static-99R indicating low risk. Conversely, the smallest reduction was observed for participants who had received positive contextual information and a Static-99R indicating high risk. In other words, the Static-99R risk estimate appears to have had a substantial impact on mitigating the effects of the risk-irrelevant contextual information.

However, we note that even after having received the Static-99R risk estimate, mock jurors overestimated the risk in relation to the Static-99R risk estimate in both the low-risk (Static-99R risk estimate of 3.3%–4.7%) and high-risk groups (Static-99R risk estimate of 10.0%–12.1%). This finding is in line with similar findings reported by Batastini, Vitacco, and colleagues (2019), in that numerical information reduced mock jurors' estimates of violent recidivism risk compared to other formats (i.e., categorical risk, risk management information), but they nevertheless overestimated risk compared to the results of an actuarial risk estimate reported by an expert. Although the Static-99R risk estimate was apparently effective in mitigating the biasing effects of risk-irrelevant contextual information on mock jurors' estimates of the SVP respondent's sexual recidivism risk, they did not reduce their risk estimates to match the risk level that was indicated by the Static-99R estimate.

Limitations and Future Research

We note that there are several limitations to our study. First, although we did find that the Static-99R had a significant effect on mock jurors' risk estimates after reviewing the Static-99R

report, we also note that our case vignette involved an SVP respondent who had a history of one sexual offense and no child victims. These two factors may have contributed to the higher accuracy and reliability ratings for the Static-99R indicating below average risk. This finding contrasts with previous research findings in which participants rated an ARAI indicating high risk as significantly more acceptable than an ARAI indicating low risk (Scurich & Krauss, 2013). It is possible that participants in our study were attentive to the fact that few of the risk factors measured by the Static-99R were present in the vignette, thereby prompting the mock jurors who received the high-risk Static-99R estimate to be skeptical of the results. Given the lower accuracy and reliability ratings of the high-risk Static-99R estimate, we cannot rule out this possibility.

As mentioned previously, it is also possible that some mock jurors did not view the sexual assault as a stereotypical rape nor the SVP respondent as a stereotypical sex offender. Therefore, the Static-99R information may have been rated more favorably in our study when it indicated low risk compared to, for instance, SVP cases that include stranger and/or child victims (Scurich & Krauss, 2013). Research suggests that a person who perpetrates a sexual crime against a child is viewed by the public as inherently more dangerous than one who sexually assaults an adult (Socia et al., 2019). Future research that varies victim type and number of previous offenses may provide further insight into the role of various features of sexual offending.

In addition, in a simulated case, there is literally no risk of subsequent sexual reoffending, which means that there are no real-world consequences to the mock jurors' decisions. This fact perhaps made them more inclined to pay attention to the Static-99R information than would occur in actual SVP cases. Furthermore, asymmetrical skepticism assumes an underlying motivation that information that is inconsistent with prior beliefs will be scrutinized more carefully because the recipient does not *want* to believe it. In actual SVP cases, as opposed to simulated cases, jurors may be (much) more motivated to disregard an actuarial risk estimate if it conflicts with their own perceptions of the SVP respondent's sexual recidivism risk, particularly if the victim is a child.

Finally, the risk estimate provided in the Static-99R report in both the low-risk and high-risk conditions was substantially lower than mock jurors' estimates of sexual recidivism risk prior to reviewing the Static-99R report. This made it difficult to identify a preference for consistent versus inconsistent information, because both Static-99R risk estimates were in a sense inconsistent with (i.e., much lower than) mock jurors' risk estimates prior to reviewing the Static-99R report. It appears that mock jurors' risk estimates, both before and after reviewing the Static-99R report were significantly influenced by their beliefs about the general rate of sexual recidivism among PCSOs (i.e., general risk estimates). In fact, a significant

proportion of the variance in case-specific risk estimates was explained by the general risk estimates both before and after reviewing the Static-99R report.

A possible experimental design to control for mock jurors' general risk estimates would be to provide mock jurors with observed sexual recidivism rates, ranging from 5–15% (e.g., Alper & Durose, 2019; Hanson et al., 2014, 2016) before they read a case vignette containing positive or negative risk-irrelevant information. The observed sexual recidivism rates could provide a scientifically-based anchor in place of mock jurors' preexisting (generally incorrect) sexual recidivism estimates among PCSOs. Mock jurors could then be provided with a case-specific Static-99R risk estimate and asked to make upward or downward adjustments to the Static-99R risk estimate in accordance with the level of risk they believe is accurate for the SVP respondent. Such a design might permit a "purer" test of the effects of an actuarial risk estimate on mock jurors' estimates of an SVP respondent's sexual recidivism risk.

Conclusions

Our findings indicate that mock jurors' general beliefs about sexual recidivism rates among PCSOs account for a large amount of the variance in estimates of a specific SVP respondent's risk of sexual recidivism, suggesting a possible anchoring effect. Presenting actual jurors with a more accurate anchor (i.e., observed sexual recidivism rates) at the beginning of an SVP trial would give them an informed starting point from which they can subsequently incorporate an actuarial risk estimate for the specific SVP respondent. In addition, our findings demonstrate the importance of ensuring that jurors pay attention to a risk estimate derived from a structured risk assessment instrument, particularly because it appears to help mitigate the effects of potentially biasing and risk-irrelevant information about the SVP respondent.

The fact that jurors typically hear sexual recidivism risk information via expert testimony may confound the impact of a risk estimate based on a structured risk assessment instrument, as shown in previous studies (Batastini, Hoeffner, et al., 2019). Therefore, it is worth exploring whether actual jurors would be more sensitive to expert testimony about an actuarial risk estimate if they were provided with written actuarial risk information, including relevant risk factors. This written information may increase their understanding of how an actuarial risk estimate applies to the case at hand as indicated by our findings (see also Scurich et al., 2012). Given the significant liberty deprivation involved in SVP civil commitment, and the poor chances of being released from this type of detention, it is of great importance that jurors are making decisions based on the most accurate information about sexual recidivism risk that is available.

Chapter 6

Who's the Most Dangerous of Them All? Risk-Irrelevant Information and Mock Jurors' Beliefs Bias Sexual Recidivism Risk Estimates

This chapter is based on the following paper:
Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. (2020). Who's the most dangerous of them all? Risk-irrelevant information and mock jurors' beliefs bias sexual recidivism risk estimates. *Manuscript in preparation.*

Abstract

Objectives: In this preregistered experiment, we examined the biasing effects of risk-irrelevant contextual information and mock jurors' attitudes toward sexual offenders on their estimates of sexual recidivism risk and commitment decisions in a Sexually Violent Predator (SVP) civil commitment case. We also examined whether an actuarial risk estimate derived from the Static-99R eliminated any significant effects of the two sources of cognitive bias.

Hypotheses: We predicted higher risk estimates and greater likelihood of recommending civil commitment both before (T1) and after (T2) reviewing Static-99R information among mock jurors who received negative risk-irrelevant contextual information about the SVP respondent, relative to those who received positive information. Similarly, we predicted higher risk estimates and a greater likelihood of a commitment recommendation among mock jurors who held more negative attitudes toward people convicted of a sexual offense (PCSOs) at T1 and at T2. **Methods:** American participants recruited via MTurk ($N = 402$) read a PCSO case vignette, general information about the Static-99R, and a Static-99R report on the case and were asked to provide sexual recidivism risk estimates and commitment recommendations.

Results: The Static-99R information did not eliminate the effects of negative contextual information on mock jurors' risk estimates but did for civil commitment decisions. Furthermore, mock jurors who held more negative preexisting attitudes toward PCSOs, and those who endorsed harsher punishment and more restrictive management of PCSOs, rendered significantly higher risk estimates and were more likely to recommend commitment, irrespective of Static-99R report. **Conclusions:** Our findings indicate that decisions about SVP civil commitment may be more influenced by negative attitudes toward PCSOs than by the objective likelihood that an SVP respondent will engage in sexual recidivism. Implications for juror selection and potential strategies to minimize risk-irrelevant information in SVP civil commitment hearings are discussed.

Introduction

In May of 1989, in what was the latest of several terrible offenses, Earl Shriner committed a violent sexual assault on a 7-year-old boy and left him for dead in the woods of Tacoma, Washington. Shriner had been released two years earlier after serving a 21-year prison sentence for killing a young girl and he had been arrested for several other crimes involving assaults on children since his release (Boerner, 1992). The case gained national coverage in the United States (U.S.), and public outrage ensued over what was perceived to be a failure of both the criminal justice system and the laws permitting civil commitment of people with mental illness (D. M. Smith, 2015). In 1990, the Washington legislature passed what would become the first of modern-day, so-called “Sexually Violent Predator” (SVP) or “Sexually Dangerous Person” (SDP) laws (Boerner, 1992). Today, 20 states, the District of Columbia, and the federal government have enacted similar statutes (Association for the Treatment of Sexual Abusers, 2010).

The SVP Civil Commitment Process

SVP statutes permit individuals who have been charged with, or convicted for, a sexually motivated offense to be confined indefinitely for treatment purposes, usually after the person has served a prison sentence for the crime (DeMatteo et al., 2015; Felthous & Ko, 2018; *Kansas v. Hendricks*, 1997; D. M. Smith, 2015). The legal process for SVP civil commitment begins when a person who is about to be released from prison is identified as potentially at risk for sexual reoffending (see, e.g., Wisconsin: Elwood et al., 2010; New Jersey: Reeves et al., 2019). A prosecutor decides whether to file a petition with the court to ask that the person be civilly committed. The person against whom the commitment petition is filed is known as the *respondent*. A hearing is required to determine whether the respondent meets the statutory criteria to be deemed an SVP and whether they should be civilly committed, based on either an evidentiary standard of “beyond a reasonable doubt” or “clear and convincing evidence” (Felthous & Ko, 2018). All states permit confinement under an SVP civil commitment statute for an indeterminate period of time, but they vary with respect to how frequently a review of the confinement can be requested or must be conducted (Felthous & Ko, 2018).

All state and federal laws must comply with the criteria set out by the United States Supreme Court before imposing SVP civil commitment, which are that the respondent must: (a) have a history of sexual offending; (b) have a mental abnormality or personality disorder; (c) exhibit an impaired ability to control their sexual behavior; and (d) is likely to engage in future sexual offending (*Kansas v. Hendricks*, 1997). Usually a jury decides whether the respondent meets the legal criteria to be deemed an SVP and should be civilly committed

(Krauss & Scurich, 2014; Zolfo, 2018). Yet, determining whether the respondent is likely to engage in future sexual offending is a specialized task, and therefore jurors are often aided in their decision-making by expert testimony about sexual reoffending risk.

Actuarial Estimates of Sexual Recidivism Risk

Although recidivism risk is not the only legal criterion that must be met for SVP civil commitment, we limit our discussion to this issue as it is the focus of the present study. The opinion of one or more mental health experts about the respondent's risk of future sexual offending is a critical component of the civil commitment hearing, and is sometimes the only evidence proffered (Guy & Edens, 2003; Janus & Prentky, 2008; Lieberman & Krauss, 2009; H. A. Miller et al., 2005). Mental health professionals often use structured risk assessment instruments (SRAIs) to evaluate the respondent's risk of sexual recidivism (Barbaree et al., 2001; Jackson & Hess, 2007; Janus & Prentky, 2003; Kelley et al., 2020; Murrie et al., 2009; Neal & Grisso, 2014b) because a substantial body of research indicates their superior reliability and accuracy in estimating sexual recidivism compared to unstructured clinical judgment (Hanson & Morton-Bourgon, 2009; Janus & Prentky, 2003).

The Static-99R (Helmus et al., 2012) is the most commonly used actuarial SRAI for sexual recidivism risk in the context of SVP laws (Archer et al., 2006; Hanson et al., 2016; Hanson & Morton-Bourgon, 2009; Jackson & Hess, 2007; Kelley et al., 2020). Actuarial SRAIs combine factors that significantly predict criminal recidivism based on observed recidivism rates (Janus & Prentky, 2003). The Static-99R contains 10 sexual recidivism risk factors, which are coded and summed to obtain a risk score and a *score-wise risk estimate* (Phenix, Fernandez, et al., 2016). The score-wise risk estimate and categorical risk level (e.g., low, moderate, high) is derived from observed sexual recidivism rates among groups with the same score (Knighton et al., 2014).

Despite professionals' widespread use of SRAIs in evaluating sexual recidivism risk, mock and actual jurors appear relatively insensitive to testimony about SRAI-based recidivism risk estimates (Boccaccini et al., 2013; Krauss et al., 2004, 2012; Lieberman et al., 2007; D. B. Turner et al., 2015). A number of possible explanations have been offered for why jurors tend to give relatively less weight to this testimony, including lack of numeracy (Krauss et al., 2004; Krauss & Lee, 2003; Krauss & Sales, 2001) and lack of scientific knowledge (Batastini, Hoeffner, et al., 2019). Boccaccini and colleagues (2013) suggested it is because recidivism risk itself is not especially important to some SVP jurors, who appear to make civil commitment decisions based on retributive motives (i.e., motivation to punish the SVP respondent) rather than concerns for public safety (see, e.g., Carlsmith et al., 2007).

However, jurors may also have little confidence in mental health experts' ability to

accurately predict sexual recidivism, and therefore rely on their own beliefs about what factors increase the likelihood of sexual recidivism (Scott et al., 2010). Jurors are not blank slates when they enter the courtroom. Perhaps jurors' beliefs and perceptions about people who have committed a sexual offense (PCSOs) contribute to their resistance to testimony based on an SRAI in SVP hearings.

Public Perceptions of People Who Committed a Sexual Offense

Public perceptions of PCSOs are often informed by extreme, but rare cases of sexual violence perpetrated by strangers against children, such as the crimes committed by Earl Shriner (Comartin et al., 2009; Ducat et al., 2009; Harper & Bartels, 2017, 2018; Harper & Hogue, 2015). Media coverage of such crimes tends to evoke powerful negative emotions of fear, disgust, and moral outrage among the public (Comartin et al., 2009; Ducat et al., 2009; Harper & Bartels, 2018; Quinn et al., 2004). Moreover, the emotional valence of these cases appears to contribute to beliefs among the public that PCSOs are a homogeneous group with a high rate of sexual recidivism (Ellman & Ellman, 2015; Katz-Schiavone et al., 2008; Levenson et al., 2007; Quinn et al., 2004; Sample & Bray, 2003).

Public Perceptions of Sexual Recidivism Rates and Knowledge of Risk Factors

Scott et al. (2010) recruited 170 members of the public to participate in an online survey, about 80% of whom said they believed someone who had committed a sexual crime would be likely to commit another offense. Additionally, other surveys of the public indicate that laypeople tend to believe sexual recidivism rates range from 74% (Levenson et al., 2007) to 98% (Katz-Schiavone et al., 2008). In contrast, observed sexual recidivism rates have been reported as ranging from 5–15% with follow-up periods of three to 10 years or more (Alper & Durose, 2019; Hanson et al., 2014, 2016; Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Fernandez, et al., 2016).

However, the general public's tendency to overestimate sexual recidivism risk does not necessarily mean they lack insight into factors that increase the likelihood of sexual recidivism. For example, in Scott et al.'s (2010) survey, a substantial majority of respondents believed that a history of violence (84.1%), a history of sexual violence (91.2%), deviant sexual preferences (80%), and child victims (77.1%) were important sexual recidivism risk factors. In fact, these factors have been shown to increase the likelihood of sexual recidivism (Hanson & Bussière, 1998).

In contrast, a minority of Scott et al.'s (2010) respondents believed that marital status (20%), stable employment (28.8%), and socioeconomic status (28.8%) were important sexual recidivism risk factors, all of which are *risk-irrelevant* and not predictive of sexual recidivism (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). That relatively fewer people endorsed these risk-irrelevant factors as important to sexual recidivism risk suggests such information about an SVP respondent should have little to no effect on jurors' perceptions of sexual recidivism risk.

Despite the fact that PCSOs tend to be among the most stigmatized offenders (Pickett et al., 2013; Tewksbury, 2005), there may be some characteristics of an SVP respondent that reinforce stereotypical views of PCSOs more than others. For example, PCSOs in the public imagination are "loners" who have major mental disorders (Levenson et al., 2007), lack a social life (Borhart & Plumm, 2015), and commit offenses against strangers and children (Wurtele et al., 1992). People whose personal qualities or lifestyle preferences "fit" with this "sex offender" stereotype, or who are perceived as more deviant from mainstream society, may be viewed as at greater risk of sexual recidivism (A. J. Harris & Socia, 2016; Imhoff, 2015; Socia et al., 2019). Hence, a person who has a positive social image as evidenced through prior good deeds, type of employment, or relationships with others may be perceived as less "dangerous" than someone who projects a less positive social image (Alicke & Zell, 2009; Harper & Bartels, 2017, 2018; P. H. Robinson et al., 2012).

Furthermore, there may be case-related information that has no bearing on the legal criteria that must be met to justify SVP civil commitment and has the potential to bias jurors' perceptions of the SVP respondent's sexual recidivism risk. Information that is legally irrelevant may influence the likelihood that an SVP juror will vote for commitment. For example, the impact of the crime on the victim or the victim's opinion of the length of the prison sentence is not legally relevant to establishing the SVP criteria. Nevertheless, jurors may be exposed to such information through the media or directly from the victim, as some jurisdictions permit victim testimony in SVP civil commitment proceedings (see, e.g., *Ellison v. Commonwealth*, 2007; Fluent & Guyer, 2008), or from presentence reports from the criminal case that are admitted as evidence in the SVP civil commitment hearing (see, e.g., *People v. Otto*, 2001).

Despite the fact that SVP jurors may have some awareness of scientifically-validated sexual recidivism risk factors (Scott et al., 2010), it is unlikely that these factors are the only ones that influence their perception of the respondent's sexual recidivism risk. Regardless of whether certain personal characteristics of an SVP respondent are relevant to recidivism risk or not, if the "sex offender" stereotype is activated, or perceptions of sexual recidivism risk are skewed by emotional victim statements, this bias may diminish the weight jurors give to risk-relevant information, such as a risk estimate derived from an SRAI. Furthermore, this bias may carry over to jurors' decisions to vote for civil commitment.

Public Endorsement of Harsh Sentencing and Restrictive Management

Many policies related to management of PCSOs (e.g., residency restrictions, registration requirements, civil commitment) have been a societal response to a small number of high-profile, heinous crimes by PCSOs, such as those committed by Earl Shriner. Indeed, there tends to be substantial public support for harsh sentencing and restrictive measures for managing PCSOs in the community (S. Brown, 1999; A. J. Harris & Socia, 2016; Katz-Schiavone et al., 2008; Levenson et al., 2007; Levenson & Cotter, 2005; Mancini et al., 2010; Pickett et al., 2013; Quinn et al., 2004; Schiavone & Jeglic, 2009; Weiner et al., 2014). Moreover, media coverage of rare but particularly tragic sexual offense cases tends to fuel punitiveness toward PCSOs (Cucolo & Perlin, 2013; Malinen et al., 2014; Quinn et al., 2004) as indicated by general support for harsh sentences and severe restrictions regardless of the efficacy of those policies in preventing sexual reoffending (King, 2019; Rydberg et al., 2018; Socia & Harris, 2016).

We are aware of one previous study that has examined how general punitive attitudes affect processing of actuarial risk estimates in SVP civil commitment decision-making (Krauss et al., 2018). Krauss and colleagues examined the effects of different forms of risk communication (i.e., categorical [e.g., low, high] versus probabilistic [e.g., 21% recidivism rate]) and mock jurors' attitudes toward vengeance and justice on mock jurors' estimates of the probability of sexual recidivism and civil commitment recommendations. Krauss et al. used the Justice-Vengeance scale (Ho et al., 2002), which measures dimensions of attitudes toward vengeance and justice. Whereas the concept of justice tends to be related to a sense of fairness and equity in punishment that is rationally-based, vengeance is associated with an emotion-based response to crime and an increase in harsh sentencing. Krauss et al. found a significant main effect for scores on the vengeance-sentence factor on civil commitment decisions, such that a 1-point increase on the vengeance-sentence factor made it 1.88 times more likely that the juror would vote for commitment. In contrast, neither mock jurors' beliefs about the acceptability of emotion in deciding criminal cases, nor their views about the importance of fairness and adhering to the law in criminal cases had a significant effect on civil commitment decisions. There was also a significant main effect of the risk communication format on commitment decisions in that mock jurors who received a categorical risk estimate were 0.39 times more likely to vote for commitment than those who received a probabilistic risk estimate. These findings suggest that vengeance-based motivations for harsh punishments can influence SVP jurors' decision-making.

There exists ample research suggesting widespread public support for harsh sentencing and substantial restrictions post-prison release for PCSOs, yet studies also indicate that support for different approaches to sentencing and management varies by individual (e.g., Mancini et

al., 2010; Mears et al., 2008) and the nature of the crime and type of victim (e.g., King & Roberts, 2017; Mears et al., 2008; Socia et al., 2019). However, there appear to be no studies that have examined how attitudes and beliefs about PCSOs influence SVP decision-making. Furthermore, we are unaware of any previous research that examines the relative influence of an actuarial risk estimate and mock jurors' attitudes toward PCSOs on sexual recidivism risk estimates or civil commitment decisions. Therefore, in the current study, we examine the extent to which an actuarial risk estimate and variations in support for sentencing and risk management strategies for PCSOs influence juror decision-making in the context of an SVP civil commitment case. In addition, we examine whether attitudes toward PCSOs moderate the effects of risk-irrelevant information on mock jurors' estimates of sexual recidivism and civil commitment decisions.

The Current Study

In this study, we used the Perceptions of Sex Offenders Scale (PSO; Harper & Hogue, 2015) to measure how general perceptions of PCSOs affected case-specific estimates of sexual recidivism risk and decisions about civil commitment. We also examined how attitudes toward sentencing and management of PCSOs, as measured by the Sentencing and Management scale of the PSO, influenced mock jurors' recommendations for civil commitment. A better understanding of beliefs and perceptions that influence juror decision-making in SVP cases, relative to actuarial estimates of sexual recidivism risk, may shed light on the factors that jurors consider important when determining whether an SVP respondent should be civilly committed.

We also evaluated the effects of risk-irrelevant information on mock jurors' perceptions of sexual recidivism risk and recommendations for civil commitment (or an alternative community disposition) both before and after they reviewed Static-99R information, including an absolute score, a risk estimate (expressed as a percentage and as a frequency), and a risk category. In addition, we examined the relationship between participants' perceptions of PCSOs, as measured by the Perceptions of Sexual Offenders (PSO; Harper & Hogue, 2015) scale, case-specific sexual recidivism risk estimates, and recommendations for how the SVP respondent should be managed. Finally, we tested for the presence of an interaction effect between mock jurors' level of support for sentencing and management strategies (measured using the Sentencing and Management subscale of the PSO) and risk-irrelevant information on their post-release disposition recommendations for an SVP respondent. The preregistration of this study can be accessed using the following link: https://osf.io/b5dr2/?view_only=7abb30111fbf4ac5aa8f918a3e684b94.

Hypotheses

Risk Estimates. We predicted that mock jurors who received negative contextual information about the respondent would provide higher estimates of his likelihood of sexual reoffending both (H1a) before (Time 1; T1) and (H1b) after receiving the risk-relevant Static-99R information (Time 2; T2) as compared to mock jurors who received positive contextual information about the respondent. Moreover, we expected that mock jurors who hold more negative (vs. positive) perceptions of PCSOs, would provide higher estimates of the respondent's risk of sexual reoffending at T1 (H2a) and at T2 (H2b). We also conducted preregistered exploratory analyses to examine whether mock jurors' perceptions of PCSOs moderate the effects of the contextual information on mock jurors' estimates (a) at T1 and (b) at T2. In other words, we examined whether contextual information had a stronger or weaker effect on risk estimates dependent upon mock jurors' preexisting perceptions.

Disposition Recommendations. We predicted that mock jurors who have received negative (vs. positive) contextual information about the respondent would be more likely to recommend civil commitment at T1 (H3a) and at T2 (H3b). We also predicted that mock jurors who endorse harsher penalties and more restrictive management of sexual offenders would be more likely to recommend civil commitment at T1 (H4a) and T2 (H4b). Finally, we conducted preregistered exploratory analyses to examine whether mock jurors' endorsement of harsh penalties and restrictive management moderates the effect of contextual information on disposition recommendations at (a) T1 and (b) at T2.

Method

Power analysis

An *a priori* power analysis was conducted using G*Power 3.1 (Faul et al., 2009) for a hierarchical multiple regression analysis with the following parameters for an *F*-test of the increase in explained variance (ΔR^2) associated with an interaction term in hierarchical multiple regression analyses: $R^2 = .02$ (a small effect according to the conventions of Cohen, 1988), $\alpha = .05$, power = .80, and total number of predictors = 4. The minimum number of participants required was 395. We planned to terminate data collection once the minimum number of participants was met (i.e., completed the study, correctly responded to both attention check questions, and indicated their data should be used).

Participant Recruitment

Participants were recruited through Amazon Mechanical Turk (MTurk) and were required to be at least 18 years of age, proficient in English, and to reside in the United States. We used these inclusion criteria to obtain a representative sample of potentially jury-eligible adults. Participants were asked to confirm that they met these three inclusion criteria after giving informed consent. We also required that participants have at least a 95% approval rating on MTurk. MTurk workers who had participated in the pretesting phase of this study or who had participated in one of two similar studies were prevented from participating in the current study.

Participants who indicated that they did not meet one of the three inclusion criteria were automatically screened out from further participation. We also planned to exclude participants whose total time spent on the study was more than three times the median absolute deviation below the sample median (Leys et al., 2013), but none met this criterion. Finally, we asked participants whether they devoted their full attention to the study and whether we should use their data (Meade & Craig, 2012), and excluded them from analyses if they responded with *no*.

A total of 713 people began the study and 21 (2.9%) did not complete it. An additional 136 people (19.1%) were screened out for indicating they did not meet one of the three inclusion criteria. We excluded from analyses the data from 153 participants (21.5%) who failed one or more attention checks and one participant (0.1%) who indicated their data should not be used (Meade & Craig, 2012). Our remaining sample size was 402.

Participant Characteristics

Participants ($N = 402$) ranged in age from 18 to 72 years ($M = 35.2$, $SD = 10.6$). Participants' self-identified gender included 204 men (50.7%), 194 women (48.3%), one transgender male (0.2%), two gender variant/non-conforming (0.5%), and one participant who declined to provide this information. The most common level of education was a bachelor degree ($n = 172$; 42.8%), followed by a master degree ($n = 51$; 12.7%), associate degree ($n = 45$; 11.2%), professional degree ($n = 8$; 2.0%), and doctoral degree ($n = 2$; 0.5%). Eighty-seven participants (21.6%) had some college education, 33 had completed high school (8.2%), and four had not completed high school (1.0%). Nearly every U.S. state and the District of Columbia was represented ($n = 45$), of which the highest number of participants resided in California ($n = 45$, 11.2%). Fifty-two participants (12.9%) indicated they had previously served on a jury in a legal case.

Procedure and Materials

The study was created using Qualtrics software and advertised on MTurk with payment of \$1 and an additional \$2 for participants who correctly answered both attention check questions. This study was approved by the ethical review committee the Faculty of Psychology and Neuroscience of Maastricht University (reference number 207_13_04_2019_A2).

Data Collection

After they had acknowledged informed consent and confirmed that they met the eligibility requirements, participants read a brief introduction to the study and were provided with contact information of the first author.

General Risk Estimate. Participants were asked, to the best of their knowledge, to provide an initial estimate (0-100% using a graphic sliding scale) of the likelihood that someone who has been convicted of a sexual offense would commit another sexual offense. This estimate will be referred to as the general risk estimate hereafter.

Perceptions of Sex Offenders Scale. The Perceptions of Sex Offenders Scale (PSO; Harper & Hogue, 2015) contains 20 items that measure an individual's affective judgments about PCSOs and how PCSOs should be punished and managed post-conviction. Participants provided a rating of their agreement with the statements on a 6-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*). Six of the items are reverse-scored. A total PSO score was obtained by subtracting a constant of 1 from each item and adding up the item scores. Therefore, the total score ranged from 0 to 100, with higher scores indicating more negative perceptions of PCSOs.

The PSO contains three subscales: Sentencing and Management, Risk Perception, and Stereotype Endorsement. There are 10 Sentencing and Management items (e.g., "Sex offenders will almost always commit further offenses."), five Risk Perception items (e.g., "Only a few sex offenders are dangerous."), and five Stereotype endorsement items (e.g., "Most sex offenders do not have close friends."). General attitudes toward PCSOs are correlated with risk perceptions and stereotype endorsement items, whereas the sentencing and management items are correlated with a generally punitive approach to criminal justice (Harper & Hogue, 2015). Subscale scores were obtained by adding up the relevant item scores.

The initial validation study of the PSO scale (Harper & Hogue, 2015) showed excellent internal consistency for the scale ($\alpha = .92$) and high levels of reliability for the three subscales (sentencing and management, $\alpha = .93$; risk perception, $\alpha = .81$; and stereotype endorsement, $\alpha = .85$). Cronbach's α for the full PSO scale in our sample was .85. The Sentencing and Management subscale in our sample showed high reliability, $\alpha = .92$, as did the Stereotype

Endorsement subscale, $\alpha = .87$. Cronbach's α for the Risk Perception subscale in our sample was $.67$.

Demographic information. Participants provided their demographic information including age, self-identified gender, education level, and the state in which they resided. We also asked participants whether they had ever served on a jury in a legal case.

Case vignette. We used a fictional case vignette containing risk-irrelevant contextual information related to social attractiveness to provide background about an SVP respondent, John Smith (see Appendix C). Risk-irrelevant variables, such as type of job, length of employment, hobbies, special talents, third party opinions about the respondent's sociability, and his marital history and relationship with children and ex-wife were varied in the vignette to evoke either a positive or negative perception of him, and participants were randomly assigned to one of the two conditions. These risk-irrelevant factors were selected based on the studies by Alicke and Zell (2009) and Robinson et al. (2012). The effect of these manipulations on perceived likability was confirmed in a pretest and confirmed in the current study with a manipulation check.

Likability Ratings. After reviewing the background information, participants rated the respondent's likability using three items: (1) likability, (2) friendliness, and (3) warmth. These three items were reproduced from the original 11-item Reysen Likability Scale (Reysen, 2005). In contrast to the original scale, which uses a 7-point Likert scale ($1 = \text{very strongly disagree}$, $7 = \text{very strongly agree}$), we did not provide a *neutral* option, and therefore used a 6-point Likert scale ($1 = \text{completely disagree}$, $6 = \text{completely agree}$). Cronbach's α for the three likability items was $.97$ in this sample. We calculated the average rating of the three items to create a composite variable for statistical analysis.

Crime Information and Ratings of Prison Sentence. After rating Mr. Smith's likability, participants read about the offense for which Mr. Smith was incarcerated (i.e., rape of a female acquaintance), his criminal history (one previous conviction for non-sexual assault related to a bar fight when he was 18 years old), public opinion about the length of his prison sentence (5 years) for the sexual assault, and the victim's opinion about the fairness of the prison sentence. Participants then indicated their opinion about the length of the 5-year prison sentence (*too long*, *too short*, or *adequate*).

Next, participants were asked to imagine that they had been selected as a juror in an SVP civil commitment hearing and read an explanation of civil commitment, the legal criteria that must be met, and the indefinite length of civil commitment (see Appendix D). Participants then responded to a multiple-choice attention check question about what they had just read (“Sex offender civil commitment most often occurs:” (a) *while a person is serving their prison sentence*, (b) *before a person is sentenced to prison*, (c) *after a person has served their prison sentence*).

Time 1 (T1) Sexual Recidivism Risk Estimate and Disposition

Recommendation. Participants were then asked to provide an estimate of the likelihood that Mr. Smith would commit another sexual offense (0–100% using a graphic sliding scale) and a disposition recommendation (1 = *community release with no conditions*, 2 = *community release on probation only*, 3 = *community release on probation and mandated sex offender treatment*, 4 = *civil commitment*) based on the information they thought was relevant. These will be referred to as the T1 ratings. Participants were also informed that they would receive additional information and have an opportunity to make any adjustment to the risk estimate and disposition recommendation.

Static-99R information. Participants read a brief description of the Static-99R, reviewed a list of the risk factors measured in the Static-99R (see Table 6.1), and read a simulated Static-99R report specific to Mr. Smith (see Appendix E). We scripted the report using the routine sample template provided in the Static-99R and Static-2002R Evaluator's Workbook (Phenix, Helmus, et al., 2016, p. 30). Participants were asked to identify the correct Static-99R score contained in the report from a 6-option multiple choice list, which served as the second attention check in the study.

Table 6.1 Risk Factors of the Static-99R

1. Age at release from index offense
2. Ever lived with a lover (at least 2 years)
3. Index non-sexual violence
4. Prior non-sexual violence
5. Prior sex offenses
6. Four or more prior sentencing dates (excluding index)
7. Any convictions for non-contact sex offenses
8. Any unrelated victims
9. Any stranger victims
10. Any male victims

Note. Adapted from (Phenix, Helmus, et al., 2016). Used with permission from the authors.

Time 2 (T2) Sexual Recidivism Risk Estimate and Disposition

Recommendation. Once again, participants provided an estimate of the likelihood that Mr. Smith would commit another sexual offense (0–100% using a graphic sliding scale) and a disposition recommendation from among the previously described options. Furthermore, we asked participants to rate how important several factors were to their T2 risk estimate on a 5-point Likert scale (1 = *not at all important*, 5 = *extremely important*), including Mr. Smith's: (a) relationships with others, (b) employment history, (c) criminal history, (d) impact of the crime on the victim, (e) length of prison sentence, and (f) score on the Static-99R. Using

the same scale, participants then rated the importance of each of the previous factors to their disposition recommendation, as well as one additional factor: the participant's own estimate of the likelihood that Mr. Smith would commit another sexual offense.

Accuracy and Reliability Ratings of the Static-99R. To obtain participants' perceptions of the reliability and accuracy of the Static-99R with respect to Mr. Smith, they were asked to rate these factors on a 6-point Likert scale (1 = *completely unreliable*, 6 = *completely reliable* and 1 = *completely inaccurate*, 6 = *completely accurate*). The definitions of reliability and accuracy were provided to participants as follows: "Reliability in this question means how much you think the results can be trusted" and "Accurate in this question means how correct you think the results are." Participants were also asked to provide an open-ended explanation for the reliability and accuracy ratings they gave.

Debriefing and Quality Checks. Participants read the following statement and responded with yes or no: "It is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise, years of effort (the researchers' and the time of other participants) could be wasted. You will receive credit for this study no matter what. In your honest opinion, should we use your data in our analyses in this study?" (Meade & Craig, 2012, pp. 441–442). Participants were then provided with debriefing information.

We used several controls for quality assurance in this study to discourage inattentive or random responding. A timer was used to track how much time participants spent on each page (they were not aware that they were being timed) and the "next" button to move forward was also delayed on most pages (participants were informed of this). Once they moved on to a subsequent page, participants could not go back and change their responses.

Data Analysis

All analyses were conducted using SPSS version 25. We used an alpha level of .05 for all statistical tests. Assessments of effect sizes as small, medium, or large rest on the conventions proposed by Cohen (1988). Participants were randomly assigned to and approximately evenly distributed between the positive ($n = 198$) and negative ($n = 204$) contextual information conditions. Confirmatory and exploratory analyses were conducted in accordance with the preregistration. Any deviations from the analysis plan or unplanned data transformations are noted below.

Results

Preliminary Analyses

We conducted an independent samples Welch's t -test to compare the likability ratings between the positive and negative contextual information conditions. On average, participants in the positive condition rated the respondent as significantly more likable ($M = 5.17, SD = 0.71$) than did participants in the negative condition ($M = 2.36, SD = 0.87$), $t(388.30) = 35.43$, $p < .001$, Hedges' $g = 3.52$, 95% CI [3.21, 3.83]. This finding confirms that before reading about the crime information, there was a large difference between the groups in the perceived likability of Mr. Smith, suggesting that the contextual information manipulation was effective.

Prior to completing the PSO scale and reading any case information, participants provided an estimate of sexual recidivism risk based on their own beliefs about sexual recidivism rates among PCSOs, which we refer to as the general risk estimate. Before hypothesis testing, we confirmed that there was no significant difference between participants in the positive and negative conditions with respect to their general risk estimates for PCSOs ($M = 69.47, SD = 18.58$ and $M = 69.38, SD = 19.20$, respectively), $t(400) = 0.05, p = .961$, Hedges' $g = 0.01$, 95% CI [-0.19, 0.20]. There was also no significant difference between these conditions with respect to their total scores on the PSO scale ($M = 51.11, SD = 13.19$ and $M = 49.61, SD = 13.15$, respectively), $t(400) = 1.14, p = .255$, Hedges' $g = 0.11$, 95% CI [-0.08, 0.31].

Confirmatory Analyses

H1a and H2a: T1 Risk Estimates

A two-step hierarchical multiple linear regression analysis was conducted on participants' ratings of the respondent's recidivism risk at T1 (i.e., before receiving the Static-99R risk estimate), with risk-irrelevant contextual information (positive vs. negative) and participants' total PSO score as the independent variables and participants' general sexual recidivism risk estimate as a covariate. An interaction term between participants' mean-centered total PSO scores and contextual information was entered in Step 2. As planned, we mean-centered participants' PSO total scores for the interaction analysis to prevent multicollinearity. Results of the regression analysis are presented in Table 6.2.

When entered in Step 1, participants' general risk estimate, PSO score, and contextual information explained a large and significant portion of the variance in T1 risk estimates, $F(3, 398) = 87.65, p < .001, R^2 = .393$, 90% CI [0.326, 0.451]. In support of H1a, participants in the negative condition ($M = 54.05, SD = 22.38$) provided risk estimates that were significantly higher than participants in the positive condition ($M = 42.00, SD = 24.49$), $t(398) = 6.96$,

$p < .001$, Hedges' $g = 0.52$, 95% CI [0.32, 0.71].⁹ In support of H2a, the total PSO score was significantly related to T1 risk estimates: higher PSO scores were associated with higher T1 risk estimates, $r = .485$, 95% CI [.407, .556].¹⁰ In addition, participants' general risk estimates accounted for a significant proportion of the variance in T1 risk estimates.

Table 6.2 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' T1 Risk Estimates

Predictor	<i>b</i>	95% CI	<i>SE_b</i>	β	<i>p</i>
Step 1 ($R^2 = .398^{***}$)					
General risk estimate	0.418	[0.308, 0.527]	0.056	0.326	< .001
PSO score	0.651	[0.493, 0.808]	0.080	0.346	< .001
Contextual information	13.100	[9.40, 16.80]	1.883	0.355	< .001
Step 2 ($\Delta R^2 = .000$)					
General risk estimate	0.418	[0.308, 0.528]	0.056	0.326	< .001
PSO score	0.691	[0.479, 0.903]	0.108	0.376	< .001
Contextual information	13.100	[9.40, 16.80]	1.884	0.271	< .001
PSO × Contextual Information	-0.080	[-0.361, 0.202]	0.143	-0.088	.577

Note. $N = 402$. Contextual information condition was dummy-coded (0 = *positive*, 1 = *negative*). Final model $F(4, 397) = 65.70$, $R^2_{\text{Adj.}} = .392$, $p < .001$.

*** $p < .001$.

An interaction term between contextual information condition and participants' mean-centered PSO total scores was entered in Step 2 of the analysis. The interaction term did not account for any significant improvement of the model. This finding indicates that the effect of the risk-irrelevant contextual information on participants' T1 risk estimates was not moderated by their preexisting perceptions of sexual offenders.

H1b and H2b: T2 Risk Estimates

A two-step hierarchical multiple linear regression analysis was conducted on participants' ratings of the respondent's recidivism risk at T2 (i.e., after receiving the Static-99R risk estimate), with contextual information condition and participants' total PSO score as the independent variables and participants' general sexual recidivism risk estimates as a covariate. As anticipated in the preregistration, we found that participants' T2 risk estimates were strongly positively skewed. Therefore, to address issues with non-normally distributed residuals, we performed a log-transformation of the risk estimates before conducting the analysis (after having added a constant of 1 to the estimates to avoid zero values). The interaction term

⁹ The reported effect size does not control for other predictors in the model.

¹⁰ The reported effect size refers to the zero-order correlation between PSO scores and T1 risk estimates.

between participants' mean-centered total PSO score and contextual information was entered in Step 2. Results of the regression analysis are presented in Table 6.3.

Table 6.3 Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' Log-Transformed T2 Risk Estimates

Predictor	b	95% CI	SE _b	β	p
Step 1 ($R^2 = .223^{***}$)					
General risk estimate	0.011	[0.006, 0.015]	0.002	0.215	< .001
PSO score	0.023	[0.016, 0.033]	0.003	0.324	< .001
Contextual information	0.208	[0.047, 0.369]	0.082	0.113	.011
Step 2 ($\Delta R^2 = .001$)					
General risk estimate	0.011	[0.006, 0.015]	0.002	0.216	< .001
PSO score	0.025	[0.016, 0.034]	0.005	0.360	< .001
Contextual information	0.208	[0.047, 0.369]	0.082	0.113	.011
PSO × Contextual Information	-0.005	[-0.017, 0.007]	0.006	-0.051	.421

Note. N = 402. Contextual information condition was dummy-coded (0 = positive, 1 = negative).

Final model $F(4, 397) = 20.71$, $R^2_{\text{Adj.}} = .216$ $p < .001$.

*** $p < .001$.

When entered in Step 1, participants' general risk estimates, PSO scores, and contextual information explained a large and significant portion of the variance in T2 risk estimates, $F(3, 396) = 38.21$, $p < .001$, $R^2 = .217$, 90% CI [.155, .274]. In support of H1b, participants in the negative condition ($M = 30.10$, $SD = 25.59$) provided risk estimates that were significantly higher than participants in the positive condition ($M = 27.01$, $SD = 25.83$), $t(398) = 2.55$, $p < .001$, Hedges' $g = 0.15$, 95% CI [-0.04, 0.35].¹¹ In support of H2b, the total PSO score was significantly related to T2 risk estimates, such that higher PSO scores were associated with higher T2 risk estimates, $r = .485$, 95% CI [.407, .556].¹²

When the contextual information and mean-centered PSO total score interaction term was entered in Step 2, it did not significantly improve the model. This finding indicates that the effect of the risk-irrelevant contextual information on participants' T2 risk estimates was not moderated by their preexisting perceptions of sexual offenders.

H3a and H4a: T1 Civil Commitment Decisions

We conducted a hierarchical binary logistic regression analysis to analyze the extent to which mock jurors' general risk estimates, scores on the Sentencing and Management factor of the PSO scale, and contextual information were associated with the likelihood of recommending

¹¹ The reported effect size does not control for other predictors in the model.

¹² The reported effect size refers to the zero-order correlation between PSO scores and T2 risk estimates.

civil commitment. An interaction term between participants' mean-centered scores on the Sentencing and Management factor and the contextual information factor was added in Step 2. Results of the logistic regression analysis are presented in Table 6.4.

Table 6.4 Results of Hierarchical Logistic Regression Analysis of T1 Civil Commitment Recommendations

Predictor	b	OR	95% CI	p
Step 1^a				
General risk estimate	0.006	1.01	[0.99, 1.02]	.438
Contextual information	0.709	2.03	[1.22, 3.40]	.007
PSO Sentencing and Management	0.093	1.10	[1.06, 1.13]	< .001
Step 2^b				
General risk estimate	0.006	1.01	[0.99, 1.02]	.466
Contextual information	0.789	2.20	[1.25, 3.89]	.007
PSO Sentencing and Management	0.104	1.11	[1.06, 1.16]	< .001
PSO Sentencing and Management × Contextual Information	-0.019	0.98	[0.93, 1.04]	.502

Note. OR = Odds Ratio. Disposition recommendation is dummy coded (0 = *recommendation other than civil commitment* and 1 = *civil commitment*).

^a $\chi^2(3) = 62.44, p < .001$, Nagelkerke $R^2 = .216$. ^b $\chi^2(4) = 62.90, p < .001$, Nagelkerke $R^2 = .218$.

At T1, 95 participants (23.6%) recommended civil commitment. In line with predictions, participants who were exposed to negative contextual information about the SVP respondent were twice as likely as participants in the positive information condition to recommend civil commitment at T1. These results provide support for H3a. In support of H4a, participants' scores on the Sentencing and Management factor of the PSO scale also significantly predicted a recommendation for civil commitment at T1, such that higher scores increased the likelihood of voting for civil commitment.

In Step 2, we examined whether participants' perspectives on punishment and community management of sexual offenders moderated the effects of contextual information on the decision to opt for civil commitment. The interaction term was not significant, indicating no significant moderating effect of participants' punishment and management perspectives on their civil commitment decisions.

H3b and H4b: T2 Civil Commitment Decisions

We conducted a second hierarchical logistic regression analysis that mirrored the model predicting civil commitment decisions at T1, except we used civil commitment decisions at T2 as the outcome variable (Table 6.5).

Table 6.5 Results of Hierarchical Logistic Regression Analysis of Civil Commitment Recommendations at T2

Predictor	b	OR	95% CI	p
Step 1 ^a				
General risk estimate	0.001	1.00	[0.98, 1.02]	.920
Contextual information	0.316	1.37	[0.76, 2.49]	.298
PSO Sentencing and Management	0.102	1.11	[1.07, 1.15]	< .001
Step 2 ^b				
General risk estimate	0.000	1.00	[0.98, 1.02]	.997
Contextual information	0.532	1.70	[0.83, 3.50]	.148
PSO Sentencing and Management	0.122	1.13	[1.07, 1.19]	< .001
PSO Sentencing and Management × Contextual Information	-0.036	0.96	[0.90, 1.03]	.271

Note. OR = Odds Ratio. Disposition recommendation is dummy coded (0 = *recommendation other than civil commitment* and 1 = *civil commitment*).

^a $\chi^2(3) = 48.03, p < .001$, Nagelkerke $R^2 = .198$. ^b $\chi^2(4) = 49.26, p < .001$, Nagelkerke $R^2 = .202$.

At T2, 60 participants (14.9%) recommended civil commitment. Participants who had been exposed to negative contextual information about the SVP respondent were not significantly more likely to recommend civil commitment than participants in the positive contextual information condition. Thus, we found no support for H3b. However, H4b was supported in that participants scoring higher on the Sentencing and Management factor of the PSO scale were significantly more likely to recommend civil commitment at T2.

In Step 2 of the analysis, we added the interaction term between participants' Sentencing and Management scores and contextual information. The interaction term was not significant, indicating that Sentencing and Management factor scores did not moderate the effects of contextual information on the decision to vote for civil commitment at T2.

Discussion

In the current study we examined whether risk-irrelevant information influenced mock jurors' perceptions of an SVP respondent's risk of sexual recidivism, both before and after they reviewed risk-relevant Static-99R information. Furthermore, we examined if there was a significant relationship between mock jurors' support for harsh sentencing and restrictive management in the community for PCSOs and civil commitment recommendations.

Mock Jurors' Estimates of Sexual Recidivism Risk

As predicted, before receiving the Static-99R risk estimate, mock jurors who were exposed to negative risk-irrelevant contextual information about the SVP respondent perceived his risk of reoffending as significantly higher than mock jurors who received positive information. This finding suggests that factors not empirically related to sexual recidivism risk can bias jurors in evaluating an SVP respondent's risk of sexual reoffending and adds support to similar findings reported in a recent study by Kamorowski et al. (2021). In addition, mock jurors' general recidivism risk estimates regarding PCSOs and their perceptions of PCSOs were directly and significantly related to case-specific risk estimates before they viewed the Static-99R information. These findings suggest that an SVP respondent who is perceived as less "likable," either because of risk-irrelevant social information or because of mock jurors' preexisting beliefs about PCSOs, tends to be viewed as more likely to engage in sexual recidivism. We found no evidence that mock jurors' perceptions of PCSOs moderated the relationship between the contextual information and mock jurors' estimates of the SVP respondent's sexual recidivism risk.

As suggested in the Introduction, it may be that the less socially attractive SVP respondent fits more closely to a "stereotypical" PCSO than the respondent who was presented as more socially attractive. A possible alternative explanation, however, is related to attributions of blameworthiness, which may act as a mediator between likability and perceptions of sexual recidivism risk. For example, previous research has shown that less socially attractive criminal defendants are likely to be perceived as more blameworthy for harm caused to a victim (Alicke & Zell, 2009). In fact, Alicke and Zell (2009) found that extenuating or mitigating information was ineffective in eliminating bias toward a less socially attractive perpetrator, unless the extenuating information was presented before his "unlikability" was established. Although blameworthiness is not legally relevant in an SVP civil commitment hearing, it would thus not be surprising to find a relationship between blameworthiness and recidivism risk in SVP cases. In fact, we found that mock jurors' ratings of the importance of the impact of the crime on the victim significantly predicted their risk estimates after receiving the Static-99R report (see Supplemental Materials, Appendix I, Table S2). Whether blameworthiness relates to recidivism risk in SVP civil commitment cases would be an interesting topic for future research.

However, contrary to findings in a similar previous study (Kamorowski, Ask, et al., 2021; Chapter 4 in this dissertation), in the present study we found that after reviewing the Static-99R report mock jurors in the negative contextual information condition again provided significantly higher estimates of risk than did those in the positive condition. Yet, we also note that the difference in risk estimates between the two conditions was relatively small (about

three percentage points). In effect, the Static-99R report nearly eliminated the biasing effects of the risk-irrelevant contextual information. Other researchers have similarly reported that numerical data (for example, recidivism risk expressed as a percentage) is effective in reducing participants' estimates of recidivism risk (e.g., Batastini, Hoeffner, et al., 2019; Scurich et al., 2012), yet the participants' risk estimates in previous studies were almost uniformly higher than the numerical estimate provided.

Laypersons' lack of numeracy (e.g., Krauss et al., 2004; Krauss & Sales, 2001; Slovic & Monahan, 1995) and knowledge about applying statistical information (for a review, see Eldridge, 2019) may partially explain jurors' apparent insensitivity to numerical risk estimates. Interestingly, Scurich and colleagues (2012) found that providing laypersons with a list of the risk factors related to violent recidivism risk particularly helped less numerate participants apply an actuarial risk estimate and decreased the likelihood that they would vote to civilly commit a low-risk individual. In the current study, we provided participants with a list of the sexual recidivism risk factors included in the Static-99R as well as the risk estimate, which may have helped them understand how the Static-99R information applied to the SVP respondent.

However, our mock jurors' preexisting perceptions of PCSOs also had a significant effect on their estimates of the SVP respondent's sexual recidivism risk, both before and after reviewing the Static-99R information. Mock jurors who had more negative perceptions of PCSOs provided significantly higher estimates of the SVP respondent's risk of sexual reoffending as compared to mock jurors with less negative perceptions of PCSOs. This finding lends support to Batastini, Hoeffner, et al.'s (2019) suggestion that individual characteristics of the juror may moderate the relative weight jurors give to actuarial estimates of recidivism risk.

Mock Jurors' Civil Commitment Recommendations

We found support for our hypothesis that negative risk-irrelevant contextual information would significantly predict a vote for civil commitment among mock jurors before receiving the Static-99R report. However, contrary to our hypothesis, negative risk-irrelevant information did not significantly predict a choice for civil commitment at T₂, a finding that indicates that the Static-99R risk estimate mitigated the effects of risk-irrelevant contextual information on disposition recommendations. Furthermore, as predicted, we found that both before and after reviewing the Static-99R information, mock jurors' scores on the Sentencing and Management factor of the PSO scale significantly predicted whether they did or did not vote for civil commitment. Again, mock jurors' perceptions did not moderate the relationship between the contextual information and civil commitment recommendations. These findings indicate

that mock jurors who endorsed more punitive sentencing and more restrictive management in the community were significantly more likely to advocate civil commitment, irrespective of the risk of recidivism indicated by the Static-99R information.

Limitations and Directions for Future Research

A few limitations to our study should be noted. First, although we asked our participants to imagine themselves as a juror in an SVP hearing, there was no deliberation component as there would be in an actual case. Some research suggests that deliberation influences juror decision-making (Lynch & Haney, 2009), yet other research is inconclusive as to whether decisions in simulated cases differ substantially from those in actual cases (Bornstein et al., 2017). We cannot exclude the possibility that, for example, an individual juror's perceptions of PCSOs would have more or less influence on their civil commitment recommendation if they were exposed to the viewpoints of other jurors.

We also note that the rate of recommendations for civil commitment in this study was quite low compared to commitment rates in other studies with mock jurors (Krauss et al., 2012; Scott et al., 2010; Scurich & Krauss, 2014) and in real SVP cases (Boccaccini et al., 2013; Boccaccini, Murrie, & Turner, 2014; Gallagher, 2009; Lu et al., 2015). Research suggests that civil commitment rates may vary as a function of victim type, whereby people appear more inclined to vote for civil commitment when the victim is a child (Scott et al., 2010). Hence, in the current study, the adult victim in the case vignette may partially explain the relatively low civil commitment rates as compared to previous studies.

Furthermore, Socia and colleagues (2019) found evidence to suggest that people who commit sexual offenses against strangers are perceived as more dangerous than those who know the victim. The victim in our vignette was an adult who knew the respondent, which may explain the relatively low rates of recommendations for civil commitment. Future researchers may consider varying victim type and the offender's relationship to the victim, to examine how these factors affect the relative influence of risk-irrelevant and actuarial risk information on civil commitment recommendations.

Finally, this study was conducted online with MTurk mock jurors, which may lessen its ecological validity. In addition, the vignette we presented lacks the emotional "pull" of participating in an actual SVP hearing. However, for the purposes of our study, we needed access to a diverse group of potentially jury-eligible people who live in the United States. In this sense, MTurk provided a more representative sample as compared to mock juror studies conducted with university students (Bartneck et al., 2015; Berinsky et al., 2012; Buhrmester et al., 2018; Casler et al., 2013). Nevertheless, Paolacci and Chandler (2014) note that MTurk

workers tend to be younger, more liberal, and more educated than the general population, suggesting that our results may actually underestimate the effects of perceptions of PCSOs and overestimate the effects of a sexual recidivism risk estimate derived from an SRAI on jurors' decisions.

On a final note, the PSO scale was intended by the developers as an outcome measure, for example, to measure perceptions before and after an educational program (Harper & Hogue, 2015). Nevertheless, the findings from our study indicate that the scale strongly predicts mock jurors' estimates of the SVP respondent's risk of sexual recidivism and their recommendations for civil commitment. Our findings lend support to the use of the PSO scale as a measure of preexisting perceptions of PCSOs and their effects on juror decision-making. The authors of the PSO scale suggested using additional attitudinal measures in conjunction with the PSO scale (Harper & Hogue, 2015), which we did not do in order to avoid overtaxing participants. However, future research may incorporate the PSO scale with other measures to further clarify the role of attitudes (e.g., punitiveness, Carlsmith et al., 2007; belief in capacity for offender change, Wevodau et al., 2016) or affect, such as the emotional valence associated with stigmatizing labels (A. J. Harris & Socia, 2016; Imhoff, 2015; Imhoff & Jahnke, 2018; Scurich et al., 2016) and how these relate to perceptions of PCSOs.

Implications and Conclusions

Our finding that jurors' sexual recidivism risk judgments appear to be easily influenced by risk-irrelevant information, and by their own preexisting views of PCSOs, offers clues as to why reliable and valid measures of sexual recidivism risk have relatively little impact in SVP civil commitment hearings. However, the results of our study also suggest that when given access to the list of risk factors and a written summary of the SRAI-derived sexual recidivism risk estimate, jurors may be inclined to adjust their own preexisting ideas about sexual recidivism risk to align more closely with that indicated by the SRAI. Perhaps when jurors read the information themselves, they make judgments based on the content of the message rather than the likability or perceived credibility of the messenger-expert who delivers it (Cramer et al., 2009; Parrott et al., 2015).

We suggest that for civil commitment to fulfill its intended purpose, screening procedures to probe prospective jurors' perceptions of PCSOs and their ability to make an objective decision in an SVP case could be implemented through *voir dire*. In addition, judges could make concerted efforts to avoid the introduction of risk-irrelevant, potentially prejudicial information during the hearing. Finally, providing jurors with a list of the risk factors relevant to sexual recidivism risk may help them to understand how an actuarial risk

estimate applies to the individual SVP respondent. To ensure that civil commitment is legally justified, identification of the small group of individuals who are “particularly dangerous” (*Kansas v. Hendricks*, 1997, p. 369) should be determined on the basis of risk-relevant factors only, using the most reliable and valid methods of sexual recidivism risk estimation.

Chapter 7

General Discussion

Introduction

Structured risk assessment instruments (SRAIs) are increasingly relied upon in both criminal justice settings and in forensic mental health settings as a presumably objective and evidence-based approach to assessing recidivism risk, although the extent to which this presumption is accurate has been questioned (e.g., Monahan & Skeem, 2016; Shepherd & Sullivan, 2017). The risk of criminal recidivism is a key factor in decisions about how to manage people in the criminal justice system, particularly in the United States (Bechtel et al., 2017; Desmarais et al., 2016; Desmarais & Singh, 2013). In addition, the risk of violent or sexual recidivism is important in forensic mental health settings for decisions related to treatment and management strategies aimed at reducing risk (Bonta & Andrews, 2007). SRAIs have been widely adopted in many countries (Singh et al., 2014). In Chapter 2 of this dissertation, I presented the results of a survey of Dutch forensic mental health practitioners about their use and perceived utility of SRAIs.

In part, SRAIs have been widely adopted to improve the accuracy of recidivism risk predictions and reduce errors in prediction due to cognitive bias. Recently, a number of researchers have specifically focused on the potential for cognitive bias to affect forensic mental health evaluations (MacLean et al., 2019; Neal & Brodsky, 2016; Zapf et al., 2018; Zapf & Dror, 2017; Zappala et al., 2017), although empirical research examining the actual effects of bias on forensic mental health evaluations is limited. The survey presented in Chapter 2 of this dissertation addressed Dutch forensic mental health practitioners' concerns about bias in forensic risk evaluations, their beliefs about sources of bias, and their ratings of the effectiveness of a number of bias correction strategies proposed by other mental health professionals (see Neal & Brodsky, 2016).

One primary aim of the experimental research in this dissertation was to examine whether mental health evaluators' ratings of violent recidivism risk using an SRAI could be influenced by cognitive bias. This question was addressed in the experimental study presented in Chapter 3, wherein the influence of an external source of bias (media exposure) and an internal source of bias (evaluator attitudes toward offenders) on variations in SRAI risk ratings was examined. Although it has long been presumed that SRAIs minimize the effects of evaluator bias (Murrie & Boccaccini, 2015), this presumption has rarely been subjected to empirical testing for either actuarial or structured professional judgment (SPJ) SRAIs. In fact, the extant findings suggest that structured measures that allow for more subjectivity in evaluator ratings (e.g., the PCL-R; Hare, 2003), including SPJ SRAIs, may have a higher susceptibility to evaluator bias than actuarial SRAIs (Grove & Meehl, 1996; Murrie et al., 2013; Shepherd & Sullivan, 2017).

Whether SRAIs mitigate the influence of bias in forensic evaluators' assessments of

recidivism risk is but one concern related to forensic risk assessment. Heilbrun, a preeminent scholar in the field of forensic risk assessment, noted the importance of risk assessment for “better-informed legal decision-making” (1997, p. 347). Hence, the second primary aim of the current dissertation was to examine whether a sexual recidivism risk estimate derived from an SRAI was effective in mitigating bias on the part of jurors. In a series of three experimental studies (Chapters 4–6), we examined mock jurors’ estimates of sexual recidivism risk before and after reviewing an SRAI risk estimate.

In the remainder of this chapter, I discuss the empirical findings from the studies presented in Chapters 2–6 in the context of previous research on bias and risk assessment. Next, I suggest some of the theoretical and practical implications of the research in this dissertation. Subsequently, I discuss some limitations of the current research, propose potential areas for future research, and provide concluding thoughts.

Empirical Findings

Forensic Practitioners’ Use of Structured Risk Assessment Instruments

7

Although SRAIs are widely used in forensic mental health settings, recent studies indicate that the use of specific SRAIs can vary by country (see, for example, McLaughlin & Kan, 2014; Nielsen et al., 2015; Pham et al., 2016; Singh et al., 2019). In the survey presented in Chapter 2, Dutch forensic mental health professionals indicated that they used an SRAI most of the time or all of the time when they conducted a forensic risk evaluation. The most commonly used SRAIs belong to the structured professional judgment (SPJ) category, such as the HCR-20 (Douglas, Hart, et al., 2014) and the HKT-30 or HKT-R (Spreen et al., 2014; Werkgroep Pilotstudy Risicotaxie, 2002). One exception is the Static-99R (Hanson et al., 2016), which is an actuarial SRAI and is the most frequently used sexual recidivism SRAI in the world (Chevalier et al., 2015; Jackson & Hess, 2007; Kelley et al., 2020). Overall, the most commonly-used SRAIs among Dutch practitioners were rated as very useful in conducting forensic risk evaluations. The ratings of the usefulness of various SRAIs in our survey (Chapter 2) add to the limited research regarding mental health professionals’ opinions of the perceived utility of specific SRAIs (Hurducas et al., 2014).

Forensic Practitioners’ Views on Bias in Forensic Risk Evaluations

The potential for bias to threaten the accuracy and objectivity of forensic evaluations has long been acknowledged (e.g., Borum et al., 1993). A number of recent publications have brought

renewed attention to this issue in forensic mental health practice (e.g., MacLean et al., 2019; Neal & Brodsky, 2016; Neal & Grisso, 2014a; Zapf et al., 2018; Zapf & Dror, 2017). In our survey of Dutch forensic mental health practitioners (Chapter 2), they rated cognitive bias as a moderate problem in forensic risk evaluations. Further, a substantial majority of the respondents agreed that cognitive bias is no less of a problem in forensic psychology than it is in other forensic sciences.

Similar to other forensic evaluators (Bonta & Motiuk, 1990; Weinberger & Sreenivasan, 2018) and judges (Hamilton, 2020; Scurich & John, 2012), the respondents to our survey were significantly more concerned about the possibility of a false negative than a false positive in a forensic risk evaluation, indicating a potential tendency toward overestimating an examinee's risk of violent or sexual reoffending. Our findings provide evidence of a possible *error management bias* in forensic risk evaluations, whereby a false positive may be perceived as a less costly error than a false negative (Haselton et al., 2015). However, one problem that stems from the error management bias is the likely increase in error rates (Haselton et al., 2015) regarding predictions of recidivism risk. Further studies could clarify whether an error management bias has a robust effect in forensic risk evaluations and whether there is an effect on related risk management decisions.

Evaluator Bias and Structured Risk Assessment Instruments

As mentioned above, one of the main aims of the research in this dissertation was to examine whether external or internal sources of bias influence forensic evaluators' ratings of risk and related judgments. In the experimental study presented in Chapter 3, we examined whether there were effects from negative pretrial publicity (external source; video compiled from reporting about the case) and forensic evaluators' attitudes toward offenders (internal source) on HCR-20^{v3} (Douglas, Hart, et al., 2014) risk ratings and final risk judgments. We hypothesized that negative pretrial publicity would increase evaluators' perceptions of the examinee's risk of future violence. We also expected that more negative evaluator attitudes toward offenders would predict significantly higher risk ratings and final risk judgments of a forensic examinee.

We found no evidence to suggest that exposure to negative media about the examinee influenced evaluators' ratings of risk or final risk judgments. This null finding could indicate that negative media does not have a biasing effect in forensic risk evaluations. In line with this, some researchers have failed to find a robust effect of pretrial publicity in creating a negative bias toward a criminal defendant (e.g., Bruschke et al., 2016). Yet, the null findings in our study could indicate that the HCR-20^{v3} was effective in eliminating negative portrayals of the examinee in the media. Unfortunately, due to some limitations of the study (e.g., small

sample size and lack of a separate manipulation check), I cannot confidently conclude which explanation is more likely.

In support of our hypothesis about the effects of evaluator attitudes toward offenders on HCR-20 completion, we found that there was a significant negative correlation between evaluator attitudes and scoring of the Clinical and Risk Management subscales, the HCR-20 total score, and the final risk judgment related to imminent violence. In other words, more negative evaluator attitudes toward offenders predicted significantly higher ratings of the examinee's risk of violent recidivism.

Although only one aspect of the final risk judgment was significantly correlated with evaluator attitudes, the judgment was about the risk of imminent violence. The risk of imminent violence is a key consideration in decisions about involuntary civil commitment because "imminence" of violence is often one of the legal criterion that must be met (Melton et al., 2007; Simon, 2006; Stone, 2018). In other words, this aspect of risk judgment can have a substantial influence in the decision about whether a person should be subject to a loss of freedom. The findings from the HCR-20 study presented in Chapter 3 add to a growing body of research suggesting SRAIs alone may not be sufficient to mitigate the influence of cognitive bias in forensic risk evaluations (e.g., Chappell et al., 2013; Chevalier et al., 2015; Guay & Parent, 2018; Murrie et al., 2009, 2013; Murrie & Boccaccini, 2015; Shepherd & Sullivan, 2017). Additional suggestions for strategies to correct for bias will be addressed in a subsequent section. Next, I discuss the findings from three studies in which we examined the extent to which a risk estimate derived from an actuarial SRAI helps mitigate bias among mock jurors in the context of a Sexually Violent Predator (SVP) civil commitment hearing.

Civil Commitment of Sexually Violent Predators

Twenty U.S. states and the federal government have enacted Sexually Violent Predator (SVP) or Sexually Dangerous Person (SDP) civil commitment statutes (Weinberger & Sreenivasan, 2018). These statutes permit indefinite confinement of a person who has been charged with, or convicted of, a sexual offense, usually after they have served a prison sentence (Knighton et al., 2014; Weinberger et al., 2018). A hearing is conducted to determine whether the person whom the government seeks to civilly commit—known as the *respondent*—meets the statutory definition of an SVP and should be subjected to civil commitment. The legal criteria are: (a) that the respondent has a history of sexual offending; (b) that the respondent has a mental abnormality or personality disorder; (c) that the respondent has some difficulty controlling their sexual behavior; and (d) as a result of these factors, the respondent is *likely to engage in future sexual offending* (emphasis added; *Kansas v. Hendricks*, 1997).

Most often, a jury decides whether the legal criteria are met and whether the respondent

should be civilly committed (Krauss & Scurich, 2014; Zolfo, 2018). The critical—and sometimes only—evidence in these cases is the testimony of a mental health professional about the respondent's diagnosis and sexual recidivism risk (Sreenivasan et al., 2003). Typically, a mental health professional uses at least one SRAI to estimate the respondent's sexual recidivism risk (Kelley et al., 2020).

One of the most commonly used sexual recidivism SRAs in SVP cases is the Static-99R (Helmus et al., 2012). However, research suggests that jurors often do not make effective use of expert testimony about recidivism risk based on SRAs (e.g., Krauss et al., 2012; Lieberman et al., 2007). Therefore, we conducted three experimental studies that examined the extent to which an actuarial risk estimate derived from the Static-99R (Hanson et al., 2016) was effective in mitigating the biasing effects of risk-irrelevant information in the context of an SVP case. By presenting the Static-99R risk estimate directly to mock jurors, we were able to eliminate a potential confounding effect of an expert's likability or credibility from the risk estimate itself.

We also examined whether individual juror characteristics influence jurors' estimates of sexual recidivism risk after they have been provided with an actuarial risk estimate. These individual characteristics include need for cognition (Chapter 4) and perceptions of sexual offenders (PSO; Chapter 6). *Need for cognition* (NFC) is the extent to which an individual tends to expend cognitive effort to understand a situation (Cacioppo & Petty, 1982). People who are low in NFC tend to rely upon emotions and generalizations in decisions, whereas those who are high in NFC tend to make a conscious effort to seek evidence and apply logic in decision-making (Epstein, 1993).

In the study presented in Chapter 5, we examined whether asymmetrical skepticism moderated mock jurors' changes in sexual recidivism risk estimates after receiving a Static-99R risk estimate. *Asymmetrical skepticism* is a tendency to interpret preference-consistent information as more accurate than preference-inconsistent information (Ditto et al., 2003; Ditto & Lopez, 1992). In the following discussion, I will refer to these studies as the NFC study, the PSO study, and the asymmetrical skepticism study, respectively. Collectively, I will refer to these three studies as the SVP studies.

Cognitive Bias, Mock Jurors' Estimates of Sexual Recidivism Risk, and the Static-99R

In the SVP studies, mock jurors were initially asked to provide a general sexual recidivism risk estimate that reflected their beliefs about how likely it is that a person who has been convicted of a sexual offense (PCSO) would sexually reoffend. The average estimates ranged between approximately 65% and 70% likelihood that a PCSO would commit another sexual offense.

Similar to other researchers' findings (e.g., King, 2019; Levenson et al., 2007), the mock jurors in the SVP studies presented in this dissertation greatly overestimated sexual recidivism rates compared to observed rates of 5–15% (e.g., Alper & Durose, 2019).

Furthermore, findings from the SVP studies revealed that mock jurors' general sexual recidivism risk estimates accounted for a significant part of the variance in their Time 1 (T1; before receiving the Static-99R report) and Time 2 (T2; after receiving the Static-99R report) case-specific risk estimates in each of the three studies. These findings indicate that jurors' preexisting general beliefs about sexual recidivism rates may carry substantial weight in their perceptions of a specific SVP respondent's sexual recidivism risk.

Misconceptions about sexual reoffending (for example, that sexual recidivism rates are high) are common among the general public and are strongly associated with punitiveness toward PCSOs (e.g., King, 2019; Pickett et al., 2013). Furthermore, punitiveness toward PCSOs fuels support for harsh sentences (Mancini & Mears, 2010; Rydberg et al., 2018; Socia et al., 2019), denial of parole (Scurich et al., 2016), and intensive supervision/monitoring (Levenson et al., 2007), including SVP civil commitment (Bojcenko & Sivasubramaniam, 2019; Carlsmith et al., 2007; Dover et al., 2012; Krauss & Scurich, 2014). Therefore, to ensure that SVP civil commitment is fulfilling its intended purpose, it is important to examine the extent to which jurors' preexisting beliefs about sexual recidivism rates and perceptions of PCSOs affect their civil commitment decisions. Ideally, bias that is created by preexisting misconceptions about sexual recidivism rates and negative perceptions of PCSOs can be eliminated by a risk estimate derived from an SRAI. I next discuss our research findings regarding the extent to which an SRAI risk estimate is effective in eliminating mock juror bias.

Mock Jurors' Estimates of Sexual Recidivism Risk without the Static-99R

In each of the SVP studies, mock jurors were randomly assigned to read background information about the SVP respondent that was intended to create either a positive or negative impression about the respondent's likability (see Appendix C). The information about the SVP respondent and his background (e.g., type of employment, length of employment, leisure activities, quality of the relationship with ex-wife and children) were manipulated to influence mock jurors' perceptions of his likability and to possibly evoke an impression of him as more or less characteristic of a "stereotypical sex offender." The manipulated factors were irrelevant to sexual recidivism risk, and I will hereafter refer to them as risk-irrelevant contextual information.

We expected that mock jurors who received negative risk-irrelevant contextual information would estimate the SVP respondent's sexual recidivism risk as significantly higher than would mock jurors in the positive information condition. As expected, when

controlling for the general sexual recidivism risk estimates discussed previously, the risk-irrelevant contextual information was a significant predictor of mock jurors' estimates of the respondent's risk of sexual recidivism before receiving the Static-99R report in each of the three SVP studies. Mock jurors who received negative risk-irrelevant contextual information consistently estimated the SVP respondent's risk of sexual recidivism as significantly higher than mock jurors who received positive risk-irrelevant contextual information.

Furthermore, in the PSO study (Chapter 6), there was also a significant direct effect of mock jurors' perceptions of sexual offenders, as measured by the Perceptions of Sexual Offenders scale (Harper & Hogue, 2015), on their estimates of the SVP respondent's sexual recidivism risk. Yet, mock jurors' perceptions of PCSOs (positive or negative), did not moderate the effects of the risk-irrelevant contextual information on their estimates of the SVP respondent's risk of sexual recidivism. In other words, the risk-irrelevant contextual information biased mock jurors' risk estimates in the expected direction (i.e., negative contextual information corresponded with higher risk estimates and positive contextual information yielded lower risk estimates) irrespective of how positive or negative were their perceptions of PCSOs.

Although the analyses we conducted in the PSO study were not a direct test of the effects of the sex offender stereotype, the significant main effect of the risk-irrelevant contextual information lends indirect support to previous research findings that indicate that the degree of similarity to a stereotypical sex offender affects community members' perceptions of the PCSO's sexual recidivism risk (e.g., Harper & Bartels, 2017, 2018). However, the finding that perceptions of sex offenders did not moderate the effects of risk-irrelevant information contrasts with previous findings suggesting that people with more negative attitudes toward sex offenders were more negatively influenced by the extent to which a sex offender "fits" the commonly-held sex offender stereotype (e.g., Harper & Bartels, 2017).

In summary, the findings from the SVP studies demonstrate the biasing effects of risk-irrelevant contextual information on mock jurors' perceptions of an SVP respondent's risk of sexual recidivism when the risk-irrelevant contextual information has a negative valence that is suggestive of a "stereotypical" sex offender. However, in each of the SVP studies, we also examined whether a sexual recidivism risk estimate from the Static-99R was effective in mitigating the biasing effects of the risk-irrelevant contextual information. Next, I present the findings related to this question, as well as the results about three possible moderators of the effects of the Static-99R risk estimate: (a) need for cognition, (b) asymmetrical skepticism, and (c) perceptions of sex offenders.

Mitigating Bias with an Actuarial Risk Estimate

In the need for cognition (NFC) study (Chapter 4), neither risk-irrelevant contextual information nor mock jurors' level of NFC significantly predicted their estimates of the SVP respondent's sexual recidivism risk after receiving the Static-99R report. Similarly, in the asymmetrical skepticism study (Chapter 5), there was no significant effect of contextual information on mock jurors' estimates of the SVP respondent's sexual recidivism risk after receiving the Static-99R report. Moreover, in the asymmetrical skepticism study, there was a significant, medium- to large-sized effect of the Static-99R risk estimate on mock jurors' estimates of the SVP respondent's risk of sexual recidivism. Mock jurors who received a Static-99R risk estimate indicating above-average risk (i.e., 10.0–12.1% sexual recidivism rate) estimated the SVP respondent's risk of sexual recidivism as significantly higher than mock jurors who received a Static-99R risk estimate indicating below-average risk (i.e., 3.3–4.7% sexual recidivism rate).

Asymmetrical skepticism was analyzed directly by examining whether the accuracy and reliability ratings of the Static-99R report were dependent upon the congruency between the risk-irrelevant contextual information and the Static-99R report indicating above-average or below-average risk. The specified relationship did not significantly predict accuracy and reliability ratings of the Static-99R report. However, mock jurors who received the Static-99R below-average risk estimate rated it as significantly more reliable and accurate than mock jurors who received the Static-99R above-average risk estimate. The fact that mock jurors believed that the below-average risk estimate was more accurate and reliable than the above-average risk estimate may indicate that when they recalled the case vignette, they realized that few of the risk factors measured by the Static-99R were present. If true, such an outcome would represent a desirable effect on jurors' understanding of sexual recidivism risk factors that future researchers should seek to confirm through additional experimental studies, for example by introducing variations in the age of the victim(s), number of victims, or offender race/ethnicity. Such studies might also include case vignettes that vary the number of risk factors measured by the Static-99R to evaluate the extent to which mock jurors are sensitive to recognizing and weighing risk-relevant factors in combination with potentially biasing risk-irrelevant information about the SVP respondent.

Contrary to the findings in the need for cognition study (Chapter 4) and the asymmetrical skepticism study (Chapter 5), risk-irrelevant contextual information significantly predicted mock jurors' estimates of the SVP respondent's sexual recidivism risk after receiving the Static-99R report in the PSO study (Chapter 6). Mock jurors in the negative risk-irrelevant contextual information condition provided significantly higher estimates of the SVP respondent's risk of sexual recidivism as compared to mock jurors in the positive information condition. However,

the mean difference in the two groups' risk estimates was about three percentage points and the effect size of the correlation between the contextual information and risk estimates was small ($r = .094$). In other words, the statistically significant difference in risk estimates seems unlikely to have much practical significance.

Of greater practical significance is the finding that mock jurors' PSO scores significantly predicted their estimates of the SVP respondent's sexual recidivism risk even after receiving the Static-99R report. Specifically, higher scores on the PSO (more negative attitudes toward PCSOs) predicted higher risk estimates. There was a moderate- to large-sized correlation ($r = .49$) between mock jurors' PSO scores and their estimates of the SVP respondent's sexual recidivism risk. In fact, combined, mock jurors' general sexual recidivism risk estimates and their PSO scores accounted for about 22% of the variance in estimates of the SVP respondent's sexual recidivism risk *after* receiving the Static-99R risk estimate. Most of the variance was attributable to perceptions of PCSOs. This finding seems to add further support to the possibility that stereotypes and perceptions of PCSOs influence the relative weight actual jurors place on a sexual recidivism risk estimate derived from an SRAI.

Yet, mock jurors' estimates of an SVP respondent's risk of sexual recidivism are not the sole point of interest with respect to decision-making in SVP cases. The second, and perhaps more important consideration, is how risk-irrelevant contextual information about the SVP respondent and a Static-99R risk estimate affect mock jurors' decisions regarding SVP civil commitment. The findings from the SVP studies related to civil commitment recommendations are discussed next.

Bias, the Static-99R, and Civil Commitment Recommendations

In each of the three SVP studies, the risk-irrelevant contextual information had a significant effect on mock jurors' votes for civil commitment. In the NFC study (Chapter 4), before reviewing the Static-99R report, nearly one-quarter of mock jurors recommended civil commitment, with a significantly greater proportion of those in the negative information condition recommending civil commitment than those in the positive information condition. After reviewing the Static-99R report, the number of mock jurors who recommended civil commitment was reduced by nearly half, with no significant difference in the number of mock jurors who recommended civil commitment between the positive and negative condition.

Similarly, in the asymmetrical skepticism study (Chapter 5), a significantly larger proportion of mock jurors in the negative contextual information condition recommended civil commitment than did mock jurors in the positive contextual information condition before reviewing the Static-99R report. However, after reviewing the Static-99R report, there was no significant effect of risk-irrelevant contextual information on civil commitment

recommendations. A significantly greater proportion of mock jurors who received a Static-99R risk estimate indicating above-average risk recommended civil commitment than mock jurors who received a Static-99R risk estimate indicating below-average sexual recidivism risk.

Finally, in the PSO study (Chapter 6), risk-irrelevant contextual information significantly predicted a vote for civil commitment before reviewing the Static-99R report, such that negative contextual information doubled the odds of a mock juror voting for civil commitment. In addition, mock jurors' scores on the PSO Sentencing and Management scale significantly predicted a civil commitment recommendation, although the odds were only slightly greater than one. After mock jurors reviewed the Static-99R report, contextual information no longer predicted a vote for civil commitment, but PSO Sentencing and Management scale scores did, again with an odds ratio only slightly greater than one.

Overall, there was a clear pattern in the three SVP studies of the effects of risk-irrelevant contextual information on votes for civil commitment before mock jurors received the Static-99R report. Consistently, mock jurors who received negative risk-irrelevant information were significantly more likely to recommend civil commitment than mock jurors who received positive risk-irrelevant information. This pattern of findings provides evidence indicating that risk-irrelevant contextual information biases mock jurors in a way that is reflected in their civil commitment recommendations. Encouragingly, there was also a clear pattern that after reviewing the Static-99R report, there was no significant difference in the likelihood of voting for civil commitment between mock jurors based on risk-irrelevant contextual information. These findings highlight the significant effect that a risk estimate derived from an SRAI can have on jurors' decisions about SVP civil commitment. Therefore, it appears to be of particular importance that jurors be permitted to weigh a risk estimate derived from an SRAI (for example, by providing them with a copy of the scoring and results obtained from an SRAI), independent of an expert's testimony about the SRAI results.

Factors Important to Mock Jurors who Recommended Civil Commitment

After reviewing the Static-99R report and providing sexual recidivism risk estimates and recommendations for the SVP respondent's release or civil commitment (i.e., disposition recommendations), mock jurors in each of the SVP studies were asked to rate how important they thought various factors were in making their disposition recommendations. These factors included: (a) the SVP respondent's relationships with others, (b) the SVP respondent's employment history, (c) the SVP respondent's criminal history, (d) the impact of the crime on the victim, (e) the mock juror's own estimate of the SVP respondent's sexual recidivism risk, (f) the length of the prison sentence to which the SVP respondent had been sentenced (i.e., five

years), and (g) the Static-99R risk estimate. The disposition options they could choose from included: (a) release with no conditions, (b) supervision only, (c) supervision and treatment, and (d) SVP civil commitment. We used binary logistic regression to analyze whether the importance ratings of any of these factors significantly predicted a vote for civil commitment.

Interestingly, mock jurors' importance ratings of the Static-99R to their disposition recommendation were significantly negatively related to recommending civil commitment only in the PSO study (Chapter 6), and not the NFC (Chapter 4), nor the asymmetrical skepticism (Chapter 5) studies. In other words, in the PSO study, as mock jurors' importance ratings of the Static-99R increased, the likelihood of recommending civil commitment decreased. Given that the Static-99R reduced the likelihood of civil commitment in each of the studies, these conflicting findings may indicate that mock jurors were not consciously aware of which factors influenced their decision. In fact, humans' lack of conscious awareness about the factors that influence their decisions is a phenomenon for which there is extensive evidence from psychological research (for a review, see Nisbett & Wilson, 1977a). Our findings indicate that the Static-99R information did affect mock jurors' likelihood of recommending civil commitment, even if they were not consciously aware of the influence.

Yet, regardless of the actual influence the Static-99R risk estimate had on reducing the total number of people who recommended civil commitment, it is concerning that among mock jurors who *did* recommend civil commitment, their importance ratings of the Static-99R to their decision had *no significant effect on that decision*. This finding seems to lend support to previous research indicating that an actuarial risk estimate was not important to mock jurors and actual jurors who voted for civil commitment relative to other factors (e.g., Boccaccini et al., 2013; D. B. Turner et al., 2015). One potential effect of jurors placing little value on the importance of the Static-99R (or other SRAI) is that the risk of false positives will increase, thereby increasing the likelihood of unjustified deprivations of liberty. Whereas the generally accepted belief is that SRAs provide the most accurate and reliable recidivism risk estimate, whether jurors effectively use this information should be an urgent priority for researchers, judges, and policy makers.

Of additional concern in the SVP civil commitment context is that in each of the three SVP studies, the only importance rating that significantly predicted a mock juror voting for civil commitment after reviewing the Static-99R report was the impact of the crime on the victim. In the NFC study (Chapter 4), a one-point increase in participants' ratings of the importance of the impact of the crime on the victim increased the odds of recommending civil commitment by 1.6 times; in the asymmetrical skepticism study (Chapter 5), 1.7 times; and in the PSO study (Chapter 6), 1.4 times. This finding suggests that the small minority of mock jurors who voted for civil commitment after reviewing the Static-99R report may have been

motivated to do so based on their perception of the impact of the crime on the victim. The impact of the crime on the victim is, in fact, irrelevant to the legal criteria that must be met to justify civilly committing an SVP respondent.

Summary of SVP Study Findings

Overall, the findings from the SVP studies provide evidence of the biasing effect that mock jurors' preexisting beliefs about sexual recidivism rates among PCSOs have on their estimates of an SVP respondent's sexual recidivism risk, irrespective of the availability of a Static-99R report. Second, mock jurors' perceptions of PCSOs affect their estimates of an SVP respondent's sexual recidivism risk. In fact, the only individual characteristic that significantly predicted mock jurors' estimates of the SVP respondent's sexual recidivism risk after reviewing the Static-99 report in the SVP studies was their PSO score (Chapter 6). Neither mock jurors' level of NFC (Chapter 4), nor an asymmetrical skepticism effect, significantly predicted or moderated mock jurors' estimates of the SVP respondent's risk of sexual recidivism after reviewing the Static-99R report.

In addition, the findings indicate that negative risk-irrelevant information about an SVP respondent may suggest a "fit" with commonly-held beliefs about the "stereotypical sex offender." Specifically, mock jurors' estimates of his sexual recidivism risk were significantly higher in the negative risk-irrelevant information condition before receiving the Static-99R report in all three SVP studies. Moreover, mock jurors in the negative risk-irrelevant information were significantly more likely to recommend civil commitment as compared to mock jurors in the positive information group.

The Static-99R report appeared mostly effective in mitigating the biasing effect of risk-irrelevant contextual information on mock jurors' estimates of the SVP respondent's risk of sexual recidivism (except for the small difference between the positive and negative group risk estimates observed in the PSO study in Chapter 6). Furthermore, there were no significant differences between the groups with respect to civil commitment recommendations in any of the SVP studies after mock jurors received the Static-99R report.

SRAIs are one strategy that may be effective in mitigating the influence of cognitive bias in forensic risk evaluations. Yet, they are not the only strategy, nor should they solely be relied upon as sufficient to mitigate bias, as indicated by our findings in the HCR-20 study presented in Chapter 3. In the next section, I briefly discuss forensic mental health practitioners' views about various bias correction strategies.

Strategies to Counteract the Effects of Cognitive Bias

In the survey of Dutch forensic mental health practitioners (Chapter 2), most practitioners agreed that an evaluator's prior beliefs and expectations can affect how they analyze a case and their ultimate opinion in the case. Yet, a majority also agreed that an evaluator's conscious effort to set aside their prior beliefs and expectations can mitigate the potential for bias. However, to correct a bias, an evaluator must first be aware of the fact that their judgment has been "contaminated," they must be motivated to correct the error, and they must be aware of the direction and magnitude of the bias (Wilson & Brekke, 1994). Yet, assuming all of those requirements are met, it remains difficult for people to sufficiently correct for the biasing effects of mental processes that occur automatically (Wilson & Brekke, 1994).

Respondents to the survey of Dutch forensic mental health practitioners rated using structured evaluation methods as the most effective "debiasing" strategy. Yet, findings from the HCR-20 study (Chapter 3), among others (e.g., Murrie et al., 2009; Murrie & Boccaccini, 2015), indicate that SRAIs may not completely eliminate the effects of external or internal sources of evaluator bias. Therefore, it is important to consider additional strategies that may be useful in preventing or mitigating the influence of bias.

The next four highest-rated bias correction strategies among our survey respondents were, in descending order: (a) taking personal responsibility for continued learning after formal education and training, (b) investigating all relevant data before forming an opinion, (c) basing conclusions and opinions on sound data, and (d) consulting with colleagues about potential issues of bias. While each of these practices are aligned with professional recommendations for best practices among forensic mental health evaluators (American Psychological Association, 2013; Glancy et al., 2015; Grisso, 2010), I am unaware of any empirical testing regarding the efficacy of these strategies to mitigate, or correct for, bias in forensic evaluations.

Theoretical Implications

Forensic Mental Health Practice

In the survey in Chapter 2 of this dissertation, most forensic mental health practitioners agreed that an evaluator's prior beliefs and expectations can affect how they analyze a case and their ultimate opinion in the case. When a forensic examiner's prior beliefs and expectations influence how they analyze a case and their ultimate opinion, this has been referred to as a *forensic confirmation bias* (Kassin et al., 2013). In fact, a number of studies have found an adversarial allegiance effect (e.g., Chevalier et al., 2015; Murrie et al., 2013; Murrie &

Boccaccini, 2015), thereby providing evidence of the effect an “expected” outcome can have on forensic risk evaluations. Relatedly, Dror and colleagues (2017) proposed a *bias cascade effect*, in which a bias that arises early in a forensic investigation as a result of contextual information influences the subsequent collection, evaluation, and interpretation of evidence. In other words, the mechanism by which a bias such as an adversarial allegiance effect arises is possibly the result of a bias cascade. Theoretical testing of a bias cascade effect would help clarify how bias affects the ultimate conclusions in a forensic risk evaluation.

Elbogen (2016) notes: “We know relatively little about how decision-making biases and heuristics affect use of violence risk assessment tools, and we have conducted virtually no empirical research aimed at describing how the aforementioned complexities of clinical contexts affect use of violence risk assessment science” (2016, p. 70). In fact, Brown and Rakow (2016) found a high level of variation among forensic clinicians with respect to how they integrated various data points into a violence risk assessment. The findings obtained in our HCR-20 study and presented in Chapter 3 indicated that forensic mental health practitioners differ in their interpretation and use of cues in violence risk assessment. Variations in how practitioners assess risk indicate the need for research to develop a unified theory about how forensic mental health clinicians identify, evaluate, and integrate information to formulate their conception of the case and their judgments of violent or sexual recidivism risk (see, e.g., Elbogen, 2002, 2016). Such research may also help clarify the effects of bias on various stages of the forensic risk evaluation process.

Finally, as the results of our survey in Chapter 2 and other surveys suggest (e.g., Neal & Brodsky, 2016; Zapf et al., 2018; Zappala et al., 2017), forensic mental health practitioners are concerned about bias in forensic evaluations. In addition, numerous scholars have written about the potential biases that can affect forensic evaluations (e.g., Borum et al., 1993; Dror & Murrie, 2018; Neal et al., 2019; Zapf & Dror, 2017). Yet, research about the ways in which cognitive bias can affect forensic risk evaluations and completion of SRAIs remains quite limited, with few exceptions (see, Murrie & Boccaccini, 2015).

Strategies to Mitigate the Effects of Bias in Forensic Risk Evaluations

While SRAIs are one potential method to reduce the influence of bias in forensic risk evaluations, the findings from our HCR-20 study (Chapter 3) indicate forensic evaluators’ risk ratings of a specific offender are influenced by the evaluator’s preexisting attitude toward offenders. These findings, combined with the prior limited research that has examined whether examiner bias may account for variations in SRAI scoring/completion (Murrie et al., 2009), suggest that using an SRAI in a forensic risk evaluation does not guarantee an objective assessment of recidivism risk. As indicated in a study by Neal and Brodsky (2016) and in our

survey of Dutch forensic mental health professionals (Chapter 2), opinions about potentially effective bias correction strategies in forensic evaluations vary among forensic mental health professionals.

However, there are two limitations to the strategies proposed by Neal and Brodsky (2016). The first is that with few exceptions, most of them have not been subjected to empirical testing. The second limitation is that very few of the strategies have been tested specifically in the context of forensic risk evaluations. This type of research is important because debiasing strategies are not necessarily effective for all types of cognitive bias, nor are they necessarily effective for all types of tasks (Fischhoff, 1982; Soll et al., 2015).

For example, Mumma and Wilson (1995) asked undergraduate psychology students to make clinical ratings of various behavioral cues as being diagnostic of bipolar disorder to test for a primacy/anchoring effect. The primacy or anchoring effect is a form of bias that occurs when information presented at an early stage of evaluation has a disproportionate influence on the ultimate judgment, as compared to the same information presented at a later stage (Hogarth & Einhorn, 1992; Tversky & Kahneman, 1974). After confirming a robust anchoring effect, Mumma and Wilson found that the strategies of “considering the opposite” and taking notes on clinical cues just prior to making judgments were highly effective in mitigating primacy effects and single-cue anchoring effects. In another example specific to forensic risk evaluations, de Vogel and de Ruiter (2004) found that consensus ratings of the HCR-20 led to significantly improved predictive validity of violence risk judgments as compared to individual evaluator ratings. This finding suggests that consulting with colleagues may mitigate the effects of single evaluator bias.

Additional strategies that have been proposed as potentially effective in mitigating cognitive bias include: receiving explicit didactic training about objectivity (e.g., Bridge & Marić, 2019; Croskerry et al., 2013), taking time to think about the evaluation before writing the report (e.g., Croskerry et al., 2013; Lilienfeld et al., 2009), and reading professional literature about the importance of objectivity (Croskerry et al., 2013). Several other strategies that have been proposed as potentially effective in mitigating bias, specifically in forensic evaluations, include: investigating all relevant data before forming an opinion and basing conclusions on sound data (Grisso, 2010; Zapf & Dror, 2017) and taking careful notes during an evaluation (Borum et al., 1993). Further, a number of researchers have suggested examining patterns of personal decision-making, such as comparing one’s decisions over time to known base rates and to agreement with referral party preferences (Brodsky, 2013; Gowensmith & McCallum, 2019; Murrie & Warren, 2005).

In summary, ongoing discussions in the literature about the potential effects of bias in forensic evaluations in general (e.g., Dror & Murrie, 2018; Neal et al., 2019; Zapf & Dror,

2017) and evaluators' concerns about the potential for bias in forensic risk evaluations in particular (e.g., Kamorowski, de Ruiter, et al., 2021; Zapf et al., 2018), highlight the need to empirically test potentially effective debiasing strategies. In addition, debiasing strategies have shown promise in other areas of forensic science (e.g., Quigley-McBride, 2020), such as linear sequential unmasking (Dror et al., 2015), which may also be useful in the context of forensic risk evaluations. An objective forensic risk evaluation can help improve accuracy of recidivism risk predictions. Moreover, it improves the likelihood of accurately identifying the highest risk offenders who will benefit the most from intensive treatment interventions. Therefore, it is important to examine additional strategies that may be effective in mitigating the influence of bias in forensic risk evaluations.

Legal Decision-Making

Numerous studies have examined how expert testimony about estimates of violent or sexual recidivism risk affect legal decision-makers' estimates of recidivism risk and legal decisions related to recidivism risk (e.g., Krauss et al., 2018; Kwartner et al., 2006; Lieberman et al., 2007; Lieberman & Krauss, 2009). Explanations for what appears to be jurors' relative lack of sensitivity to expert testimony about risk estimates derived from SRAIs have tended to focus on mock jurors' information processing tendencies (e.g., Lieberman et al., 2007; McCabe et al., 2010) or apparent preferences for how risk is communicated (e.g., Krauss et al., 2018; Scurich & John, 2012; Slovic et al., 2000; Varela et al., 2014). Like the preceding discussion about the gap in research that examines how external and internal factors may bias forensic clinicians' evaluations of violent or sexual recidivism risk, research that examines external and internal factors that may bias jurors' evaluations of risk is largely absent.

The three SVP studies we conducted (Chapters 4–6) represent an attempt at a more unified approach to examining how risk-irrelevant contextual information and individual juror characteristics can influence their estimates of sexual recidivism risk and decision-making about SVP civil commitment. The impetus for this research was the idea that there is complexity in juror decision-making in SVP cases and they apparently do not make decisions solely based on sexual recidivism risk estimates derived from an SRAI (Boccaccini et al., 2013; D. B. Turner et al., 2015).

The findings from the three SVP studies indicate that risk-irrelevant information can bias jurors with respect to their estimates of the SVP respondent's risk of sexual recidivism and civil commitment decisions, but that for most, an actuarial risk estimate appears to reduce their estimates of the SVP respondent's risk of sexual recidivism. Nevertheless, the variation in estimates of the SVP respondent's sexual recidivism risk and jurors' votes for civil commitment appear to be influenced by three primary factors: (a) their general estimates of

the sexual recidivism rates among PCSOs; (b) their perceptions of PCSOs; and (c) the impact of the crime on the victim. Replication studies and future research about the influence of individual characteristics and contextual information about the case can help clarify how jurors in SVP cases make their decisions. Relatedly, SRAIs are increasingly being used in a variety of decisions in the criminal justice system, including pretrial detention (Bechtel et al., 2017), sentencing (Monahan, 2017), and parole (Desmarais et al., 2016) decisions. Therefore, analogous research on how judges and jurors process risk estimates derived from SRAIs in the context of other criminal case factors is an important gap to be addressed.

Practical Implications

Forensic Mental Health Practice

The first practical implication of the research presented in this dissertation for forensic mental health practice is related to evaluators' greater concerns about a false negative than a false positive in a forensic risk evaluation, as illustrated by our findings in the survey presented in Chapter 2. There are at least three concerns related to overestimating an examinee's risk of violent or sexual reoffending. First, an individual who is rated as "high risk" is likely to experience unnecessary loss of liberty (Boccaccini et al., 2013) or delayed release from confinement (Guy, Kusaj, et al., 2015).

Second, there are limited resources to provide community-based treatment and facilities in which to confine people, when appropriate from a community safety perspective. For example, in a recent study, 75% of judges surveyed in Virginia ($N = 85$) said that the availability of outpatient drug or mental health programs in their jurisdiction was either *less than adequate* or *virtually non-existent* (Monahan et al., 2018). Using limited resources and space when it is unwarranted is costly, wasteful, and can result in other individuals who exhibit greater risks and needs not obtaining appropriate services (Bonta & Andrews, 2007). In fact, some studies suggest it is common for practitioners to seek more restrictive supervision options or target treatment needs not indicated by an SRAI (e.g., J. Miller & Maloney, 2013). Finally, studies indicate that when low risk offenders receive intensive services or supervision, it can actually increase their criminal behavior (e.g., Bonta et al., 2000). Therefore, identifying factors that may lead to bias and errors in forensic risk evaluations and decisions related to recidivism risk should be a research priority.

The second practical implication of the studies in this dissertation are that bias has the potential to influence forensic evaluations, including risk evaluations conducted with an SRAI (Chapter 3). As the results of the HCR-20 study indicated (Chapter 3), forensic risk evaluations can be affected by internal sources of bias (Gowensmith & McCallum, 2019). Although we did

not find support for the effects of an external source of bias (i.e., negative media depictions of the suspect), this finding cannot be considered definitive due to study limitations (particularly limitations on statistical power). One important step in avoiding the influence of bias is to be aware of how it may interfere with objective observations and inferences (e.g., Dror & Murrie, 2018; Neal & Grisso, 2014a; Zapf & Dror, 2017). However, the operation of bias typically occurs outside of conscious awareness (Nisbett & Wilson, 1977a), which means that introspection is not an effective debiasing strategy (Pronin & Kugler, 2007). Yet, there are some steps that may be helpful in avoiding bias.

For example, although collateral sources of information can be important to a forensic evaluation, evaluators might take steps to avoid being exposed to information that has likely originated from a biased source, such as a victim impact statement (Jackson & Hess, 2007; Lynett & Rogers, 2000) or is irrelevant to the forensic evaluation task (Neal & Saks, 2016). Another suggestion is to consider alternative hypotheses or engaging in hypothesis testing in the forensic evaluation (Heilbrun et al., 2009; Melton et al., 2007; Zapf & Dror, 2017). Finally, a consensus approach to violence risk assessment has been shown to lead to improved accuracy related to violence predictions (de Vogel & de Ruiter, 2004).

Some bias correction or bias avoidance strategies may be easier to incorporate in forensic mental health practice (e.g., considering alternative hypotheses, consulting with colleagues) than others (e.g., avoiding exposure to task-irrelevant contextual information). Understanding how bias operates and knowing potential sources of bias in forensic evaluations (Zapf & Dror, 2017) is an affirmative step that can be taken by forensic mental health practitioners to improve their awareness of bias. Therefore, future research should be aimed at identifying strategies that can be effective for practitioners to minimize the potential for bias.

Furthermore, the results of the SVP studies in this dissertation suggest that a risk estimate derived from an SRAI can mitigate some bias on the part of mock jurors. It is essential that the risk estimate they receive is valid. Yet, if a risk estimate derived from an SRAI reflects bias on the part of the forensic evaluator, then all that will be accomplished is substituting evaluator bias for juror bias. I next discuss some of the implications of our research for legal decision-making, and possible strategies to help legal decision-makers make better-informed decisions.

Legal Decision-Making

There are several implications of the research findings presented in this dissertation for legal decision-makers. The first is related to the judicial role as gatekeeper of the admissibility of expert testimony (*Daubert v. Merrell Dow Pharmaceuticals*, 1993), specifically related to SRAs. As suggested by the findings in the HCR-20 study (Chapter 3) and others (e.g., Murray

et al., 2014; Murrie et al., 2009, 2013; Murrie & Boccaccini, 2015), use of an SRAI in estimating recidivism risk is not a guarantee of objectivity. The HCR-20 has been entered as evidence in parole hearings, SVP civil commitment hearings, capital sentencing hearings, and hearings about transfer of a juvenile to the jurisdiction of an adult court (Cox et al., 2018; Vitacco et al., 2012). For example, Cox et al. (2018) surveyed federal criminal cases and state court appeals cases between 2010 through 2016 in which the HCR-20 was entered as evidence ($N = 134$). In 88% of cases, the HCR-20 was introduced as evidence by the prosecution and was challenged—unsuccessfully—in only 11 cases. In fact, at least 19 state courts have concluded that evidentiary standards such as those outlined in *Daubert* do not apply to expert testimony based on SRAs (Krauss & Scurich, 2013).

It should be of concern to legal professionals that studies indicate that experts who use the same SRAI to evaluate the same individual can arrive at different results (e.g., Boccaccini et al., 2012; Chevalier et al., 2015; Murrie & Boccaccini, 2015), to which our findings in the HCR-20 study (Chapter 3) lend support. Furthermore, practicing professionals do not always use SRAs in the manner prescribed by the instrument developers (Guay & Parent, 2018; Krauss & Scurich, 2013; J. Miller & Maloney, 2013). As discussed in the previous section related to implications in forensic mental health settings and our findings in the survey of Dutch forensic mental health practitioners (Chapter 2), practitioners also exhibit a tendency to overestimate rather than underestimate risk (Bonta & Motiuk, 1990). Moreover, in some settings, upward increases in risk estimates tend to be applied more often to racial minorities (Chappell et al., 2013; Schmidt et al., 2016) or to certain types of offenders (e.g., PCSOs; Schmidt et al., 2016; Storey et al., 2012; Wormith et al., 2012).

Judges might consider that the question of whether expert testimony based on an SRAI meets scientific standards requires some inquiry about the underlying methods employed by the evaluator to complete the SRAI. It should constitute no additional burden on the evaluator to answer questions about the process they undertook to collect information, their approach to integrating the information, and the reasoning upon which they ultimately based their conclusions *specific to completion of the SRAI*. I do not suggest judges should be able to detect an underlying bias in the SRAI results, but rather, that they should at least inquire about the process. In furtherance of this suggestion, I note that a number of recommendations about how to help judges and lawyers understand and evaluate forensic evidence have been offered and may be useful (Evans et al., 2019; Shooter & Cooper, 2019).

Some additional implications of the studies in this dissertation apply to juror decision-making and risk estimates derived from an SRAI. First, SVP statutes depend upon the accurate and reliable identification of those few individuals who are at the highest risk of sexual recidivism (*Kansas v. Crane*, 2002; *Kansas v. Hendricks*, 1997; Levenson, 2004). The most

valid approach currently known to identifying who is at high risk for sexual recidivism are SRAIs. Therefore, it is promising that the SVP studies conducted in this dissertation (Chapters 4–6) indicate that a risk estimate from an SRAI has a mitigating effect on the biasing effects of risk-irrelevant factors about the SVP respondent and other case-related factors.

The first implication of our findings is that it may be helpful to provide jurors with a copy of the SRAI report in cases where the risk of recidivism is at issue (e.g., SVP civil commitment hearings). Previous studies have found that mock jurors and actual jurors give relatively little weight to expert testimony about actuarial risk estimates (e.g., Boccaccini et al., 2013; Krauss et al., 2012; D. B. Turner et al., 2015). One proposed reason has been juror lack of numeracy (e.g., Krauss et al., 2004; Krauss & Sales, 2001). However, Scurich and colleagues (2012) found that providing less numerate jurors with a list of relevant risk factors helped them apply an actuarial risk estimate to the case at hand. Similarly, the findings from the three SVP studies in this dissertation suggest that a written version of an SRAI report may be a useful supplement to expert testimony to inform jurors' perceptions of recidivism risk and related decisions.

Nevertheless, I would also note that the risk estimates provided by the mock jurors after reviewing the Static-99R report were still much higher in each of the three SVP studies (NFC study: 25.6%; asymmetrical skepticism study: 25.3%; PSO study: 29.0%) than that indicated by the Static-99R estimate (7.0–8.8% in the NFC and PSO studies; 3.3–4.7% or 10.0–12.1% in the asymmetrical skepticism study).

Yet, as a point of comparison, Scurich and Krauss (2014) reported that after their mock jurors reviewed expert testimony based on an average-risk Static-99R estimate, their average estimate of the SVP respondent's sexual recidivism risk was 74%. However, the victims in Scurich and Krauss's vignette were two children. In contrast, the victim in our SVP studies was an adult female. It seems plausible that the mock jurors in our SVP studies were more influenced by the Static-99R risk estimate because an adult victim of sexual assault does not have the same negative emotional impact as a child victim of sexual abuse. Furthermore, people who sexually offend against children tend to be viewed by the public as the most dangerous and most deserving of punishment (Comartin et al., 2009; Ferguson & Ireland, 2006; Socia et al., 2019). Hence, the effectiveness of the Static-99R report in our SVP studies in reducing mock jurors' estimates of the SVP respondent's sexual recidivism risk may have been related to the nature of the vignette we chose (i.e., adult female victim, acquaintance rape). Therefore, it cannot be presumed that the influence the Static-99R report had on mock jurors' estimates of the SVP respondent's sexual recidivism risk estimates in our studies would reproduce in a case that involved a child victim.

The second practical implication from the SVP studies is that it may be useful to provide jurors in SVP civil commitment cases with the observed base rate of sexual recidivism before

any evidence is presented. The general public tends to overestimate the risk of sexual recidivism among PCSOs (e.g., Chapters 4–6 in this dissertation; Katz-Schiavone et al., 2008; Levenson et al., 2007) as compared to observed recidivism rates, a finding that was replicated in each of the SVP studies we conducted. Moreover, in each of the three SVP studies, a substantial amount of the variation in mock jurors' estimates of the SVP respondent's sexual recidivism risk was accounted for by their initial estimate of sexual reoffending among PCSOs in general (hereafter, general risk estimate).

The effect of mock jurors' general risk estimates was reduced, but not eliminated, after they reviewed the Static-99R report in all three studies. It appears that mock jurors' general risk estimates may have had an *anchoring effect*, a form of cognitive bias in which an initial value has a disproportionate effect on subsequent probability estimates (Tversky & Kahneman, 1974). Therefore, as a corollary measure to providing jurors with a written version of the SRAI report, it may be useful to provide jurors with an initial value, or base rate of observed recidivism, on which to anchor their case-specific risk estimates.

However, this approach should be empirically tested because some studies have found that mock jurors tend to ignore base rate information (e.g., Batastini, Hoeffner, et al., 2019; S. L. Miller & Brodsky, 2011; Pennycook et al., 2014; Walters et al., 2014). Moreover, the public tends to be misinformed about a number of aspects related to sexual abuse and perpetrators of this abuse (e.g., Fedoroff & Moran, 1997; Katz-Schiavone et al., 2008; Levenson et al., 2007). Providing basic education to jurors about sexual offending can help them make better-informed decisions and may mitigate the potential biasing effects of any misconceptions they may have (King, 2019; Malinen et al., 2014).

Although there does not appear to be a direct relationship between jurors' estimates of the SVP respondent's sexual recidivism risk and a vote for civil commitment (e.g., Scurich & Krauss, 2014), higher risk estimates tend to increase the likelihood of civil commitment. In fact, misconceptions about PCSOs, including about sexual recidivism rates, have been found to be significantly associated with punitive attitudes toward this group (King, 2019). Therefore, it is important that jurors are basing their estimates on the most scientifically-valid risk estimate available (i.e., the estimate derived from an SRAI) rather than their preexisting beliefs about sexual recidivism rates among PCSOs.

The final practical implication of the SVP studies is related to SVP civil commitment decisions and the basis upon which jurors vote for civil commitment. Although most state statutes contain some language about the likelihood of sexual reoffending (e.g., "likely," "substantially probable," "more likely than not"), only the state of Washington has specifically assigned a percentage—more than 50%—to the meaning of likely (for an overview of state statutes and applicable case law, see Knighton et al., 2014). Some state courts have explicitly

declined to declare a percentage value that should be associated with the term “likely” (e.g., North Dakota and Massachusetts; Knighton et al., 2014). Therefore, it is typically left to a jury or a judge to decide what level of sexual reoffending risk is sufficient to meet an amorphous legal likelihood standard.

In the three SVP studies in this dissertation, despite the Static-99R report in the NFC study (Chapter 4) and the PSO study (Chapter 6) indicating a sexual recidivism rate of 7–8.8% and the Static-99R report in the asymmetrical skepticism study (Chapter 5) indicating a sexual recidivism rate of 3.3–4.7% and 10–12.1%, there were jurors who nevertheless voted for civil commitment. Specifically, in the NFC study, 52 (12.2%) mock jurors recommended civil commitment, 67 (17.8%) in the asymmetrical skepticism study, and 60 (14.9%) in the PSO study. Similarly, Knighton and colleagues (2014) found that a large majority of SVP jurors in Texas indicated that a 15% chance or less of sexual recidivism meant the SVP respondent was “likely” to reoffend. All jurisdictions that have SVP civil commitment laws (except for the State of Washington) lack a defined level of sexual recidivism risk that is deemed a sufficient probability to justify civil commitment. This decision is largely left to jurors in these cases.

Furthermore, some researchers have suggested that a sexual recidivism risk estimate is not particularly important to jurors’ recommendations for civil commitment (e.g., Boccaccini et al., 2013; Varela et al., 2014). For example, there is evidence of motivated reasoning among SVP jurors who vote for civil commitment (e.g., Carlsmith et al., 2007; Krauss et al., 2018; Scurich & Krauss, 2013), such that some will decide that even a 1% chance of sexual reoffending is sufficient to meet the statutory requirement of “likely” to reoffend (Knighton et al., 2014). Relatedly, in the PSO study (Chapter 6), we found that mock jurors’ endorsement of harsher sentencing and restrictive management in the community for PCOS significantly predicted a vote for civil commitment after receiving the Static-99R report.

Moreover, previous research has also found that punitiveness seems to underlie some jurors’ decisions to vote for civil commitment, particularly when they think the initial term of imprisonment was insufficient (Carlsmith et al., 2007; Krauss & Scurich, 2014). We also found that across all three SVP studies, mock jurors’ importance ratings of the impact of the crime on the victim significantly increased the odds of recommending civil commitment. The findings from the three SVP studies in this dissertation add to previous research that suggests that decisions about SVP civil commitment are likely motivated by factors other than an objective estimate of the SVP respondent’s likelihood of engaging in sexual reoffending.

It is difficult to square the growing body of evidence about case factors and internal factors that have the potential to bias SVP jurors with the constitutional directive that SVP civil commitment be applied only to those who meet the requisite legal criteria, including that it be “more likely than not” that the SVP respondent will engage in sexual reoffending if

released to the community (Knighton et al., 2014). Once a person has been civilly committed as an SVP, it can be difficult, if not impossible, to ever regain their freedom as evidenced by the low rates of release observed in many jurisdictions that have SVP statutes (see, e.g., Fanniff et al., 2010; Gookin, 2007; Krauss & Scurich, 2014). The findings from the three SVP studies in this dissertation, in addition to previous studies (e.g., Carlsmith et al., 2007; Krauss & Scurich, 2014; Socia et al., 2019), provide growing evidence that casts doubt on whether civil commitment is imposed for non-punitive purposes and only for those few who are at the highest risk of sexual recidivism (*Kansas v. Crane*, 2002; *Kansas v. Hendricks*, 1997).

Limitations and Directions for Future Studies

The experiment testing whether there was a biasing effect of media exposure and forensic mental health practitioners' attitudes toward offenders on HCR-20 scoring (Chapter 3) was limited by a small sample size. Although there were observed differences in the HCR-20 scoring related to practitioners' attitudes toward offenders, it would be informative to replicate the study with a larger sample size. One way to obtain a larger sample size would be to replicate the study through a multi-lab collaboration.

In addition, the HCR-20 study does not account for expected variability between examiners, which is an issue related to reliability between experts (see, for example, Dror & Murrie, 2018; Huang & Bull, 2020). Dror and Murrie (2018) distinguish between *reliability* and *biasability* of forensic evaluator's decisions. As defined by Dror and Murrie, reliability "refers to the consistency, reproducibility, or repeatability of decisions, regardless of bias" and biasability "refers to the potential effects of irrelevant contextual information and other biases that may impact the decision" of forensic science or legal practitioners (p. 11). Both reliability and biasability are important in explaining variations in expert decision making in terms of both the process and the outcome (Dror & Murrie, 2018; Mossman, 2013). In other words, when variations are obtained between forensic experts related to a forensic risk evaluation, these are likely attributable to both reliability and biasability issues. Additional experimental studies, such as that conducted by Murrie and colleagues (2009), could help disentangle variations in SRAI scoring that are attributable to problems in reliability from variations that may reflect the operation of bias.

With the exception of a handful of studies about adversarial allegiance (Murrie et al., 2008, 2009, 2013; Murrie & Boccaccini, 2015), there is limited research regarding biasability in experts regarding their conclusions in forensic evaluations. We attempted to address this shortcoming as it relates to conclusions reached with the use of the HCR-20 and the effects of two types of bias: (1) the influence of negative media portrayals of the subject of the evaluation

(an external source of bias); and (2) the influence of forensic evaluator's attitudes toward offenders (an internal source of bias). However, we note that inter-rater reliability of an SRAI (a facet of reliability between forensic evaluators) and the standard error of measurement should be accounted for when analyzing differences between expert ratings (see, for example, Murrie et al., 2009). Although the HCR-20^{v3} has demonstrated good-to-excellent interrater reliability in a few studies (Douglas & Belfrage, 2014; Kötter et al., 2014), we could not exclude the possibility that the variations in completion of the HCR-20 in our study may have been due to limitations in reliability rather than the influence of bias because we did not account for the standard error of measurement or inter-rater reliability of the HCR-20^{v3} in our statistical model. Future researchers should undertake experimental research that enables them to differentiate between the effects of reliability and biasability on variations in SRAI completion and outcomes.

Second, the series of three SVP studies (Chapters 4–6) lacked a deliberation component, which *may* have implications for ecological validity. While some studies have found that deliberation influences jury decision making (Lynch & Haney, 2009), others have found that a jury's ultimate verdict is strongly predicted by the distribution of verdicts among individual jurors (see Kovera & Levett, 2015). Therefore, it would be useful to examine whether a written copy of an SRAI report has a similarly mitigating effect on bias in a group of jurors.

Relatedly, in the three SVP studies, there was quite literally no risk of sexual recidivism and no real-world consequences stemming from the jurors' decision to vote for civil commitment or not. Therefore, the extent to which a written version of the SRAI risk results affects jurors' perceptions of the SVP respondent's sexual recidivism risk is uncertain. However, conducting a study in which jurors in actual SVP cases are or are not provided with a written version of the SRAI risk estimate seems quite feasible. Such a study would obviously have much higher ecological validity.

It may also be possible to carry out an experimental study in actual SVP trials in which cases are randomly assigned to determine whether jurors will or will not receive a text version of the risk factors and associated calculations made by the forensic evaluator(s). However, researchers should carefully weigh the risks and benefits of conducting an experimental study in actual cases, as recent controversy over research conducted by Harvard Law School's Access to Justice Lab illustrates (Charles Hamilton Houston Institute for Race & Justice at Harvard Law, 2021; Karp, 2021). In a study conducted in Dane County, Wisconsin, the Access to Justice Lab (A2J Lab) employed random assignment to determine whether the judge charged with deciding an accused's type of bail and level of pretrial supervision was provided with a paper printout of the results of a pretrial risk assessment instrument (Harvard Law School Access to Justice Lab, 2016). This study, and other randomly-controlled trial research conducted by the

A2J Lab in criminal cases has been criticized, in part, for the potential to disadvantage certain groups (Charles Hamilton Houston Institute for Race & Justice at Harvard Law, 2021; Karp, 2021). Although research about pretrial release may raise different ethical and Constitutional concerns (e.g., access to counsel; presumption of innocence) than SVP civil commitment trials, researchers should nevertheless consider the effect of scientific experimentation in cases in which an individual's liberty is at stake.

Finally, I note what is a limitation of not only the SVP studies we conducted—but also those of others (e.g., Krauss & Scurich, 2014; Scurich & Krauss, 2014)—which is that we asked mock jurors to estimate the *SVP respondent's* risk of sexual recidivism based on group information. Yet, the Static-99R only provides information about the observed sexual recidivism rate among individuals with the same score as the SVP respondent. Therefore, asking mock jurors to estimate the SVP respondent's risk of sexual recidivism based on observed recidivism rates from a Static-99R score in a normative sample, although it essentially *mimics what occurs in actual SVP cases*, is problematic. An illustration can help explain: a Static-99R score of four is associated with five-year recidivism rates of between 10–12% (Hanson et al., 2016). In other words, 10-12 people out of a group of 100 sex offenders with a score of four are predicted to be convicted of another sexual offense within five years after release to the community.

What the Static-99R does not provide is information to help professionals or legal decision-makers discern whether a specific SVP respondent is more likely to be one of the 10-12 who sexually recidivate or one of the 88 to 90 people who do not. For example, one SVP evaluator in Florida (DeClue & Rice, 2016) reported that out of 73 men he evaluated for sexual recidivism risk, he predicted 54 were likely to engage in sexual violence if released. After a 10-year period, only five (9%) of those 54 men had a new sexual *charge or conviction*. Comparatively, among the 19 men the evaluator estimated as unlikely to engage in sexual recidivism if released, 18 did not (one was charged but not convicted).

This example demonstrates the inherent limitations of attempting to predict an individual's behavior based on group information. Nevertheless, whereas jurors in SVP cases are asked to decide about an SVP respondent's likelihood of sexual recidivism risk and thereby make a dichotomous decision to vote for civil commitment or release, it is still preferable for those decisions to be based on an SRAI rather than not.

Conclusion

Forensic mental health professionals play a significant role in legal cases and judges tend to follow their recommendations (Dolores & Redding, 2009; Gowensmith & McCallum, 2019; MacLean et al., 2019). Recently, numerous researchers in the field of forensic psychology have called for increased attention to the potential for bias to influence forensic evaluations (e.g., Dror & Murrie, 2018; Neal & Brodsky, 2016; Zapf & Dror, 2017). Therefore, the first aim of the research in this dissertation was to examine whether bias can influence SRAIs commonly used in forensic risk evaluations. The experimental findings from the HCR-20 study (Chapter 3) provide evidence that indicates that SRAI ratings can be affected by bias on the part of the evaluator. While SRAIs may be useful in correcting some types of bias, they should not be viewed as a “cure,” nor should they be the sole strategy evaluators rely upon. Hence, there is a pressing need for research regarding the effectiveness of bias correction strategies that have been suggested in the literature.

The second aim of this dissertation was to examine the joint effects of biasing information and a risk estimate derived from an SRAI on mock jurors’ perceptions of sexual recidivism risk and decisions about SVP civil commitment. The SRAI risk estimate was somewhat successful in eliminating the biasing effects of the SVP respondent’s “likability,” but was not effective in eliminating the biasing effects of mock jurors’ preexisting beliefs about sexual recidivism rates among PCSOs, nor their preexisting perceptions of PCSOs.

Cognitive bias is an inherent part of the human condition. Yet, the risk of reoffending is crucial to decisions about how to manage and reduce risk in forensic and justice applications. As recidivism risk and SRAIs are playing an increasingly important role in forensic mental health and legal settings, there is a need to examine the extent to which SRAIs can be relied upon to produce an objective estimate of recidivism risk. Forensic risk assessment cannot achieve the goal of helping legal decision-makers make “better-informed” decisions (Heilbrun, 1997) if what is thought to be an objective measure of risk is compromised by bias. The findings from the research in this dissertation indicate that an SRAI is not a certificate of objectivity, and that risk may be in the eye of the beholder. Researchers should therefore continue efforts to develop ways to mitigate the influence of bias in forensic risk assessment and legal decision-making; justice and fairness depend on it.

References

References

- Abel, M. H., & Watters, H. (2005). Attributions of guilt and punishment as functions of physical attractiveness and smiling. *The Journal of Social Psychology*, 145(6), 687–702. <https://doi.org/10.3200/SOCP.145.6.687-703>
- Acklin, M. W., Fuger, K., & Gowensmith, W. (2015). Examiner agreement and judicial consensus in forensic mental health evaluations. *Journal of Forensic Psychology Practice*, 15(4), 318–343. <https://doi.org/10.1080/15228932.2015.1051447>
- Adviesbureau Van Montfoort & Reklassering Nederland. (2004). *RISc versie 1.0. Recidive Inschattings Schalen. Handleiding [RISc version 1.0. Risk Assessment Scales. Manual]*. Flevodruk.
- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34(3), 341–382. <https://doi.org/10.1177/0011000005285875>
- Alicke, M. D., & Zell, E. (2009). Social attractiveness and blame. *Journal of Applied Social Psychology*, 39(9), 2089–2105. <https://doi.org/10.1111/j.1559-1816.2009.00517.x>
- Alper, M., & Durose, M. R. (2019). *Recidivism of sex offenders released from state prison: A 9-year follow-up (2005-14)* (NCJ 251773). Bureau of Justice Statistics. <https://www.bjs.gov/content/pub/pdf/rsorspspyfu0514.pdf>
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist*, 68(1), 7–19. <https://doi.org/10.1037/a0029889>
- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory-Revised*. Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17(1), 19–52.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7–27. <https://doi.org/10.1177/001128705281756>
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07
- Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*, 49(3), 323–330.
- Ask, K., & Granhag, P. A. (2007). Motivational bias in criminal investigators' judgments of witness reliability. *Journal of Applied Social Psychology*, 37(3), 561–591. <https://doi.org/10.1111/j.1559-1816.2007.00175.x>
- Ask, K., Rebelius, A., & Granhag, P. A. (2008). The 'elasticity' of criminal evidence: A moderator of investigator bias. *Applied Cognitive Psychology*, 22(9), 1245–1259. <https://doi.org/10.1002/acp.1432>
- Ask, K., Reinhard, M.-A., Marksteiner, T., & Granhag, P. A. (2011). Elasticity in evaluations of criminal evidence: Exploring the role of cognitive dissonance. *Legal and Criminological Psychology*, 16(2), 289–306. <https://doi.org/10.1348/135532510X510153>
- Association for the Treatment of Sexual Abusers. (2010). *Civil commitment of sexually violent predators*. ATSA. <https://www.atsa.com/civil-commitment-sexually-violent-predators>
- Baker v. DesLauriers, 6:14-cv-01356-JTM-KGG (U.S. District Court, D. Kansas 2014). <https://www.clearinghouse.net/chDocs/public/MH-KS-0001-0001.pdf>

- Bakhshay, S., & Haney, C. (2018). The media's impact on the right to a fair trial: A content analysis of pretrial publicity in capital cases. *Psychology, Public Policy, and Law*, 24(3), 326–340. <https://doi.org/10.1037/law0000174>
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28(4), 490–521. <https://doi.org/10.1177/009385480102800406>
- Bartneck, C., Duenser, A., Molchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLOS ONE*, 10(4), e0121595. <https://doi.org/10.1371/journal.pone.0121595>
- Batastini, A. B., Hoeffner, C. E., Vitacco, M. J., Morgan, R. D., Coaker, L. C., & Lester, M. E. (2019). Does the format of the message affect what is heard? A two-part study on the communication of violence risk assessment data. *Journal of Forensic Psychology Research and Practice*, 19(1), 44–71. <https://doi.org/10.1080/24732850.2018.1538474>
- Batastini, A. B., Vitacco, M. J., Coaker, L. C., & Lester, M. E. (2019). Communicating violence risk during testimony: Do different formats lead to different perceptions among jurors? *Psychology, Public Policy, and Law*, 25(2), 92–106. <https://doi.org/10.1037/law0000196>
- Beale, S. S. (2006). The news media's influence on criminal justice policy: How market-driven news promotes punitiveness. *William and Mary Law Review*, 48(2), 397–482.
- Bechtel, K., Holsinger, A. M., Lowenkamp, C. T., & Warren, M. J. (2017). A meta-analytic review of pretrial research: Risk assessment, bond type, and interventions. *American Journal of Criminal Justice*, 42(2), 443–467. <https://doi.org/10.1007/s12103-016-9367-1>
- 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 and Human Behavior*, 13(1), 79–87. <https://doi.org/10.1007/BF01056164>
- Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sex Abuse*, 19(2), 135–153. <https://doi.org/10.1007/s11194-007-9044-5>
- Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368. <https://doi.org/10.1093/pan/mpr057>
- Bertrams, A., & Dickhäuser, O. (2010). University and school students' motivation for effortful thinking: Factor structure, reliability, and validity of the German Need for Cognition scale. *European Journal of Psychological Assessment*, 26(4), 263–268. <https://doi.org/10.1027/1015-5759/a000035>
- Blais, J. (2015). Preventative detention decisions: Reliance on expert assessments and evidence of partisan allegiance within the Canadian context. *Behavioral Sciences & the Law*, 33(1), 74–91. <https://doi.org/10.1002/bls.2155>
- Blumenthal, S., Huckle, C., Czornyj, R., Craissati, J., & Richardson, P. (2010). The role of affect in the estimation of risk. *Journal of Mental Health*, 19(5), 444–451. <https://doi.org/10.3109/09638231003728083>
- Boccaccini, M. T., Chevalier, C. S., Murrie, D. C., & Varela, J. G. (2017). Psychopathy Checklist-Revised use and reporting practices in sexually violent predator evaluations. *Sexual Abuse: A Journal of Research and Treatment*, 29(6), 592–614. <https://doi.org/10.1177/1079063215612443>

-
- Boccaccini, M. T., Murrie, D. C., Rufino, K. A., & Gardner, B. O. (2014). Evaluator differences in Psychopathy Checklist-Revised factor and facet scores. *Law and Human Behavior*, 38(4), 337–345. <https://doi.org/10.1037/lhb0000069>
- Boccaccini, M. T., Murrie, D. C., & Turner, D. B. (2014). Jurors' views on the value and objectivity of mental health experts testifying in sexually violent predator trials. *Behavioral Sciences & the Law*, 32(4), 483–495. <https://doi.org/10.1002/bls.2129>
- Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual offenders. *Psychological Assessment*, 29(6), 611–623. <https://doi.org/10.1037/pas0000377>
- Boccaccini, M. T., Rufino, K. A., Jeon, H., & Murrie, D. C. (2017). Does the predictive validity of psychopathy ratings depend on the clinical experience of the raters? *International Journal of Forensic Mental Health*, 16(2), 130–138. <https://doi.org/10.1080/14999013.2016.1266421>
- Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). Do some evaluators report consistently higher or lower PCL-R scores than others? Findings from a statewide sample of sexually violent predator evaluations. *Psychology, Public Policy, and Law*, 14(4), 262–283. <https://doi.org/10.1037/a0014523>
- Boccaccini, M. T., Turner, D. B., Murrie, D. C., Henderson, C. E., & Chevalier, C. (2013). Do scores from risk measures matter to jurors? *Psychology, Public Policy, and Law*, 19(2), 259–269. <https://doi.org/10.1037/a0031354>
- Boccaccini, M. T., Turner, D. B., Murrie, D. C., & Rufino, K. A. (2012). Do PCL-R scores from state or defense experts best predict future misconduct among civilly committed sex offenders? *Law and Human Behavior*, 36(3), 159–169. <https://doi.org/10.1037/h0093949>
- Bodenhausen, G. V., & Wyer, R. S. (1985). Effects of stereotypes on decision making and information-processing strategies. *Journal of Personality and Social Psychology*, 48(2), 267–282.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. British Columbia Institute Against Family Violence.
- Boerner, D. (1992). Confronting violence: In the act and in the word. *University of Puget Sound Law Review*, 15(3), 526–577.
- Bogaerts, S., Spreen, M., ter Horst, P., & Gerlsma, C. (2018). Predictive validity of the HKT-R risk assessment tool: Two and 5-year violent recidivism in a nationwide sample of Dutch forensic psychiatric patients. *International Journal of Offender Therapy and Comparative Criminology*, 62(8), 2259–2270. <https://doi.org/10.1177/0306624X17717128>
- Bojcenko, M. N., & Sivasubramaniam, D. (2019). The role of prior punishment in preventive detention decisions. *Psychology, Crime & Law*, 26(5), 479–506. <https://doi.org/10.1080/1068316X.2019.1690654>
- Bonta, J., & Andrews, D. A. (2007). *Risk-Need-Responsivity model for offender assessment and rehabilitation* (No. 2007–06). <https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf>
- Bonta, J., & Motiuk, L. L. (1990). Classification to halfway houses: A quasi-experimental evaluation. *Criminology*, 28(3), 497–506.
- Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior*, 27(3), 312–329.
- Boots, D. P., & Heide, K. M. (2006). Parricides in the media: A content analysis of available reports across cultures. *International Journal of Offender Therapy and Comparative Criminology*, 50(4), 418–445. <https://doi.org/10.1177/0306624X05285103>

- Borhart, H. M., & Plumm, K. M. (2015). The effects of sex offender stereotypes on potential juror beliefs about conviction, victim blame and perceptions of offender mental stability. *Applied Psychology in Criminal Justice*, 11(3), 207–219.
- Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyaries, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and Human Behavior*, 41(1), 13–28. <https://doi.org/10.1037/lhb0000223>
- Borum, R., Bartel, P., & Forth, A. E. (2002). *Manual for the Structured Assessment of Violence Risk in Youth (SAVRY)* (consultation edition, version 1). University of South Florida.
- Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *The Journal of Psychiatry and Law*, 21(1), 35–76.
- Bourget, D., Gagne, P., & Labelle, M.-E. (2007). Parricide: A comparative study of matricide versus patricide. *The Journal of the American Academy of Psychiatry and the Law*, 35(3), 306–312.
- Boyll, J. R. (1991). Psychological, cognitive, personality and interpersonal factors in jury verdicts. *Law and Psychology Review*, 15, 163–184.
- Bridge, C., & Marić, M. (2019). Cognitive biases in forensic science training and education. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and Fake News in Education*. Information Age Publishing.
- Brodsky, S. L. (2013). The pull to affiliate and allegiance effects. In *Testifying in court: Guidelines and maxims for the expert witness* (2nd ed., pp. 161–165). American Psychological Association. <https://doi.org/10.1037/14037-039>
- Brown, B., & Rakow, T. (2016). Understanding clinicians' use of cues when assessing the future risk of violence: A clinical judgement analysis in the psychiatric setting. *Clinical Psychology & Psychotherapy*, 23(2), 125–141. <https://doi.org/10.1002/cpp.1941>
- Brown, S. (1999). Public attitudes toward the treatment of sex offenders. *Legal and Criminological Psychology*, 4(2), 239–252. <https://doi.org/10.1348/135532599167879>
- Bruschke, J., Gonis, A., Hill, S. A., Fiber-Ostrow, P., & Loges, W. (2016). The influence of heterogeneous exposure and pre-deliberation queries on pretrial publicity effects. *Communication Monographs*, 83(4), 521–534. <https://doi.org/10.1080/03637751.2016.1182639>
- Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, 13(2), 149–154. <https://doi.org/10.1177/1745691617706516>
- Burgess, E. W. (1928). Factors making for success or failure on parole. In A. A. Bruce, E. W. Burgess, & A. J. Harno (Eds.), *The workings of the indeterminate sentence law and the parole system in Illinois* (pp. 221–234). Illinois State Board of Parole.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459.
- Butler, B. (2007). The role of death qualification in jurors' susceptibility to pretrial publicity. *Journal of Applied Social Psychology*, 37(1), 115–123. <https://doi.org/10.1111/j.0021-9029.2007.00150.x>
- Butler, B. (2008). The role of death qualification in venirepersons' susceptibility to victim impact statements. *Psychology, Crime & Law*, 14(2), 133–141. <https://doi.org/10.1080/10683160701483534>
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>

-
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119(2), 197–253.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13
- Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology*, 45(4), 805–818.
- Camilleri, A., Abarno, D., Bird, C., Coxon, A., Mitchell, N., Redman, K., Sly, N., Wills, S., Silenieks, E., Simpson, E., & Lindsay, H. (2019). A risk-based approach to cognitive bias in forensic science. *Science & Justice*, 59(5), 533–543. <https://doi.org/10.1016/j.scijus.2019.04.003>
- Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36(6), 567–590. <https://doi.org/10.1177/0093854809333610>
- Carlsmith, K. M., Monahan, J., & Evans, A. (2007). The function of punishment in the “civil” commitment of sexually violent predators. *Behavioral Sciences & the Law*, 25(4), 437–448. <https://doi.org/10.1002/bls.761>
- Carroll, A. (2007). Are violence risk assessment tools clinically useful? *Australian & New Zealand Journal of Psychiatry*, 41(4), 301–307. <https://doi.org/10.1080/00048670701213237>
- Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, 29(6), 2156–2160. <https://doi.org/10.1016/j.chb.2013.05.009>
- Cauley, D. R. (2007). Risk assessment of civilly committed sexually violent predators: Static versus dynamic orientations. *Forensic Examiner*, 16(1), 18–24.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chapleau, K. M., & Oswald, D. L. (2013). Status, threat, and stereotypes: Understanding the function of rape myth acceptance. *Social Justice Research*, 26(1), 18–41. <https://doi.org/10.1007/s11211-013-0177-z>
- Chappell, A. T., Maggard, S. R., & Higgins, J. L. (2013). Exceptions to the rule? Exploring the use of overrides in detention risk assessment. *Youth Violence and Juvenile Justice*, 11(4), 332–348. <https://doi.org/10.1177/1541204012470849>
- Charles Hamilton Houston Institute for Race & Justice at Harvard Law. (2021, January 26). *THREAD: there is much left intentionally unsaid in this piece about efforts at HLS and elsewhere to run RCTs in the criminal legal system, deceptively & inaccurately framed as applying scientific rigor to tricky public policy unknowns. Let's talk about it.* [Tweet]. Twitter. <https://twitter.com/houstoninst/status/1354197891427028998>
- Charman, S. D. (2013). The forensic confirmation bias: A problem of evidence integration, not just evidence evaluation. *Journal of Applied Research in Memory and Cognition*, 2(1), 56–58. <https://doi.org/10.1016/j.jarmac.2013.01.010>
- 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>

- Childs, K., Frick, P. J., Ryals, J. S., Lingonblad, A., & Villio, M. J. (2014). A comparison of empirically based and structured professional judgment estimation of risk using the Structured Assessment of Violence Risk in Youth. *Youth Violence and Juvenile Justice*, 12(1), 40–57. <https://doi.org/10.1177/1541204013480368>
- Cocozza, J. J., & Steadman, H. J. (1976). The failure of psychiatric predictions of dangerousness: Clear and convincing evidence. *Rutgers Law Review*, 29(5), 1084–1101.
- Cocozza, J. J., & Steadman, H. J. (1977). Prediction in psychiatry: An example of misplaced confidence in experts. *Social Problems*, 25(3), 265–276.
- Coffey, C. A., Batastini, A. B., & Vitacco, M. J. (2018). Clues from the digital world: A survey of clinicians' reliance on social media as collateral data in forensic evaluations. *Professional Psychology: Research and Practice*, 49(5–6), 345–354. <https://doi.org/10.1037/pro0000206>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.
- Cohen, T. H., Lowenkamp, C. T., Bechtel, K., & Flores, A. W. (2020). Risk assessment overrides: Shuffling the risk deck without any improvements in prediction. *Criminal Justice and Behavior*, 47(12), 1609–1629. <https://doi.org/10.1177/0093854820953449>
- Cohen, T. H., Pendergast, B., & VanBenschoten, S. W. (2016). *Examining overrides of risk classifications for offenders on federal supervision*. 80(1), 12–21.
- Comartin, E. B., Kernsmith, P. D., & Kernsmith, R. M. (2009). Sanctions for sex offenders: Fear and public policy. *Journal of Offender Rehabilitation*, 48(7), 605–619. <https://doi.org/10.1080/10509670903196066>
- Commons, M. L., Miller, P. M., & Gutheil, T. G. (2004). Expert witness perceptions of bias in experts. *The Journal of the American Academy of Psychiatry and the Law*, 32(1), 6.
- Cooper, B., & Rizzo, T. (2013, September 14). Sex offender treatment programs are under scrutiny. *The Kansas City Star*. <https://www.kansascity.com/news/local/article327407.html>
- Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: A systematic review. *Forensic Science International*, 297, 35–46. <https://doi.org/10.1016/j.forsciint.2019.01.016>
- Corteen, K., & Steele, R. (2018). A criminal injustice system? Sex offender suspects and defendants. *Liverpool Law Review*, 39(3), 265–277. <https://doi.org/10.1007/s10991-018-9219-9>
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.
- Cox, J., Fairfax-Columbo, J., DeMatteo, D., Vitacco, M. J., Kopkin, M. R., Parrott, C. T., & Bownes, E. (2018). An update and expansion on the role of the Violence Risk Appraisal Guide and Historical Clinical Risk Management-20 in United States case law. *Behavioral Sciences & the Law*, 36(5), 517–531. <https://doi.org/10.1002/blsl.2376>
- Craig, L. A., Browne, K. D., & Stringer, I. (2003). Risk scales and factors predictive of sexual offence recidivism. *Trauma, Violence, & Abuse*, 4(1), 45–69. <https://doi.org/10.1177/1524838002238945>
- Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, 11(1), 65–84. <https://doi.org/10.1080/13552600410001667733>
- Cramer, R. J., Brodsky, S. L., & DeCoster, J. (2009). Expert witness confidence and juror personality: Their impact on credibility and persuasion in the courtroom. *The Journal of the American Academy of Psychiatry and the Law*, 37(1), 63–74.

-
- Crosbie, K. (2018, November 30). In memoriam: Lives lost while imprisoned after their release date in Kansas. *Just Future Project*. <https://ajustfuture.org/in-memoriam-lives-lost-while-imprisoned-after-their-release-date-in-kansas/>
- Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality & Safety*, 22(Supplement 2), ii65–ii72. <https://doi.org/10.1136/bmjqs-2012-001713>
- Cucolo, H. E., & Perlin, M. L. (2013). “They’re planting stories in the press”: The impact of media distortions on sex offender law and policy. *University of Denver Criminal Law Review*, 3(1), 185–245.
- Cunningham, M. D., & Reidy, T. J. (2001). A matter of life or death: Special considerations and heightened practice standards in capital sentencing evaluations. *Behavioral Sciences & the Law*, 19(4), 473–490. <https://doi.org/10.1002/bls.460>
- Cunningham, M. D., & Reidy, T. J. (2002). Violence risk assessment at federal capital sentencing: Individualization, generalization, relevance, and scientific standards. *Criminal Justice and Behavior*, 29(5), 512–537. <https://doi.org/10.1177/009385402236731>
- Curley, L. J., Munro, J., Lages, M., MacLean, R., & Murray, J. (2020a). Assessing cognitive bias in forensic decisions: A review and outlook. *Journal of Forensic Sciences*, 65(2), 354–360. <https://doi.org/10.1111/1556-4029.14220>
- Curley, L. J., Munro, J., Lages, M., MacLean, R., & Murray, J. (2020b). Authors’ response. *Journal of Forensic Sciences*, 65(2), 668–670. <https://doi.org/10.1111/1556-4029.14273>
- Daftary-Kapur, T., Dumas, R., & Penrod, S. D. (2010). Jury decision-making biases and methods to counter them. *Legal and Criminological Psychology*, 15(1), 133–154. <https://doi.org/10.1348/135532509X465624>
- Daftary-Kapur, T., Penrod, S. D., O’Connor, M., & Wallace, B. (2014). Examining pretrial publicity in a shadow jury paradigm: Issues of slant, quantity, persistence and generalizability. *Law and Human Behavior*, 38(5), 462–477. <https://doi.org/10.1037/lhb0000081>
- D’Agostino, P. R., & Fincher-Kiefer, R. (1992). Need for cognition and the correspondence bias. *Social Cognition*, 10(2), 151–163. <https://doi.org/10.1521/soco.1992.10.2.151>
- Dantas, S., Santos, A., Dias, I., Dinis-Oliveira, R. J., & Magalhães, T. (2014). Parricide: A forensic approach. *Journal of Forensic and Legal Medicine*, 22(1), 1–6. <https://doi.org/10.1016/j.jflm.2013.11.008>
- Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993).
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668–1674.
- de Ruiter, C. (2016). Risk assessment in the Netherlands. In J. P. Singh, S. Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 270–279). Oxford University Press.
- de Ruiter, C., & Hildebrand, M. (2007). Risk assessment and treatment in Dutch forensic psychiatry. *Netherlands Journal of Psychology*, 63(4), 152–160. <https://doi.org/10.1007/BF03061078>
- de Ruiter, C., & Kaser-Boyd, N. (2015). *Forensic psychological assessment in practice: Case studies* (1st ed.). Routledge. <https://doi.org/10.4324/9781315723310>
- de Vogel, V., & de Ruiter, C. (2004). Differences between clinicians and researchers in assessing risk of violence in forensic psychiatric patients. *Journal of Forensic Psychiatry & Psychology*, 15(1), 145–164. <https://doi.org/10.1080/14788940410001655916>

- de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law*, 12(3), 321–336. <https://doi.org/10.1080/10683160600569029>
- de Vogel, V., de Ruiter, C., Bouman, Y. H. A., & de Vries Robb  , M. (2009). *SAPROF. Guidelines for the assessment of protective factors for violence risk*. Forum Educatief.
- de Vogel, V., de Ruiter, C., Bouman, Y. H. A., & de Vries Robb  , M. (2012). *SAPROF. Guidelines for the assessment of protective factors for violence risk* (2nd ed.). Forum Educatief.
- de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3(2), 149–165. <https://doi.org/10.1080/14999013.2004.10471204>
- de Vogel, V., van den Broek, E., & de Vries Robb  , M. (2014). The use of the HCR-20^{v3} in Dutch forensic psychiatric practice. *International Journal of Forensic Mental Health*, 13(2), 109–121. <https://doi.org/10.1080/14999013.2014.906518>
- de Vries Robb  , M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry & Psychology*, 24(4), 440–457. <https://doi.org/10.1080/14789949.2013.818162>
- DeClue, G., & Rice, A. (2016). Florida's released "Sexually Violent Predators" are not "high risk." *Open Access Journal of Forensic Psychology*, 8(1), 22–51.
- Deitchman, M. A., Kennedy, W. A., & Beckham, J. C. (1991). Self-selection factors in the participation of mental health professionals in competency for execution evaluations. *Law and Human Behavior*, 15(3), 287–303. <https://doi.org/10.1007/BF01061714>
- DeMatteo, D., Edens, J. F., Galloway, M., Cox, J., Smith, S. T., & Formon, D. (2014). The role and reliability of the Psychopathy Checklist—Revised in U.S. sexually violent predator evaluations: A case law survey. *Law and Human Behavior*, 38(3), 248–255. <https://doi.org/10.1037/lhb0000059>
- DeMatteo, D., Hart, S. D., Heilbrun, K., Boccaccini, M. T., Cunningham, M. D., Douglas, K. S., Dvoskin, J. A., Edens, J. F., Guy, L. S., Murrie, D. C., Otto, R. K., Packer, I. K., & Reidy, T. J. (2020). Statement of concerned experts on the use of the Hare Psychopathy Checklist—Revised in capital sentencing to assess risk for institutional violence. *Psychology, Public Policy, and Law*, 26(2), 133–144. <https://doi.org/10.1037/law0000223>
- DeMatteo, D., Murphy, M., Galloway, M., & Krauss, D. A. (2015). A national survey of United States Sexually Violent Person legislation: Policy, procedures, and practice. *International Journal of Forensic Mental Health*, 14(4), 245–266. <https://doi.org/10.1080/14999013.2015.1110847>
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, 66(5), 819–829.
- Dernevik, M., Falkheim, M., Holmqvist, R., & Sandell, R. (2001). Implementing risk assessment procedures in a forensic psychiatric setting: Clinical judgement revisited. In D. Farrington, C. R. Hollin, & M. McMurran, *Sex and violence: The psychology of crime and risk assessment* (pp. 83–101). Routledge.
- Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services*, 13(3), 206–222. <https://doi.org/10.1037/ser0000075>

-
- Desmarais, S. L., & Singh, J. P. (2013). *Risk assessment instruments validated and implemented in correctional settings in the United States*. <https://csgjusticecenter.org/wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf>
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568–584.
- Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, 29(9), 1120–1132. <https://doi.org/10.1177/0146167203254536>
- Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, 75(1), 53–69.
- Dolores, J. C., & Redding, R. E. (2009). The effects of different forms of risk communication on judicial decision making. *International Journal of Forensic Mental Health*, 8(2), 142–146. <https://doi.org/10.1080/14999010903199407>
- Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Sage.
- Douglas, K. S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *International Journal of Forensic Mental Health*, 13(2), 130–139. <https://doi.org/10.1080/14999013.2014.908429>
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *HCR-20 (Version 3): Assessing risk of violence—User guide*. Mental Health, Law, and Policy Institute, Simon Fraser University.
- Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{v3}): Development and overview. *International Journal of Forensic Mental Health*, 13(2), 93–108. <https://doi.org/10.1080/14999013.2014.906519>
- Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, 29(5), 617–658. <https://doi.org/10.1177/009385402236735>
- Douglas, K. S., Shaffer, C., Blanchard, A. J. E., Guy, L. S., Reeves, K. A., & Weir, J. (2014). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography* (HCR-20 Violence Risk Assessment White Paper Series, #1, p. 220). Mental Health, Law, and Policy Institute, Simon Fraser University. <http://hcr-20.com/hcr/wp-content/uploads/2013/03/HCR-20-Annotated-Bibliography-Version-12-January-2014.pdf>
- Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, 11(3), 347–383. <https://doi.org/10.1037/1076-8971.11.3.347>
- Dover, T. L., Matthews, M., Krauss, D. A., & Levin, S. (2012). Just world beliefs, expert psychological testimony, and verdicts: A mediational model. *Analyses of Social Issues and Public Policy*, 12(1), 340–363. <https://doi.org/10.1111/j.1530-2415.2011.01278.x>
- Doyle, M., & Dolan, M. (2007). Standardized risk assessment. *Psychiatry*, 6(10), 409–414.
- Doyle, M., Power, L. A., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting post-discharge community violence in England and Wales using the HCR-20^{v3}. *International Journal of Forensic Mental Health*, 13(2), 140–147. <https://doi.org/10.1080/14999013.2014.906517>

- Dror, I. E. (2009). On proper research and understanding of the interplay between bias and decision outcomes. *Forensic Science International*, 191(1–3), e17–e18. <https://doi.org/10.1016/j.forsciint.2009.03.012>
- Dror, I. E. (2012). Expectations, contextual information, and other cognitive influences in forensic laboratories. *Science & Justice*, 52(2), 132. <https://doi.org/10.1016/j.scijus.2012.03.005>
- Dror, I. E. (2018). Biases in forensic experts. *Science*, 360(6386), 243.
- Dror, I. E. (2020). Cognitive and human factors in expert decision making: Six fallacies and the eight sources of bias. *Analytical Chemistry*, 92(12), 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., Morgan, R. M., Rando, C., & Nakhaeizadeh, S. (2017). Letter to the editor: 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., & Murrie, D. C. (2018). A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law*, 24(1), 11–23. <https://doi.org/10.1037/law0000140>
- Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Letter to the editor: 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>
- Ducat, L., Thomas, S., & Blood, W. (2009). Sensationalising sex offenders and sexual recidivism: Impact of the Serious Sex Offender Monitoring Act 2005 on media reportage. *Australian Psychologist*, 44(3), 156–165. <https://doi.org/10.1080/00050060903127499>
- Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, 25(5), 582–602. <https://doi.org/10.1521/soco.2007.25.5.582>
- Edens, J. F., Buffington-Vollum, J. K., Keilen, A., Roskamp, P., & Anthony, C. (2005). Predictions of future dangerousness in capital murder trials: Is it time to “disinvent the wheel?” *Law and Human Behavior*, 29(1), 55–86. <https://doi.org/10.1007/s10979-005-1399-x>
- 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(2), 447–456. <https://doi.org/10.1037/pas0000048>
- Edens, J. F., Smith, S. T., Magyar, M. S., Mullen, K., Pitta, A., & Petrila, J. (2012). “Hired guns,” “charlatans,” and their “voodoo psychobabble”: Case law references to various forms of perceived bias among mental health expert witnesses. *Psychological Services*, 9(3), 259–271. <https://doi.org/10.1037/a0028264>
- Edmond, G., Tangen, J. M., Searston, R. A., & Dror, I. E. (2015). Contextual bias and cross-contamination in the forensic sciences: The corrosive implications for investigations, plea bargains, trials and appeals. *Law, Probability and Risk*, 14(1), 1–25. <https://doi.org/10.1093/lpr/mgu018>
- Elbogen, E. B. (2002). The process of violence risk assessment: A review of descriptive research. *Aggression and Violent Behavior*, 7(6), 591–604.
- Elbogen, E. B. (2016). The process and context of violence risk assessment. In J. P. Singh, S. Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 53–75). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199386291.003.0005>

-
- Eldridge, H. (2019). Juror comprehension of forensic expert testimony: A literature review and gap analysis. *Forensic Science International: Synergy*, 1(1), 24–34. <https://doi.org/10.1016/j.fsisyn.2019.03.001>
- Ellison v. Commonwealth, 639 S.E.2d 209 (Supreme Court of VA 2007).
- Ellman, I. M., & Ellman, T. (2015). “Frightening and high”: The Supreme Court’s crucial mistake about sex crime statistics. *Constitutional Commentary*, 30(3), 495–508. <https://doi.org/10.2139/ssrn.2616429>
- Elwood, R. W. (2019). Agreement between courts and SVP evaluators in the state of Wisconsin. *Criminal Justice and Behavior*, 46(6), 853–865. <https://doi.org/10.1177/0093854819839746>
- Elwood, R. W., Doren, D. M., & Thornton, D. (2010). Diagnostic and risk profiles of men detained under Wisconsin’s Sexually Violent Person law. *International Journal of Offender Therapy and Comparative Criminology*, 54(2), 187–196. <https://doi.org/10.1177/0306624X08327305>
- Epstein, S. (1993). Implications of cognitive-experiential self-theory for personality and developmental psychology. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), *Studying lives through time: Personality and development*. (pp. 399–438). American Psychological Association. <https://doi.org/10.1037/10127-033>
- Evans, A., Kim, C., Laraia, N., Lutz, Z., Osborne, A., Parchuke, E., Williams, L., Zook, A., & Quarino, L. (2019). Toward a more effective use and understanding of forensic evidence in courts of law: Developing strategies for the scientific education of legal practitioners. *Widener Law Review*, 25(1), 1–22.
- Everson, M. D., & Sandoval, J. M. (2011). Forensic child sexual abuse evaluations: Assessing subjectivity and bias in professional judgements. *Child Abuse & Neglect*, 35(4), 287–298. <https://doi.org/10.1016/j.chab.2011.01.001>
- Ewing, C. P. (1982). “Dr. Death” and the case for an ethical ban on psychiatric and psychological predictions of dangerousness in capital sentencing proceedings. *American Journal of Law & Medicine*, 8(4), 407–428.
- Ewing, C. P. (1991). Preventive detention and execution: The constitutionality of punishing future crimes. *Law and Human Behavior*, 15(2), 139–163. <https://doi.org/10.1007/BF01044615>
- Expertisecentrum Forensische Psychiatrie. (n.d.). About EFP. efp.nl/en/about-efp
- Fanniff, A. M., Otto, R. K., & Petrala, J. (2010). Competence to proceed in SVP commitment hearings: Irrelevant or fundamental due process right? *Behavioral Sciences & the Law*, 28(5), 647–670. <https://doi.org/10.1002/bls.957>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science*, 241(4861), 31–35. <https://doi.org/10.1126/science.3291114>
- 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 24,827 people: Systematic review and meta-analysis. *BMJ*, 345(e4692), 1–12. <https://doi.org/10.1136/bmj.e4692>
- Fedoroff, J. P., & Moran, B. (1997). Myths and misconceptions about sex offenders. *The Canadian Journal of Human Sexuality*, 6(4), 263–276.
- Felthous, A. R., & Ko, J. (2018). Sexually violent predator law in the United States. *East Asian Archives of Psychiatry*, 28(4), 159–173. <https://doi.org/10.12809/eaap1835>

- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26(2), 105–109. <https://doi.org/10.3969/j.issn.1002-0829.2014.02.009>
- Ferguson, K., & Ireland, C. A. (2006). Attitudes towards sex offenders and the influence of offence type: A comparison of staff working in a forensic setting and students. *The British Journal of Forensic Practice*, 8(2), 10–19.
- Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge University Press.
- Fluent, T., & Guyer, M. (2008). Violent Sexual Predator (SVP) civil hearings and the application of collateral estoppel: SVP civil hearings and double jeopardy. *Journal of the American Academy of Psychiatry and the Law*, 36(1), 136–138.
- Ford, M. C. (1986). The role of extralegal factors in jury verdicts. *The Justice System Journal*, 11(1), 16–39.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.socjec.2010.10.008>
- Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology*, 36(4), 384–409. <https://doi.org/10.1006/jesp.1999.1409>
- Galinsky, A. D., Moskowitz, G. B., & Skurnik, I. (2000). Counterfactuals as self-generated primes: The effect of prior counterfactual activation on person perception judgments. *Social Cognition*, 18(3), 252–280. <https://doi.org/10.1521/soco.2000.18.3.252>
- Gallagher, M. P. (2009, November 2). Sex offenders sue over state refusal to allow setup of legal-help group. *New Jersey Law Journal*, 198(5), 468.
- Gardner, B. O., Kelley, S., Murrie, D. C., & Dror, I. E. (2019). What do forensic analysts consider relevant to their decision making? *Science & Justice*, 59(5), 516–523. <https://doi.org/10.1016/j.scijus.2019.04.005>
- Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54(5), 733–740.
- Glancy, G. D., Ash, P., Bath, E. P., Buchanan, A., Fedoroff, P., Frierson, R. L., Harris, V. L., Hatters Friedman, S. J., Hauser, M. J., Knoll, J., Norko, M., Pinals, D. A., Price, M., Recupero, P., Scott, C. L., & Zonana, H. V. (2015). AAPL practice guideline for the forensic assessment. *The Journal of the American Academy of Psychiatry and the Law*, 32(2 Supplement), S3–S53.
- Goldyne, A. J. (2007). Minimizing the influence of unconscious bias in evaluations: A practical guide. *The Journal of the American Academy of Psychiatry and the Law*, 35(1), 60–66.
- Gookin, K. (2007). *Comparison of state laws authorizing involuntary commitment of sexually violent predators: 2006 Update, revised* (No. 07-08-1101). Washington Institute for Public Policy. https://www.wsipp.wa.gov/ReportFile/989/Wsipp_Comparison-of-State-Laws-Authorizing-Involuntary-Commitment-of-Sexually-Violent-Predators-2006-Update-Revised_Full-Report.pdf
- Gordon, S. (2016). Crossing the line: Daubert, dual roles, and the admissibility of forensic mental health testimony. *Cardozo Law Review*, 37(4), 1345–1400.
- Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, 52(1), 178–200. <https://doi.org/10.1177/0011128705281748>
- Gowensmith, W. N., Bryant, A. E., & Vitacco, M. J. (2014). Decision-making in post-acquittal hospital release: How do forensic evaluators make their decisions? *Behavioral Sciences & the Law*, 32(5), 596–607. <https://doi.org/10.1002/bls.2135>

-
- Gowensmith, W. N., & McCallum, K. E. (2019). Mirror, mirror on the wall, who's the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations. *South African Journal of Psychology*, 49(2), 165–176. <https://doi.org/10.1177/0081246319835117>
- Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2012). Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior*, 36(2), 130–139. <https://doi.org/10.1037/h0093958>
- Gowensmith, W. N., Murrie, D. C., Boccaccini, M. T., & McNichols, B. J. (2017). Field reliability influences field validity: Risk assessments of individuals found not guilty by reason of insanity. *Psychological Assessment*, 29(6), 786–794. <https://doi.org/10.1037/pas0000376>
- Grabe, M. E., Zhou, S., & Barnett, B. (2001). Explicating sensationalism in television news: Content and the bells and whistles of form. *Journal of Broadcasting & Electronic Media*, 45(4), 635–655. https://doi.org/10.1207/s15506878jobem4504_6
- Graber, M. L., Kissam, S., Payne, V. L., Meyer, A. N. D., Sorensen, A., Lenfestey, N., Tant, E., Henriksen, K., LaBresh, K., & Singh, H. (2012). Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Quality & Safety*, 21(7), 535–557. <https://doi.org/10.1136/bmjqqs-2011-000149>
- Greenberg, S. A., & Shuman, D. W. (2007). When worlds collide: Therapeutic and forensic roles. *Professional Psychology: Research and Practice*, 38(2), 129–132. <https://doi.org/10.1037/0735-7028.38.2.129>
- Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology*, 2(1), 102–115.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293–323.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10.1037//1040-3590.12.1.19>
- 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(2), 143–152. <https://doi.org/10.1037/tpso000114>
- Guay, J. P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior*, 45(1), 82–100.
- Gunnell, J. J., & Ceci, S. J. (2010). When emotionality trumps reason: A study of individual processing style and juror bias. *Behavioral Sciences & the Law*, 28(6), 850–877. <https://doi.org/10.1002/bls.939>
- Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2001). Inside the judicial mind. *Cornell Law Review*, 86(4), 777–830. <https://doi.org/10.2139/ssrn.257634>
- Guy, L. S., Douglas, K. S., & Hart, S. D. (2015). Risk assessment and communication. In B. L. Cutler & P. A. Zapf (Eds.), *APA handbook of forensic psychology, Vol. 1: Individual and situational influences in criminal and civil contexts*. (Vol. 1, pp. 35–86). American Psychological Association. <https://doi.org/10.1037/14461-003>
- Guy, L. S., & Edens, J. F. (2003). Juror decision-making in a mock sexually violent predator trial: Gender differences in the impact of divergent types of expert testimony. *Behavioral Sciences & the Law*, 21(2), 215–237. <https://doi.org/10.1002/bls.529>
- Guy, L. S., Kusaj, C., Packer, I. K., & Douglas, K. S. (2015). Influence of the HCR-20, LS/CMI, and PCL-R on decisions about parole suitability among lifers. *Law and Human Behavior*, 39(3), 232–243. <https://doi.org/10.1037/lhb0000111>

- Guy, L. S., Nelson, R. J., Fusco-Morin, S. L., & Vincent, G. M. (2014). What do juvenile probation officers think of using the SAVRY and YLS/CMI for case management, and do they use the instruments properly? *International Journal of Forensic Mental Health*, 13(3), 227–241. <https://doi.org/10.1080/14999013.2014.939789>
- Hamilton, M. (2020). Judicial gatekeeping on scientific validity with risk assessment tools. *Behavioral Sciences & the Law*, 38(3), 226–245. <https://doi.org/10.1002/bls.2456>
- Hamnett, H. J., & Dror, I. E. (2020). The effect of contextual information on decision-making in forensic toxicology. *Forensic Science International: Synergy*, 2(1), 339–348. <https://doi.org/10.1016/j.fsisyn.2020.06.003>
- Hannah-Moffat, K., Maurutto, P., & Turnbull, S. (2009). Negotiated risk: Actuarial illusions and discretion in probation. *Canadian Journal of Law & Society/Revue Canadienne Droit et Societe (University of Toronto Press)*, 24(3), 391–409. <https://doi.org/10.1017/S0829320100010097>
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348–362.
- Hanson, R. K., Harris, A. J. R., Helmus, L., & Thornton, D. (2014). High-risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence*, 29(15), 2792–2813. <https://doi.org/10.1177/0886260514526062>
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Public Safety Canada. <https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf>
- Hanson, R. K., Helmus, L.-M., & Harris, A. J. R. (2015). Assessing the risk and needs of supervised sexual offenders: A prospective study using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior*, 42(12), 1205–1224. <https://doi.org/10.1177/0093854815602094>
- Hanson, R. K., Morton-Bourgon, K., Canada, & Public Safety and Emergency Preparedness Canada. (2004). *Predictors of sexual recidivism: An updated meta-analysis*. Public Works and Government Services Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73(6), 1154–1163. <https://doi.org/10.1037/0022-006X.73.6.1154>
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1–21. <https://doi.org/10.1037/a0014421>
- Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders* (No. 1999-02). Solicitor General Canada.
- Hanson, R. K., Thornton, D., Helmus, L.-M., & Babchishin, K. M. (2016). What sexual recidivism rates are associated with Static-99R and Static-2002R scores? *Sexual Abuse: A Journal of Research and Treatment*, 28(3), 218–252. <https://doi.org/10.1177/1079063215574710>
- Hare, R. D. (2003). *Manual for the revised Psychopathy Checklist* (2nd ed.). Multi-Health Systems.
- Harper, C. A., & Bartels, R. M. (2017). The influence of implicit theories and offender characteristics on judgements of sexual offenders: A moderated mediation analysis. *Journal of Sexual Aggression*, 23(2), 139–150. <https://doi.org/10.1080/13552600.2016.1250963>
- Harper, C. A., & Bartels, R. M. (2018). Implicit theories and offender representativeness in judgments about sexual crime. *Sexual Abuse: A Journal of Research and Treatment*, 30(3), 276–295. <https://doi.org/10.1177/1079063216658019>

-
- Harper, C. A., & Hogue, T. E. (2015). Measuring public perceptions of sex offenders: Reimagining the Community Attitudes Toward Sex Offenders (CATSO) scale. *Psychology, Crime & Law*, 21(5), 452–470. <https://doi.org/10.1080/1068316X.2014.989170>
- Harris, A. J., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question* (No. 2004–03; p. 29). Public Safety and Emergency Preparedness Canada.
- Harris, A. J., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *STATIC-99 Coding Rules Revised—2003*. http://www.static99.org/pdfdocs/static-99-coding-rules_e.pdf
- Harris, A. J., & Socia, K. M. (2016). What's in a name? Evaluating the effects of the "sex offender" label on public opinions and beliefs. *Sexual Abuse: A Journal of Research and Treatment*, 28(7), 660–678. <https://doi.org/10.1177/1079063214564391>
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20(4), 315–335.
- Harris, P. M. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgment. *Federal Probation*, 70(2), 8–14.
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version*. Multi-Health Systems.
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments. *British Journal of Psychiatry*, 190(S49), s60–s65. <https://doi.org/10.1192/bjp.190.5.s60>
- Harte, J. M., & Breukink, M. D. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie [Objectivity or apparent security? Quality, possibilities and limitations of instruments for risk assessment]. *Tijdschrift Voor Criminologie*, 52(1), 52–72.
- Hartvig, P., Ostberg, B., Alfarnes, S., Moger, T. A., Skjønberg, M., & Bjørkly, S. (2007). *Violent Risk Screen 10 (V-RISK-10)*. www.forensic-psychiatry.no www.forensic-psychiatry.no
- Harvard Law School Access to Justice Lab. (2016, May 25). *Pretrial release*. The Access to Justice Lab. <https://a2jlab.org/pretrial-release/>
- Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In *The handbook of evolutionary psychology* (pp. 1–20). American Cancer Society. <https://doi.org/10.1002/9781119125563.evpsych241>
- Haugtvedt, C. P., Petty, R. E., & Cacioppo, J. T. (1992). Need for cognition and advertising: Understanding the role of personality variables in consumer behavior. *Journal of Consumer Psychology*, 1(3), 239–260. [https://doi.org/10.1016/S1057-7408\(08\)80038-1](https://doi.org/10.1016/S1057-7408(08)80038-1)
- Hawes, S. W., Boccaccini, M. T., & Murrie, D. C. (2013). Psychopathy and the combination of psychopathy and sexual deviance as predictors of sexual recidivism: Meta-analytic findings using the Psychopathy Checklist—Revised. *Psychological Assessment*, 25(1), 233–243. <https://doi.org/10.1037/a0030391>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). Guilford Press.
- Heide, K. M. (2013). *Understanding parricide: When sons and daughters kill parents*. Oxford University Press.
- Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, 21(4), 347–359.
- Heilbrun, K., Goldstein, A. M., & Grisso, T. (2009). *Foundations of forensic mental health assessment*. Oxford University Press.

- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment*, 24(1), 64–101. <https://doi.org/10.1177/1079063211409951>
- Hildebrand, M., de Ruiter, C., & de Vogel, V. (2004). Psychopathy and sexual deviance in treated rapists: Association with sexual and nonsexual recidivism. *Sexual Abuse: A Journal of Research and Treatment*, 16(1), 1–24.
- Hill, D., & Demetriooff, S. (2019). Clinical-forensic psychology in Canada: A survey of practitioner characteristics, attitudes, and psychological assessment practices. *Canadian Psychology/Psychologie Canadienne*, 60(1), 55–63. <https://doi.org/10.1037/cap0000152>
- Hilton, N. Z., Harris, G. T., Rawson, K., & Beach, C. A. (2005). Communicating violence risk information to forensic decision makers. *Criminal Justice and Behavior*, 32(1), 97–116. <https://doi.org/10.1177/0093854804270630>
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist*, 34(3), 400–409. <https://doi.org/10.1177/0011000005285877>
- Hilton, N. Z., Scurich, N., & Helmus, L.-M. (2015). Communicating the risk of violent and offending behavior: Review and introduction to this special issue. *Behavioral Sciences & the Law*, 33(1), 1–18. <https://doi.org/10.1002/bls.2160>
- Ho, R., ForsterLee, L., ForsterLee, R., & Crofts, N. (2002). Justice versus vengeance: Motives underlying punitive judgements. *Personality and Individual Differences*, 33(3), 365–377. [https://doi.org/10.1016/S0191-8869\(01\)00161-1](https://doi.org/10.1016/S0191-8869(01)00161-1)
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1), 1–55. [https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/10.1016/0010-0285(92)90002-J)
- Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior*, 29(4), 380–396.
- Homant, R. J., & Kennedy, D. B. (1986). Judgment of legal insanity as a function of attitude toward the insanity defense. *International Journal of Law and Psychiatry*, 8(1), 67–81.
- Homant, R. J., & Kennedy, D. B. (1987a). Subjective factors in clinicians' judgments of insanity: Comparison of a hypothetical case and an actual case. *Professional Psychology: Research and Practice*, 18(5), 439–446.
- Homant, R. J., & Kennedy, D. B. (1987b). Subjective factors in the judgment of insanity. *Criminal Justice and Behavior*, 14(1), 38–61.
- Hope, L., Memon, A., & McGeorge, P. (2004). Understanding pretrial publicity: Predecisional distortion of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, 10(2), 111–119. <https://doi.org/10.1037/1076-898X.10.2.111>
- Huang, C.-Y., & Bull, R. (2020). Applying Hierarchy of Expert Performance (HEP) to investigative interview evaluation: Strengths, challenges and future directions. *Psychiatry, Psychology and Law*. Advance online publication. <https://doi.org/10.1080/13218719.2020.1770634>
- Huls, L., Nentjes, L., Rinne, T., & Verschueren, B. (2018). Gebruikte versus werkelijke voorspellers van seksuele recidive: Invloed van morele verwerpelijkheid op oordeel pro Justitia rapporteurs [How reliable are the predictors of sexual recidivism? Moral considerations can color the judgement of pro Justitia reporters]. *Tijdschrift Voor Psychiatrie*, 60(2), 78–86.
- Hurducas, C. C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of surveys. *International Journal of Forensic Mental Health*, 13(3), 181–192. <https://doi.org/10.1080/14999013.2014.942923>

-
- Huss, M. T., & Zeiss, R. A. (2004). Clinical assessment of violence from inpatient records: A comparison of individual and aggregate decision making across risk strategies. *International Journal of Forensic Mental Health*, 3(2), 139–147. <https://doi.org/10.1080/14999013.2004.10471203>
- Iman, R. L. (1974). A power study of a rank transform for the two-way classification model when interaction may be present. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 2(2), 227–239. <https://doi.org/10.2307/3314695>
- Imhoff, R. (2015). Punitive attitudes against pedophiles or persons with sexual interest in children: Does the label matter? *Archives of Sexual Behavior*, 44(1), 35–44. <https://doi.org/10.1007/s10508-014-0439-3>
- Imhoff, R., & Jahnke, S. (2018). Determinants of punitive attitudes toward people with pedophilia: Dissecting effects of the label and intentionality ascriptions. *Archives of Sexual Behavior*, 47(2), 353–361. <https://doi.org/10.1007/s10508-017-1048-8>
- Ismail, G., & Looman, J. (2018). Field inter-rater reliability of the Psychopathy Checklist—Revised. *International Journal of Offender Therapy and Comparative Criminology*, 62(2), 468–481. <https://doi.org/10.1177/0306624X16652452>
- Jaccard, J. J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Sage.
- Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sex Abuse*, 19(4), 425–448. <https://doi.org/10.1007/s11194-007-9062-3>
- Jackson, R. L., & Richards, H. J. (2007). Diagnostic and risk profiles among civilly committed sex offenders in Washington State. *International Journal of Offender Therapy and Comparative Criminology*, 51(3), 313–323. <https://doi.org/10.1177/0306624X06292874>
- Jackson, R. L., Rogers, R., & Shuman, D. W. (2004). The adequacy and accuracy of sexually violent predator evaluations: Contextualized risk assessment in clinical practice. *International Journal of Forensic Mental Health*, 3(2), 115–129. <https://doi.org/10.1080/14999013.2004.10471201>
- Janus, E. S. (2013). Preventive detention of sex offenders: The American experience versus international human rights norms. *Behavioral Sciences & the Law*, 31(3), 328–343. <https://doi.org/10.1002/bls.2059>
- Janus, E. S., & Meehl, P. E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy, and Law*, 3(1), 33–64.
- Janus, E. S., & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility and accountability. *American Criminal Law Review*, 40(4), 1443–1499.
- Janus, E. S., & Prentky, R. A. (2008). Sexual predator laws: A two-decade retrospective. *Federal Sentencing Reporter*, 21(2), 90–97. <https://doi.org/10.1525/fsr.2008.21.2.90>
- Jeanguenat, A. M., Budowle, B., & Dror, I. E. (2017). Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. *Science & Justice*, 57(6), 415–420. <https://doi.org/10.1016/j.scijus.2017.07.005>
- Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, 28(8), 474–481. <https://doi.org/10.1016/j.tree.2013.05.014>
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3), 582–591.
- Kamorowski, J., Ask, K., Schreuder, M., Jelíćic, M., & de Ruiter, C. (2021). ‘He seems odd’: The effects of risk-irrelevant information and actuarial risk estimates on mock jurors’ perceptions of sexual recidivism risk. *Psychology, Crime & Law*. Advance online publication. <https://doi.org/10.1080/1068316X.2021.1909016>

- Kamorowski, J., de Ruiter, C., Schreuder, M., Ask, K., & Jelćic, M. (2021). Forensic mental health practitioners' use of structured risk assessment instruments, views about bias in risk evaluations, and strategies to counteract it. *International Journal of Forensic Mental Health*. Advance online publication. <https://doi.org/10.1080/14999013.2021.1895377>
- Kansas v. Crane, 534 U.S. 407 (2002).
- Kansas v. Hendricks, 521 U.S. 346 (1997).
- Karp, J. (2021, January 24). *Studying justice or hurting it: The fight over A2J research*. Law360 Access to Justice. <https://www.law360.com/articles/1347474/studying-justice-or-hurting-it-the-fight-over-a2j-research>
- 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>
- Katz-Schiavone, S. M., Levenson, J. S., & Ackerman, A. R. (2008). Myths and facts about sexual violence: Public perceptions and implications for prevention. *Journal of Criminal Justice and Popular Culture*, 15(3), 291–311.
- Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2020). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse*, 32(1), 3–29. <https://doi.org/10.1177/1079063218800474>
- King, L. L. (2019). Perceptions about sexual offenses: Misconceptions, punitiveness, and public sentiment. *Criminal Justice Policy Review*, 30(2), 254–273. <https://doi.org/10.1177/0887403416660150>
- King, L. L., & Roberts, J. J. (2017). The complexity of public attitudes toward sex crimes. *Victims & Offenders*, 12(1), 71–89. <https://doi.org/10.1080/15564886.2015.1005266>
- Knighton, J. C., Murrie, D. C., Boccaccini, M. T., & Turner, D. B. (2014). How likely is “likely to reoffend” in sex offender civil commitment trials? *Law and Human Behavior*, 38(3), 293–304. <https://doi.org/10.1037/lhb0000079>
- Koeppl, B. (2019, June 4). Modern-day gulag in the Golden State. *The Washington Spectator*. <https://washingtonspectator.org/koeppl-gulags/>
- Korteling, J. E., Brouwer, A.-M., & Toet, A. (2018). A neural network framework for cognitive bias. *Frontiers in Psychology*, 9(1), 1561–1573. <https://doi.org/10.3389/fpsyg.2018.01561>
- Korva, N., Porter, S., O'Connor, B. P., Shaw, J., & Brinke, L. (2013). Dangerous decisions: Influence of juror attitudes and defendant appearance on legal decision-making. *Psychiatry, Psychology & Law*, 20(3), 384–398. <https://doi.org/10.1080/13218719.2012.692931>
- Kötter, S., von Franqué, F., Bolzmacher, M., Eucker, S., Holzinger, B., & Müller-Isberner, R. (2014). The HCR-20^{v3} in Germany. *International Journal of Forensic Mental Health*, 13(2), 122–129. <https://doi.org/10.1080/14999013.2014.911784>
- Kovera, M. B. (2002). The effects of general pretrial publicity on juror decisions: An examination of moderators and mediating mechanisms. *Law and Human Behavior*, 26(1), 43–72.
- Kovera, M. B., & Levett, L. M. (2015). Jury decision making. In B. L. Cutler & P. A. Zapf (Eds.), *APA handbook of forensic psychology, Vol. 2: Criminal investigation, adjudication, and sentencing outcomes*. American Psychological Association. <https://doi.org/10.1037/14462-000>
- Krahé, B., Temkin, J., & Bieneck, S. (2007). Schema-driven information processing in judgements about rape. *Applied Cognitive Psychology*, 21(5), 601–619. <https://doi.org/10.1002/acp.1297>

-
- Krauss, D. A., Cook, G. I., & Klapatch, L. (2018). Risk assessment communication difficulties: An empirical examination of the effects of categorical versus probabilistic risk communication in sexually violent predator decisions. *Behavioral Sciences & the Law*, 36(5), 532–553. <https://doi.org/10.1002/bls.2379>
- Krauss, D. A., & Lee, D. H. (2003). Deliberating on dangerousness and death: Jurors' ability to differentiate between expert actuarial and clinical predictions of dangerousness. *International Journal of Law and Psychiatry*, 26(2), 113–137. [https://doi.org/10.1016/S0160-2527\(02\)00211-X](https://doi.org/10.1016/S0160-2527(02)00211-X)
- Krauss, D. A., Lieberman, J. D., & Olson, J. (2004). The effects of rational and experiential information processing of expert testimony in death penalty cases. *Behavioral Sciences & the Law*, 22(6), 801–822. <https://doi.org/10.1002/bls.621>
- Krauss, D. A., McCabe, J. G., & Lieberman, J. D. (2012). Dangerously misunderstood: Representative jurors' reactions to expert testimony on future dangerousness in a sexually violent predator trial. *Psychology, Public Policy, and Law*, 18(1), 18–49. <https://doi.org/10.1037/a0024550>
- Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, 7(2), 267–310. <https://doi.org/10.1037/1076-8971.7.2.267>
- Krauss, D. A., & Scurich, N. (2013). Risk assessment in the law: Legal admissibility, scientific validity, and some disparities between research and practice. *Behavioral Sciences & the Law*, 31(2), 215–229. <https://doi.org/10.1002/bls.2065>
- Krauss, D. A., & Scurich, N. (2014). The impact of case factors on jurors' decisions in a sexual violent predator hearing. *Psychology, Public Policy, and Law*, 20(2), 135–145. <https://doi.org/10.1037/law0000007>
- Krivacska, J. (2011). Reliability of sex-offense risk assessments at civil commitment: Threshold levels in New Jersey SVP proceedings. *Open Access Journal of Forensic Psychology*, 3(1), 107–145.
- Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *The Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): User manual*. ProActive ReSolutions, Inc.
- Kukucka, J. (2014). The journey or the destination? Disentangling process and outcome in forensic identification. *Forensic Science Policy & Management*, 5(3–4), 112–114.
- Kukucka, J. (2020). People who live in ivory towers shouldn't throw stones: A refutation of Curley et al. *Forensic Science International: Synergy*, 2(1), 110–113. <https://doi.org/10.1016/j.fsisyn.2020.03.001>
- Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, 38(3), 256–270. <https://doi.org/10.1037/lhb0000066>
- 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>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498.
- Kwartner, P. P., Lyons, P. M., & Boccaccini, M. T. (2006). Judges' risk communication preferences in risk for future violence cases. *International Journal of Forensic Mental Health*, 5(2), 185–194. <https://doi.org/10.1080/14999013.2006.10471242>
- LaFree, G., Reskin, B. F., & Visher, C. A. (1985). Jurors' responses to victims' behavior and legal issues in sexual assault trials. *Social Problems*, 32(4), 389–407.
- Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 34(5), 491–498. <https://doi.org/10.1037/0735-7028.34.5.491>

- Landy, D., & Aronson, E. (1969). The influence of the character of the criminal and his victim on the decisions of simulated jurors. *Journal of Experimental Social Psychology*, 5(2), 141–152. [https://doi.org/10.1016/0022-1031\(69\)90043-2](https://doi.org/10.1016/0022-1031(69)90043-2)
- Langan, P. A., Schmitt, E. L., & Durose, M. R. (2003). *Recidivism of sex offenders released from prison in 1994* (NCJ 198281). U.S. Department of Justice, Bureau of Justice Statistics. <https://www.bjs.gov/content/pub/pdf/rsorp94.pdf>
- Lanterman, J. L., Boyle, D. J., & Ragusa-Salerno, L. M. (2014). Sex offender risk assessment, sources of variation, and the implications of misuse. *Criminal Justice and Behavior*, 41(7), 822–843. <https://doi.org/10.1177/0093854813515237>
- Latessa, E. J., & Lovins, B. (2010). The role of offender risk assessment: A policy maker guide. *Victims & Offenders*, 5(3), 203–219. <https://doi.org/10.1080/15564886.2010.485900>
- Leij, J. B. J. van der, Jackson, J. L., Malsch, M., & Nijboer, J. F. (2001). Residential mental health assessment within Dutch criminal cases: A discussion. *Behavioral Sciences & the Law*, 19(5–6), 691–702. <https://doi.org/10.1002/bls.465>
- Levenson, J. S. (2004). Reliability of sexually violent predator civil commitment criteria in Florida. *Law and Human Behavior*, 28(4), 357–368. <https://doi.org/10.1023/B:LAHU.0000039330.22347.ad>
- Levenson, J. S., Brannon, Y. N., Fortney, T., & Baker, J. (2007). Public perceptions about sex offenders and community protection policies. *Analyses of Social Issues and Public Policy*, 7(1), 137–161. <https://doi.org/10.1111/j.1530-2415.2007.00119.x>
- Levenson, J. S., & Cotter, L. P. (2005). The effect of Megan's Law on sex offender reintegration. *Journal of Contemporary Criminal Justice*, 21(1), 49–66.
- Levinson, R. M., & Ramsay, G. (1979). Dangerousness, stress, and mental health evaluations. *Journal of Health and Social Behavior*, 20(2), 178–187. <https://doi.org/10.2307/2136438>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Lieberman, J. D. (2002). Head over the heart or heart over the head? Cognitive experiential self-theory and extralegal heuristics in juror decision making. *Journal of Applied Social Psychology*, 32(12), 2526–2553. <https://doi.org/10.1111/j.1559-1816.2002.tb02755.x>
- Lieberman, J. D., & Krauss, D. A. (2009). The effects of labelling, expert testimony, and information processing mode on juror decisions in SVP civil commitment trials. *Journal of Investigative Psychology & Offender Profiling*, 6(1), 25–41. <https://doi.org/10.1002/jip.91>
- Lieberman, J. D., Krauss, D. A., Kyger, M., & Lehoux, M. (2007). Determining dangerousness in sexually violent predator evaluations: Cognitive–experiential self-theory and juror judgments of expert testimony. *Behavioral Sciences & the Law*, 25(4), 507–526. <https://doi.org/10.1002/bls.771>
- Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, 4(4), 390–398. <https://doi.org/10.1111/j.1745-6924.2009.01144.x>
- Liptak, A. (2017, March 6). Did the Supreme Court base a ruling on a myth? *The New York Times*. <https://www.nytimes.com/2017/03/06/us/politics/supreme-court-repeat-sex-offenders.html>
- Liu, J. Z., & Li, X. (2019). Legal techniques for rationalizing biased judicial decisions: Evidence from experiments with real judges. *Journal of Empirical Legal Studies*, 16(3), 630–670. <https://doi.org/10.1111/jels.12229>
- Lord, C. G., Lepper, M. R., & Preston, E. (1985). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243. <https://doi.org/10.1037/0022-3514.47.6.1231>

-
- Lu, Y., Freeman, N. J., & Sandler, J. C. (2015). Predictors of the sex offender civil commitment trial outcomes in New York State. *Law and Human Behavior*, 39(5), 514–524. <https://doi.org/10.1037/lhb0000143>
- Lynch, M., & Haney, C. (2009). Capital jury deliberation: Effects on death sentencing, comprehension, and discrimination. *Law and Human Behavior*, 33(6), 481–496. <https://doi.org/10.1007/s10979-008-9168-2>
- Lynett, E., & Rogers, R. (2000). Emotions overriding forensic opinions? The potentially biasing effects of victim statements. *The Journal of Psychiatry & Law*, 28(4), 449–457. <https://doi.org/10.1177/009318530002800402>
- MacLean, N., Neal, T. M. S., Morgan, R. D., & Murrie, D. C. (2019). Forensic clinicians' understanding of bias. *Psychology, Public Policy, and Law*, 25(4), 323–330. <https://doi.org/10.1037/law0000212>
- Malinen, S., Willis, G. M., & Johnston, L. (2014). Might informative media reporting of sexual offending influence community members' attitudes towards sex offenders? *Psychology, Crime & Law*, 20(6), 535–552. <https://doi.org/10.1080/1068316X.2013.793770>
- Mancini, C., & Mears, D. P. (2010). To execute or not to execute? Examining public support for capital punishment of sex offenders. *Journal of Criminal Justice*, 38(5), 959–968. <https://doi.org/10.1016/j.jcrimjus.2010.06.013>
- Mancini, C., Shields, R. T., Mears, D. P., & Beaver, K. M. (2010). Sex offender residence restriction laws: Parental perceptions and public policy. *Journal of Criminal Justice*, 38(5), 1022–1030. <https://doi.org/10.1016/j.jcrimjus.2010.07.004>
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, 22(2), 191–217. <https://doi.org/10.1177/1079063210366039>
- 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(4), 541–547. <https://doi.org/10.1002/acp.1719>
- Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology*, 59(1), 27–37.
- McCabe, J. G., Krauss, D. A., & Lieberman, J. D. (2010). Reality check: A comparison of college students and a community sample of mock jurors in a simulated sexual violent predator civil commitment. *Behavioral Sciences & the Law*, 28(6), 730–750. <https://doi.org/10.1002/bls.902>
- McCallum, K. E., Boccaccini, M. T., & Bryson, C. N. (2017). The influence of risk assessment instrument scores on evaluators' risk opinions and sexual offender containment recommendations. *Criminal Justice and Behavior*, 44(9), 1213–1235. <https://doi.org/10.1177/0093854817707232>
- McKune v. Lile, 536 U.S. 24 (2002). <https://supreme.justia.com/cases/federal/us/536/24/case.pdf>
- McLaughlin, J. L., & Kan, L. Y. (2014). Test usage in four common types of forensic mental health assessment. *Professional Psychology: Research and Practice*, 45(2), 128–135. <https://doi.org/10.1037/a0036318>
- McNeil, D. E., Lam, J. N., & Binder, R. L. (2000). Relevance of interrater agreement to violence risk assessment. *Journal of Consulting and Clinical Psychology*, 68(6), 1111–1115. <https://doi.org/10.1037/0022-006X.68.6.1111>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>

- Mears, D. P., Mancini, C., Gertz, M., & Bratton, J. (2008). Sex crimes, children, and pornography: Public views and public policy. *Crime & Delinquency*, 54(4), 532–559. <https://doi.org/10.1177/0011128707308160>
- Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook* (3rd ed.). Guilford Press.
- Melvin, K. B., Gramling, L. K., & Gardner, W. M. (1985). A scale to measure attitudes toward prisoners. *Criminal Justice and Behavior*, 12(2), 241–253. <https://doi.org/10.1177/0093854885012002006>
- Mesler, J. L., Anderson, G., & Calkins, C. (2016). Sex offender policy and prevention. In M. K. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (pp. 217–248). Springer International Publishing. https://doi.org/10.1007/978-3-319-29406-3_7
- Messina, E., Ferracuti, S., Nicolò, G., Ruggeri, M., Kooijmans, T., & Meynen, G. (2019). Forensic psychiatric evaluations of defendants: Italy and the Netherlands compared. *International Journal of Law and Psychiatry*, 66(1), Article 101473. <https://doi.org/10.1016/j.ijlp.2019.101473>
- Metzner, J. L., & Ash, P. (2010). Commentary: The mental status examination in the age of the Internet—Challenges and opportunities. *The Journal of the American Academy of Psychiatry and the Law*, 38(1), 27–31.
- Meyer, L. F., & Valençā, A. M. (2021). Factors related to bias in forensic psychiatric assessments in criminal matters: A systematic review. *International Journal of Law and Psychiatry*, 75(1), Article 101681. <https://doi.org/10.1016/j.ijlp.2021.101681>
- Michelin, R. L., & Snodgrass, S. R. (1980). Defendant characteristics and juridic decisions. *Journal of Research in Personality*, 14(3), 340–350. [https://doi.org/10.1016/0092-6566\(80\)90017-3](https://doi.org/10.1016/0092-6566(80)90017-3)
- Miller, A. K., Rufino, K. A., Boccaccini, M. T., Jackson, R. L., & Murrie, D. C. (2011). On individual differences in person perception: Raters' personality traits relate to their Psychopathy Checklist-Revised scoring tendencies. *Assessment*, 18(2), 253–260. <https://doi.org/10.1177/1073191111402460>
- Miller, H. A., Amenta, A. E., & Conroy, M. A. (2005). Sexually violent predator evaluations: Empirical evidence, strategies for professionals, and research directions. *Law and Human Behavior*, 29(1), 29–54. <https://doi.org/10.1007/s10979-005-1398-y>
- Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior*, 40(7), 716–736.
- Miller, S. L., & Brodsky, S. L. (2011). Risky business: Addressing the consequences of predicting violence. *The Journal of the American Academy of Psychiatry and the Law*, 39(3), 396–401. <https://doi.org/PMID: 21908757>
- Monahan, J. (1981). *The clinical prediction of violent behavior*. U. S. Department of Health and Human Services.
- Monahan, J. (2017). *Risk assessment in sentencing* (No. 2017–44; Public Law and Legal Theory Research Paper Series, p. 14). University of Virginia School of Law. <https://ssrn.com/abstract=3024016>
- Monahan, J., Metz, A. L., & Garrett, B. L. (2018). Judicial appraisals of risk assessment in sentencing. *Behavioral Sciences & the Law*, 36(5), 565–575. <https://doi.org/10.1002/bls.2380>
- Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, 12(1), 489–513. <https://doi.org/10.1146/annurev-clinpsy-021815-092945>
- Monahan, J., Steadman, H. J., Appelbaum, P. S., Grisso, T., Mulvey, E. P., Roth, L. H., Robbins, P. C., Banks, S., & Silver, E. (2006). The Classification of Violence Risk. *Behavioral Sciences and the Law*, 24(6), 721–730. <https://doi.org/10.4135/9781412959537.n35>

-
- Moran, G., & Cutler, B. L. (1991). The prejudicial impact of pretrial publicity. *Journal of Applied Social Psychology*, 21(5), 345–367. <https://doi.org/10.1111/j.1559-1816.1991.tb00524.x>
- Mossman, D. (2013). When forensic examiners disagree: Bias, or just inaccuracy? *Psychology, Public Policy, and Law*, 19(1), 40–55. <https://doi.org/10.1037/a0029242>
- Mulvey, E. P., & Lidz, C. W. (1984). Clinical considerations in the prediction of dangerousness in mental patients. *Clinical Psychology Review*, 4(4), 379–401. [https://doi.org/10.1016/0272-7358\(84\)90018-7](https://doi.org/10.1016/0272-7358(84)90018-7)
- Mulvey, E. P., & Lidz, C. W. (1985). Back to basics: A critical analysis of dangerousness research in a new legal environment. *Law and Human Behavior*, 9(2), 209–219. <https://doi.org/10.1007/BF01067052>
- Mumma, G. H., & Wilson, S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of Clinical Psychology*, 51(6), 841–853.
- Murray, J., Charles, K. E., Cooke, D. J., & Thomson, M. E. (2014). Investigating the influence of causal attributions on both the worksheet and checklist versions of the HCR-20. *International Journal of Forensic Mental Health*, 13(1), 8–17. <https://doi.org/10.1080/14999013.2014.890978>
- Murray, J., & Thomson, Dr. M. E. (2010). Clinical judgement in violence risk assessment. *Europe's Journal of Psychology*, 6(1), 128–149. <https://doi.org/10.5964/ejop.v6i1.175>
- Murrie, D. C., & Balusek, K. (2008). Forensic assessment of violence risk in adversarial proceedings: Pursuing objectivity and avoiding bias. *Journal of Forensic Psychology Practice*, 7(4), 141–153. https://doi.org/10.1300/J158v07n04_06
- Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science*, 11(1), 37–55. <https://doi.org/10.1146/annurev-lawsocsci-120814-121714>
- 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>
- Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, 32(4), 352–362. <https://doi.org/10.1007/s10979-007-9097-5>
- 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, and Law*, 15(1), 19–53. <https://doi.org/10.1037/a0014897>
- Murrie, D. C., & Warren, J. I. (2005). Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Professional Psychology: Research and Practice*, 36(5), 519–524. <https://doi.org/10.1037/0735-7028.36.5.519>
- Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice*, 54(3), 208–214. <https://doi.org/10.1016/j.scijus.2013.11.003>
- National District Attorneys Association. (2012). *Sex offender civil commitment laws*. <https://ndaa.org/wp-content/uploads/Sex-Offender-Civil-Commitment-April-2012.pdf>
- National Research Council (Ed.). (2009). *Strengthening forensic science in the United States: A path forward*. National Academies Press.

- Neal, T. M. S. (2016). Are forensic experts already biased before adversarial legal parties hire them? *PLOS ONE*, 11(4), e0154434. <https://doi.org/10.1371/journal.pone.0154434>
- 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. (2014a). 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>
- Neal, T. M. S., & Grisso, T. (2014b). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, 41(12), 1406–1421. <https://doi.org/10.1177/0093854814548449>
- Neal, T. M. S., Hight, M., Howatt, B. C., & Hamza, C. (2019). The cognitive and social psychological bases of bias in forensic mental health judgments. In M. K. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (Vol. 3, pp. 151–175). Springer International Publishing. https://doi.org/10.1007/978-3-319-75859-6_5
- Neal, T. M. S., & Saks, M. J. (2016). *The science of science offers solutions to cognitive bias in forensic psychology*. American Psychology-Law Society Annual Conference, Atlanta, GA.
- Nederlands Instituut voor Forensische Psychiatrie en Psychologie. (2018). *NIFP-richtlijn: Ambulant forensisch psychologisch onderzoek en rapportage in het strafrecht*. https://www.nifp.nl/binaries/richtlijn-ambulant-forensisch-onderzoek-en-rapportage-in-strafzaken_tcm106-351968.pdf
- Neimark, G., Hurford, M. O., & DiGiacomo, J. (2006). The Internet as collateral informant. *American Journal of Psychiatry*, 163(10), 1842.
- Nhan, J., Polzer, K., & Ferguson, J. (2012). "More dangerous than hitmen": Judicial perceptions of sexual offenders. *International Journal of Criminology & Sociological Theory*, 5(1), 823–836.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Nielsen, L. H., Mastrigt, S. van, Otto, R. K., Seewald, K., Ruiter, C. de, Rettenberger, M., Reeves, K. A., Rebocho, M. F., Pham, T. H., Mei Yee Ho, R., Grann, M., Godoy-Cervera, V., Folino, J. O., Doyle, M., Desmarais, S. L., Condemarin, C., Arbach-Lucioni, K., & Singh, J. P. (2015). Violence risk assessment practices in Denmark: A multidisciplinary national survey. *Scandinavian Journal of Forensic Science*, 21(2), 103–110. <https://doi.org/10.1515/sjfs-2015-0003>
- Nijman, H., Merckelbach, H., & Cima, M. (2009). Performance intelligence, sexual offending and psychopathy. *Journal of Sexual Aggression*, 15(3), 319–330. <https://doi.org/10.1080/13552600903195057>
- Nisbett, R. E., & Wilson, T. D. (1977a). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Nisbett, R. E., & Wilson, T. D. (1977b). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35(4), 250–256. <https://doi.org/10.1037/0022-3514.35.4.250>
- Norton, R., & Grant, T. (2008). Rape myth in true and false rape allegations. *Psychology, Crime & Law*, 14(4), 275–285. <https://doi.org/10.1080/10683160701770286>
- Osberg, T. M. (1987). The convergent and discriminant validity of the Need for Cognition scale. *Journal of Personality Assessment*, 51(3), 441–450. https://doi.org/10.1207/s15327752jpa5103_11

-
- Otto, A. L., Penrod, S. D., & Dexter, H. R. (1994). The biasing impact of pretrial publicity on juror judgments. *Law and Human Behavior*, 18(4), 453–469.
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184–188. <https://doi.org/10.1177/0963721414531598>
- Parker, G. (2016). Come see the bias inherent in the system! *The Journal of the American Academy of Psychiatry and the Law*, 44(4), 411–414.
- Parrott, C. T., Neal, T. M. S., Wilson, J. K., & Brodsky, S. L. (2015). Differences in expert witness knowledge: Do mock jurors notice and does it matter? *The Journal of the American Academy of Psychiatry and the Law*, 43(1), 69–81.
- Paternoster, R., & Deise, J. (2011). A heavy thumb on the scale: The effect of victim impact evidence on capital decision making. *Criminology*, 49(1), 129–161. <https://doi.org/10.1111/j.1745-9125.2010.00220.x>
- Pennycook, G., Trippas, D., Handley, S. J., & Thompson, V. A. (2014). Base rates: Both neglected and intuitive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 544–554. <https://doi.org/10.1037/a0034887>
- People v. Otto, 26 P.3d 1061 (CA Supreme Court 2001). <https://law.justia.com/cases/california/supreme-court/4th/26/200.html>
- Persson, M., Belfrage, H., Fredriksson, B., & Kristiansson, M. (2017). Violence during imprisonment, forensic psychiatric care, and probation: Correlations and predictive validity of the risk assessment instruments COVR, LSI-R, HCR-20^{v3}, and SAPROF. *International Journal of Forensic Mental Health*, 16(2), 117–129. <https://doi.org/10.1080/14999013.2016.1266420>
- Pham, T., Ducro, C., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., ... Singh, J. P. (2016). Enquête internationale sur les pratiques d'évaluation du risque de violence: Présentation des données belges [International survey of violence risk assessment practices: Presentation of Belgian data]. *Annales Médico-Psychologiques, Revue Psychiatrique*, 174(7), 539–543. <https://doi.org/10.1016/j.amp.2015.10.018>
- Phenix, A., Fernandez, Y., Harris, A. J. R., Helmus, M., Hanson, R. K., & Thornton, D. (2016). *Static-99R Coding Rules Revised – 2016* (No. 2017-R012). Public Safety Canada. http://www.static99.org/pdfdocs/Coding_manual_2016_v2.pdf
- Phenix, A., Helmus, L. M., & Hanson, R. K. (2016). *Static-99R & Static-2002R evaluators' workbook*. http://www.static99.org/pdfdocs/Evaluators_Workbook_2016-10-19.pdf
- Phenix, A., & Jackson, R. L. (2015). Evaluations for sexual offender civil commitment. In R. L. Jackson & R. Roesch (Eds.), *Learning forensic assessment: Research and practice* (2nd ed., pp. 162–201). Routledge.
- Pickett, J. T., Mancini, C., & Mears, D. P. (2013). Vulnerable victims, monstrous offenders, and unmanageable risk: Explaining public opinion on the social control of sex crime. *Criminology*, 51(3), 729–759. <https://doi.org/10.1111/1745-9125.12018>
- Pirelli, G., Hartigan, S., & Zapf, P. A. (2018). Using the Internet for collateral information in forensic mental health evaluations. *Behavioral Sciences & the Law*, 36(2), 157–169. <https://doi.org/10.1002/bls.2334>
- Pirelli, G., Otto, R. K., & Estoup, A. (2016). Using Internet and social media data as collateral sources of information in forensic evaluations. *Professional Psychology: Research and Practice*, 47(1), 12–17. <https://doi.org/10.1037/pro0000061>

- Porter, S., Brinke, L., & Wilson, K. (2009). Crime profiles and conditional release performance of psychopathic and non-psychopathic sexual offenders. *Legal and Criminological Psychology*, 14(1), 109–118. <https://doi.org/10.1348/135532508X284310>
- Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43(4), 565–578. <https://doi.org/10.1016/j.jesp.2006.05.011>
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369–381. <https://doi.org/10.1177/0146167202286008>
- Pyszczynski, T. A., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). Academic Press.
- Quigley-McBride, A. (2020). Practical solutions to forensic contextual bias. *Zeitschrift Für Psychologie*, 228(3), 162–174. <https://doi.org/10.1027/2151-2604/a000409>
- Quinn, J. F., Forsyth, C. J., & Mullen-Quinn, C. (2004). Societal reaction to sex offenders: A review of the origins and results of the myths surrounding their crimes and treatment amenability. *Deviant Behavior*, 25(3), 215–232. <https://doi.org/10.1080/01639620490431147>
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998a). Fifteen arguments against actuarial risk appraisal. In V. L. Quinsey, G. T. Harris, M. E. Rice, & C. A. Cormier, *Violent offenders: Appraising and managing risk*. (pp. 171–190). American Psychological Association. <https://doi.org/10.1037/10304-009>
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998b). *Violent offenders: Appraising and managing risk*. American Psychological Association.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006a). Clinical judgment. In V. L. Quinsey, G. T. Harris, M. E. Rice, & C. A. Cormier, *Violent offenders: Appraising and managing risk* (2nd ed.). (pp. 61–81). American Psychological Association. <https://doi.org/10.1037/11367-004>
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006b). Criticisms of actuarial risk assessment. In V. L. Quinsey, G. T. Harris, M. E. Rice, & C. A. Cormier, *Violent offenders: Appraising and managing risk* (2nd ed.). (pp. 197–223). American Psychological Association. <https://doi.org/10.1037/11367-009>
- 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>
- Recupero, P. R. (2008). Forensic evaluation of problematic Internet use. *The Journal of the American Academy of Psychiatry and the Law*, 36(4), 505–514.
- Recupero, P. R. (2010). The mental status examination in the age of the Internet. *The Journal of the American Academy of Psychiatry and the Law*, 38(1), 15–26.
- Reeves, R., Cevasco, R., & De Crisce, D. (2019). The pathway from prison to civil commitment to release in New Jersey. In W. T. O'Donohue & D. S. Bromberg (Eds.), *Sexually violent predators: A clinical science handbook* (pp. 357–366). Springer.
- Reskin, B. F., & Visher, C. A. (1986). The impacts of evidence and extralegal factors in jurors' decisions. *Law & Society Review*, 20(3), 423–438. <https://doi.org/10.2307/3053582>
- Rettenberger, M., Rice, M. E., Harris, G. T., & Eher, R. (2017). Actuarial risk assessment of sexual offenders: The psychometric properties of the Sex Offender Risk Appraisal Guide (SORAG). *Psychological Assessment*, 29(6), 624–638. <https://doi.org/10.1037/pas0000390>

-
- Reyzen, S. (2005). Construction of a new scale: The Reyzen likability scale. *Social Behavior and Personality*, 33(2), 201–208. <https://doi.org/10.2224/sbp.2005.33.2.201>
- Rhine, E. E., Mitchell, K. L., & Reitz, K. R. (2018). *Levers of change in parole release and revocation* (p. 36). Robina Institute of Criminal Law and Criminal Justice. https://robinainstitute.umn.edu/sites/robinainstitute.umn.edu/files/parole_landscape_report.pdf
- Richardson, D., & Campbell, J. L. (1982). Alcohol and rape: The effect of alcohol on attributions of blame for rape. *Personality and Social Psychology Bulletin*, 8(3), 468–476.
- Robinson, A. L., & Howarth, E. (2012). Judging risk: Key determinants in British domestic violence cases. *Journal of Interpersonal Violence*, 27(8), 1489–1518. <https://doi.org/10.1177/0886260511425792>
- Robinson, P. H., Jackowitz, S. E., & Bartels, D. M. (2012). Extralegal punishment factors: A study of forgiveness, hardship, good deeds, apology, remorse, and other such discretionary factors in assessing criminal punishment. *Vanderbilt Law Review*, 65(3), 735–826. <https://doi.org/10.2139/ssrn.1798725>
- Ruhland, E. L., Rhine, E. E., Robey, J. P., & Mitchell, K. L. (2016). *The continuing leverage of releasing authorities: Findings from a national survey*. Robina Institute of Criminal Law and Criminal Justice. https://robinainstitute.umn.edu/sites/robinainstitute.umn.edu/files/final_national_parole_survey_2017.pdf
- Ruva, C. L. (2018). From the headlines to the jury room: An examination of the impact of pretrial publicity on jurors and juries. In M. K. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (Vol. 3, pp. 1–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-75859-6_1
- Ruva, C. L., Dickman, M., & Mayes, J. L. (2014). Exposure to both positive and negative pretrial publicity reduces or eliminates mock-juror bias. *International Journal of Psychology and Behavioral Sciences*, 4(1), 30–40.
- Ruva, C. L., Guenther, C. C., & Yarbrough, A. (2011). Positive and negative pretrial publicity: The roles of impression formation, emotion, and predecisional distortion. *Criminal Justice and Behavior*, 38(5), 511–534. <https://doi.org/10.1177/0093854811400823>
- Ruva, C. L., & Hudak, E. M. (2013). Pretrial publicity and juror age affect mock-juror decision making. *Psychology, Crime & Law*, 19(2), 179–202. <https://doi.org/10.1080/1068316X.2011.616509>
- Ruva, C. L., McEvoy, C., & Bryant, J. B. (2007). Effects of pre-trial publicity and jury deliberation on juror bias and source memory errors. *Applied Cognitive Psychology*, 21(1), 45–67. <https://doi.org/10.1002/acp.1254>
- Rydberg, J., Dum, C. P., & Socia, K. M. (2018). Nobody gives a #%%!: A factorial survey examining the effect of criminological evidence on opposition to sex offender residence restrictions. *Journal of Experimental Criminology*, 14(4), 541–550. <https://doi.org/10.1007/s11292-018-9335-5>
- Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the Need for Cognition scale. *Perceptual and Motor Skills*, 74(2), 610.
- Sample, L. L., & Bray, T. M. (2003). Are sex offenders dangerous? *Criminology & Public Policy*, 3(1), 59–82. <https://doi.org/10.1111/j.1745-9133.2003.tb00024.x>
- Sanghara, K. K., & Wilson, J. C. (2006). Stereotypes and attitudes about child sexual abusers: A comparison of experienced and inexperienced professionals in sex offender treatment. *Legal and Criminological Psychology*, 11(2), 229–244. <https://doi.org/10.1348/135532505X68818>
- Sawilowsky, S. S. (1990). Nonparametric tests of interaction in experimental design. *Review of Educational Research*, 60(1), 91–126. <https://doi.org/10.2307/1170226>

- Schiavone, S. K., & Jeglic, E. L. (2009). Public perception of sex offender social policies and the impact on sex offenders. *International Journal of Offender Therapy and Comparative Criminology*, 53(6), 679–695. <https://doi.org/10.1177/0306624X08323454>
- Schmidt, F., Sinclair, S. M., & Thomasdóttir, S. (2016). Predictive validity of the Youth Level of Service/Case Management Inventory with youth who have committed sexual and non-sexual offenses: The utility of professional override. *Criminal Justice and Behavior*, 43(3), 413–430.
- Schwalbe, C. S. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review*, 30(12), 1458–1464. <https://doi.org/10.1016/j.childyouth.2007.11.021>
- Scott, S., Gilcrist, B., Thurston, N., & Huss, M. T. (2010). Risk communication in sexually violent predator hearings. *Behavioral Sciences & the Law*, 28(3), 322–336. <https://doi.org/10.1002-bsl.903>
- Seurich, N., Gongola, J., & Krauss, D. A. (2016). The biasing effect of the “sexually violent predator” label on legal decisions. *International Journal of Law and Psychiatry*, 47(1), 109–114. <https://doi.org/10.1016/j.ijlp.2016.05.002>
- Seurich, N., & John, R. S. (2012). Prescriptive approaches to communicating the risk of violence in actuarial risk assessment. *Psychology, Public Policy, and Law*, 18(1), 50–78. <https://doi.org/10.1037/a0024592>
- Seurich, N., & Krauss, D. A. (2013). The effect of adjusted actuarial risk assessment on mock-jurors’ decisions in a sexual predator commitment proceeding. *Jurimetrics*, 53(4), 395–414.
- Seurich, N., & Krauss, D. A. (2014). The presumption of dangerousness in sexually violent predator commitment proceedings. *Law, Probability and Risk*, 13(1), 91–104.
- Seurich, N., Monahan, J., & John, R. S. (2012). Innumeracy and unpacking: Bridging the nomothetic/idiographic divide in violence risk assessment. *Law and Human Behavior*, 36(6), 548–554. <https://doi.org/10.1037/h0093994>
- Shapiro, D. L., & Walker, L. E. A. (2016). Multiple relationships in forensic settings. In O. Zur (Ed.), *Multiple relationships in psychotherapy and counseling: Unavoidable, common, and mandatory dual relations in therapy* (pp. 82–91). Taylor & Francis.
- Shepherd, S. M., & Sullivan, D. (2017). Covert and implicit influences on the interpretation of violence risk instruments. *Psychiatry, Psychology and Law*, 24(2), 292–301. <https://doi.org/10.1080/13218719.2016.1197817>
- Shestowsky, D., & Horowitz, L. M. (2004). How the Need for Cognition Scale predicts behavior in mock jury deliberations. *Law and Human Behavior*, 28(3), 305–337. <https://doi.org/10.1023/B:LAHU.0000029141.46850.fb>
- Shooter, A., & Cooper, S. L. (2019). A template for enhancing the impact of the National Academy of Sciences’ reporting on forensic science. *British Journal of American Legal Studies*, 8(3), 441–464. <https://doi.org/10.2478/bjals-2019-0009>
- Simon, R. I. (2006). The myth of “imminent” violence in psychiatry and the law. *University of Cincinnati Law Review*, 75(2), 631–644.
- Singh, J. P., Assy, R., & Serpa, K. I. (2019). Violence risk assessment practices in Israel: A preliminary survey investigation. *Journal of Aggression, Conflict and Peace Research*, 11(2), 116–126. <https://doi.org/10.1108/JACPR-05-2018-0358>

-
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., & Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13(3), 193–206. <https://doi.org/10.1080/14999013.2014.922141>
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, 31(3), 499–513. <https://doi.org/10.1016/j.cpr.2010.11.009>
- Skorinko, J. L., & Spellman, B. A. (2013). Stereotypic crimes: How group-crime associations affect memory and (sometimes) verdicts and sentencing. *Victims & Offenders*, 8(3), 278–307. <https://doi.org/10.1080/15564886.2012.755140>
- Slovic, P., & Monahan, J. (1995). Probability, danger, and coercion: A study of risk perception and decision making in mental health law. *Law and Human Behavior*, 19(1), 49–65.
- Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24(3), 271–296.
- Smith, D. M. (2015). Dangerous diagnoses, risky assumptions, and the failed experiment of Sexually Violent Predator commitment. *Oklahoma Law Review*, 67(4), 619–732.
- Smith, K. J., O'Rourke, S., & Macpherson, G. (2020). The predictive validity of the HCR-20^{v3} within Scottish forensic inpatient facilities: A closer look at key dynamic variables. *International Journal of Forensic Mental Health*, 19(1), 1–17. <https://doi.org/10.1080/14999013.2019.1618999>
- Smith, S. M. (2013). *Evaluating racial bias in inpatient risk assessments* [Unpublished doctoral dissertation]. The City University of New York.
- Socia, K. M., & Harris, A. J. (2016). Evaluating public perceptions of the risk presented by registered sex offenders: Evidence of crime control theater? *Psychology, Public Policy, and Law*, 22(4), 375–385. <https://doi.org/10.1037/law0000081>
- Socia, K. M., Rydberg, J., & Dum, C. P. (2019). Punitive attitudes toward individuals convicted of sex offenses: A vignette study. *Justice Quarterly*. Advance online publication. <https://doi.org/10.1080/0741825.2019.1683218>
- Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 924–951). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118468333.ch33>
- Sommers, S. R., & Kassin, S. M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin*, 27(10), 1368–1377. <https://doi.org/10.1177/01461672012710012>
- Spector, R. (2001). Is there racial bias in clinicians' perceptions of the dangerousness of psychiatric patients? A review of the literature. *Journal of Mental Health*, 10(1), 5–15. <https://doi.org/10.1080/09638230020023570>
- Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). *Handleiding en Methodologische Verantwoording HKT-R, Historische, Klinische en Toekomstige – Revisie* [Guidelines and Methodological Research of the HKT-R, Historical, Clinical and Future – Revision]. Dr. van Mesdag kliniek.

- Sreenivasan, S., Weinberger, L. E., & Garrick, T. (2003). Expert testimony in sexually violent predator commitments: Conceptualizing legal standards of “mental disorder” and “likely to reoffend.” *The Journal of the American Academy of Psychiatry and the Law*, 31(4), 471–485. <https://doi.org/PMID: 14974803>
- Stalans, L. J. (1993). Citizens’ crime stereotypes, biased recall, and punishment preferences in abstract cases: The educative role of interpersonal sources. *Law and Human Behavior*, 17(4), 451–470.
- Stark, J. S., Bentley, R. J., Lowther, M. A., & Shaw, K. M. (1991). The student goals exploration: Reliability and concurrent validity. *Educational and Psychological Measurement*, 51(2), 413–422. <https://doi.org/10.1177/0013164491512015>
- Steblay, N., Besirevic, J., Fulero, S. M., & Jimenez-Lorente, B. (1999). The effects of pretrial publicity on juror verdicts: A meta-analytic review. *Law and Human Behavior*, 23(2), 219–235. <https://doi.org/10.1023/A:1022325019080>
- Stevenson, M., & Doleac, J. L. (2019). Algorithmic Risk Assessment in the Hands of Humans. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3489440>
- Stevenson, M. T. (2018). Assessing risk assessment. *Minnesota Law Review*, 103(1), 303–384. <https://dx.doi.org/10.2139/ssrn.3016088>
- Stone, D. (2018). Dangerous minds: Myths and realities behind the violent behavior of the mentally ill, public perceptions, and the judicial response through involuntary civil commitment. *Law & Psychology Review*, 42(1), 59–83.
- Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse: A Journal of Research and Treatment*, 24(3), 289–302. <https://doi.org/10.1177/1079063211423943>
- Strub, D. S., Douglas, K. S., & Nicholls, T. L. (2014). The validity of Version 3 of the HCR-20 violence risk assessment scheme amongst offenders and civil psychiatric patients. *International Journal of Forensic Mental Health*, 13(2), 148–159. <https://doi.org/10.1080/14999013.2014.911785>
- Studebaker, C. A., & Penrod, S. D. (1997). The media, the law, and common sense. *Psychology, Public Policy, and Law*, 3(2/3), 428–460.
- Tewksbury, R. (2005). Collateral consequences of sex offender registration. *Journal of Contemporary Criminal Justice*, 21(1), 67–81. <https://doi.org/10.1177/1043986204271704>
- Thompson, W. C. (2020). Commentary on: Curley LJ, Munro J, Lages M, MacLean R, Murray J. Assessing cognitive bias in forensic decisions: A review and outlook. *Journal of Forensic Sciences*, 65(2), 666–667. <https://doi.org/10.1111/1556-4029.14272>
- Truman, J. L., & Rand, M. R. (2009). *Criminal victimization, 2009* (NCJ 231327; National Crime Victimization Survey). U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. <https://www.bjs.gov/content/pub/pdf/cv09.pdf>
- Tully, J. (2017). HCR-20 shows poor field validity in clinical forensic psychiatry settings. *Evidence Based Mental Health*, 20(3), 95–96.
- Turner, D. B., Boccaccini, M. T., Murrie, D. C., & Harris, P. B. (2015). Jurors report that risk measure scores matter in sexually violent predator trials, but that other factors matter more: Importance of risk measures to SVP jurors. *Behavioral Sciences & the Law*, 33(1), 56–73. <https://doi.org/10.1002/bl.2154>
- Turner, S., Chamberlain, A. W., Jannetta, J., & Hess, J. (2015). Does GPS improve recidivism among high risk sex offenders? Outcomes for California’s GPS pilot for high risk sex offender parolees. *Victims & Offenders*, 10(1), 1–28. <https://doi.org/10.1080/15564886.2014.953236>

-
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (Eds.). (1982). Judgments of and by representativeness. In *Judgment under uncertainty: Heuristics and biases* (pp. 84–100). Cambridge University Press.
- Urbanik, F. (2007). *FOTRES: Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System [FOTRES: Forensic Operationalized Therapy/Risk Evaluation System]*. Zytglogge.
- van Horn, J., Eisenberg, M., & Uzieblo, K. (2016). Risicotaxatie in de Nederlandse ambulante forensische geestelijke gezondheidszorg [Risk assessment in the Dutch ambulatory forensic clinic]. *Tijdschrift voor Psychiatrie*, 58, 583–592.
- Vanginderen, K. G. (1998). Kansas v. Hendricks: Throwing away the key. *Thomas Jefferson Law Review*, 20(2), 357–375.
- Varela, J. G., Boccaccini, M. T., Cuervo, V. A., Murrie, D. C., & Clark, J. W. (2014). Same score, different message: Perceptions of offender risk depend on Static-99R risk communication format. *Law and Human Behavior*, 38(5), 418–427. <https://doi.org/10.1037/lhb0000073>
- Vera, L. M., Boccaccini, M. T., Laxton, K., Bryson, C., Pennington, C., Ridge, B., & Murrie, D. C. (2019). How does evaluator empathy impact a forensic interview? *Law and Human Behavior*, 43(1), 56–68. <https://doi.org/10.1037/lhb0000310>
- Verplanken, B. (1991). Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment. *Personality and Social Psychology Bulletin*, 17(2), 188–193. <https://doi.org/10.1177/014616729101700211>
- Vess, J., & Skelton, A. (2010). Sexual and violent recidivism by offender type and actuarial risk: Reoffending rates for rapists, child molesters and mixed-victim offenders. *Psychology, Crime & Law*, 16(7), 541–554. <https://doi.org/10.1080/10683160802612908>
- Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., Vargen, L. M., & Vincent, G. M. (2019). Impact of risk assessment instruments on rates of pretrial detention, postconviction placements, and release: A systematic review and meta-analysis. *Law and Human Behavior*, 43(5), 397–420. <https://doi.org/10.1037/lhb0000344>
- Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17(3), 377–395. <https://doi.org/10.1177/1073191109359587>
- Vitacco, M. J., Erickson, S. K., Kurus, S., & Apple, B. N. (2012). The role of the Violence Risk Appraisal Guide and Historical, Clinical, Risk-20 in U.S. courts: A case law survey. *Psychology, Public Policy, and Law*, 18(3), 361–391. <https://doi.org/10.1037/a0025834>
- Vitacco, M. J., Gottfried, E. D., & Batastini, A. B. (2018). Using technology to improve the objectivity of criminal responsibility evaluations. *The Journal of the American Academy of Psychiatry and the Law*, 46(1), 71–77.
- Walters, G. D., Kroner, D. G., DeMatteo, D., & Locklair, B. R. (2014). The impact of base rate utilization and clinical experience on the accuracy of judgments made with the HCR-20. *Journal of Forensic Psychology Practice*, 14(4), 288–301. <https://doi.org/10.1080/15228932.2014.941726>
- Webster, C. D., Douglas, K. S., Eaves, D. C., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Simon Fraser University, Mental Health Law and Policy Institute.
- Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short-Term Assessment of Risk and Treatability (START)*. St. Joseph's Healthcare, Hamilton and British Columbia Mental Health and Addiction Services.

- Wegener, D. T., Kerr, N. L., Fleming, M. A., & Petty, R. E. (2000). Flexible corrections of juror judgments: Implications for jury instructions. *Psychology, Public Policy, and Law*, 6(3), 629–654. <https://doi.org/10.1037/1076-8971.6.3.629>
- Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naive theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68(1), 36–51. <https://doi.org/10.1037/0022-3514.68.1.36>
- Weinberger, L. E., & Sreenivasan, S. (2018). Addressing ethics dilemmas in violence-risk assessment: A forensic psychologist perspective. In E. Griffith (Ed.), *Ethics challenges in forensic psychiatry and psychology practice* (pp. 284–303). Columbia University Press. <https://doi.org/10.7312/grif18330-019>
- Weinberger, L. E., Sreenivasan, S., Azizian, A., & Garrick, T. (2018). Linking mental disorder and risk in sexually violent person assessments. *The Journal of the American Academy of Psychiatry and the Law*, 46(1), 63–70.
- Weiner, C., Shackley, M., Willis, G. M., & Day, A. (2014). The influence of offence severity and risk of re-offence on judgements towards sex offenders. *Sexual Abuse in Australia and New Zealand*, 6(1), 3–11.
- Werkgroep Pilotstudy Risicotaxie. (2002). *Findings of a nationwide pilot study on the HKT-30*. Ministerie van Justitie.
- Werth, R. (2019). Risk and punishment: The recent history and uncertain future of actuarial, algorithmic, and “evidence-based” penal techniques. *Sociology Compass*, 13(2), e12659. <https://doi.org/10.1111/soc4.12659>
- Wevodau, A. L., Cramer, R. J., Clark, J. W., & Kehn, A. (2014). The role of emotion and cognition in juror perceptions of victim impact statements. *Social Justice Research*, 27(1), 45–66. <https://doi.org/10.1007/s11211-014-0203-9>
- Wevodau, A. L., Cramer, R. J., Gemberling, T. M., & Clark, J. W. (2016). A psychometric assessment of the Community Attitudes Toward Sex Offenders (CATSO) Scale: Implications for public policy, trial, and research. *Psychology, Public Policy, and Law*, 22(2), 211–220. <https://doi.org/10.1037/law0000066>
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116(1), 117–142. <https://doi.org/10.1037/0033-295X.116.1.117>
- Witt, P. H., & DeMatteo, D. (2019). Sexually violent predator laws: Historical development and evolution. In W. T. O'Donohue & D. S. Bromberg (Eds.), *Sexually Violent Predators: A clinical science handbook* (pp. 9–20). Springer International Publishing. https://doi.org/10.1007/978-3-030-04696-5_2
- Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law*, 12(3), 279–309. <https://doi.org/10.1037/1076-8971.12.3.279>
- Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39(12), 1511–1538. <https://doi.org/10.1177/0093854812455741>
- Wurtele, S. K., Kvaternick, M., & Franklin, C. F. (1992). Sexual abuse prevention for preschoolers: A survey of parents' behaviors, attitudes, and beliefs. *Journal of Child Sexual Abuse*, 1(1), 113–128.
- Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136(5), 740–767. <https://doi.org/10.1037/a0020473>

-
- Yanich, D. (2004). Crime creep: Urban and suburban crime on local TV news. *Journal of Urban Affairs*, 26(5), 535–563. <https://doi.org/10.1111/j.0735-2166.2004.00214.x>
- Yung, C. R. (2010). The emerging criminal war on sex offenders. *Harvard Civil Rights-Civil Liberties Law Review*, 45(2), 435–481.
- Zapf, P. A., & Dror, I. E. (2017). Understanding and mitigating bias in forensic evaluation: Lessons from forensic science. *International Journal of Forensic Mental Health*, 16(3), 227–238. <https://doi.org/10.1080/14999013.2017.1317302>
- Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheeles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice*, 4(1), 27–44. https://doi.org/10.1300/J158v04n01_02
- 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. (2017). 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>
- Zolfo, M. (2018). Commitment through fear: Mandatory jury trials and substantive due process violations in the civil commitment of sex offenders in Illinois. *Chicago-Kent Law Review*, 93(2), 593–627.

R

Impact Statement

Impact Statement

The human brain has an incredible capacity to make quick and often accurate judgments in complex situations. Yet, psychological research has repeatedly demonstrated that our initial judgments based on intuition and automatic processing can be less than optimal, or even incorrect. When decisions have far reaching consequences, such as in legal decision making, intuitive judgments are often not desirable. It is precisely the circumstances of high stakes and complex information in which decisions are made by judges, jurors, and mental health experts who are called upon to provide an evaluation of an individual's risk of violent or sexual recidivism. Furthermore, research about cognitive bias and heuristics has demonstrated that adjusting initial intuitive impressions when new information becomes available takes focused and intentional cognitive effort that may be difficult to summon, particularly in emotionally-charged situations.

The research findings in this dissertation related to forensic risk evaluations have several implications for forensic mental health evaluators. First, forensic mental health professionals should consider the potential for contextual factors and task-irrelevant information to influence their perceptions of recidivism risk, including when they are relying upon a structured risk assessment instrument. Affirmative steps that may help minimize errors in judgment related to bias include avoiding exposure to task-irrelevant information and employing a consensus-based approach to risk evaluations. Second, the use of a structured risk assessment instrument may provide a guide for focusing on relevant risk factors but should not be viewed as proof that a risk evaluation has not been affected by bias. Nevertheless, based on current standards and guidelines, the use of a structured instrument to assess recidivism risk is preferred to relying upon unstructured clinical judgment. Finally, where a forensic evaluator is called upon to provide a report or testimony to legal decision-makers, transparency about the information obtained and the process employed to arrive at a risk judgment may help inform decisions in a way that making only conclusory statements does not.

The findings from the mock juror experimental studies in this dissertation, in contrast to the results from comparable previous studies, suggest that providing lay jurors with factual information contained in an actuarial risk assessment instrument (the Static-99R) can be effective in mitigating the influence of their initial impression of a defendant in a sexually violent predator (SVP) civil commitment case (hereafter, SVP respondent). Unlike previous studies that have typically used expert testimony about the results of a risk assessment instrument, leaving open the possibility that the credibility or likability of the expert affects the weight jurors give the testimony, we provided mock jurors with a list of the risk factors contained in the Static-99R, the total score, and the observed recidivism rates derived from it. Providing this "pared-down" version of sexual recidivism risk assessment resulted in mock

jurors' giving more realistic (and lower) estimates of the SVP respondent's risk of sexual recidivism, regardless of their initial impressions of his likability. Furthermore, the actuarial risk information eliminated observed between-group differences in civil commitment recommendations, wherein mock jurors who initially read negative risk-irrelevant contextual information about the SVP respondent were significantly more likely to recommend civil commitment than those who read initial positive risk-irrelevant information about the SVP respondent. The encouraging real-world implications of these findings suggest that providing jurors directly with the list of risk factors and associated recidivism rates may decrease the likelihood of jurors making civil commitment decisions based on their initial emotionally laden impressions of an SVP respondent, rather than the relevant criteria that must be satisfied as legal justification for civil commitment.

However, the findings from the three experimental studies about mock jurors' judgments in SVP civil commitment cases indicate that jurors are unlikely to base their judgments solely on a sexual recidivism risk estimate. Specifically, we found that only a small minority of mock jurors ultimately rendered an estimate of sexual recidivism risk that was within the range of observed recidivism rates indicated by the Static-99R. In fact, we found that mock jurors' preexisting, grossly overestimated sexual recidivism rates were significantly related to their estimates of sexual recidivism risk in the individual case, both before and after receiving information about actual, observed sexual recidivism rates provided in the Static-99R. Moreover, mock jurors who endorsed harsh sentencing and restrictive management strategies for people convicted of a sexual offense exhibited an increased likelihood of recommending civil commitment, irrespective of the Static-99R information or their own estimates of the SVP respondent's risk of sexual recidivism. In fact, among mock jurors who recommended civil commitment, we found that their own ratings of the importance of the Static-99R to their recommendation did not significantly predict whether they recommended civil commitment. Perhaps most concerning was our finding across all three studies that the more importance mock jurors assigned to the impact of the crime on the victim in their decision making, the more likely they were to vote for civil commitment.

Taken together, the findings from the series of SVP civil commitment studies have implications specific to SVP civil commitment cases. First, providing jurors with a list of sexual recidivism risk factors and the risk estimate derived from a structured risk assessment instrument for the case at hand, may help mitigate the influence of risk-irrelevant factors (such as initial impressions of the SVP respondent and jurors' preexisting beliefs about sexual recidivism rates). Second, screening methods might be employed in an effort to dismiss jurors who are likely to be unwilling or unable to objectively weigh the evidence as it relates to the legal criteria required for civil commitment decisions. Finally, whereas the impact of

the crime on the victim appears to significantly affect commitment decisions, courts should seriously consider emphasizing that SVP civil commitment is not intended as punishment, and therefore jurors should avoid punishment-seeking motives in their decisions (although I acknowledge that such an instruction may have limited effects).

More broadly, there are several implications for legal systems employing structured risk assessment instruments to inform legal decision-making. First, the extent to which estimated recidivism rates indicated by a structured risk assessment instrument influence the judgments made by the decision-maker (i.e., a judge or juror) are likely to vary depending on factors that are unrelated to recidivism risk (such as risk-irrelevant characteristics of the defendant) and the personal attributes of the decision-maker (for example, their preexisting beliefs and attitudes). In other words, when it comes to human judgment, statistical information about recidivism rates may be expected to have limited effects on the decision given the presence of more salient factors, even if the decision-maker is not consciously aware of how such factors influence their decisions. Therefore, an expectation that risk assessment instruments will necessarily result in less biased and more objective decision-making should be tempered by the realities of how the human mind processes information. In fact, it is more likely that a risk estimate derived from a structured risk assessment instrument, like other pieces of evidence, will be weighed according to the judgment of the decision-maker about the weight it deserves. Structured risk assessment instruments can inform human judgment; they do not replace it.

Finally, as risk assessment instruments are lauded and implemented in various decision-making tasks in the forensic mental health and legal arena, researchers should conduct research that focuses on how structured measures of risk and resulting risk assessment affect human judgments and decisions. Whether the decision is about the appropriate type of treatment indicated, the level of supervision to which an individual should be subjected, or whether an individual can be released to the community with low risk to the public, the likelihood of recidivism indicated by a structured risk assessment instrument will always be viewed from the perspective of the decision-maker.

IS

Summary

Summary

There were two primary aims of this dissertation. First, was to investigate whether external or internal sources of bias influence trained clinical mental health evaluators' ratings and judgments of an examinee's violence recidivism risk using an SRAI. The findings presented in the HCR-20 study (Chapter 3) indicated that HCR-20 scores and risk judgments were significantly related to evaluator attitudes toward offenders, but not to media exposure. The second aim of the research in this dissertation was to investigate empirically whether bias affects mock jurors' use of SRAI information in making estimates of an individual's likelihood of recidivism risk and whether individual juror characteristics affect this relationship. As indicated by the findings in the studies in Chapters 4–6, risk-irrelevant contextual information had a biasing effect on mock jurors' estimates of sexual recidivism risk and recommendations for civil commitment. A risk estimate derived from an SRAI was mostly, but not completely, effective in mitigating the effects of bias on mock jurors' estimates of sexual recidivism risk. Finally, I obtained forensic clinicians' ratings of potential bias correction strategies to examine what they believe may be effective, which may guide further research efforts to empirically test strategies to mitigate bias in forensic evaluations.

Structured risk assessment instruments (SRAIs) have gained increased influence over the past several decades as a valid method for discriminating between people who are at high risk of criminal, violent, or sexual recidivism from those who are at low risk of recidivism. In the introductory chapter to this dissertation (Chapter 1), I discussed the importance of violent and sexual recidivism risk estimates to forensic mental health clinicians and legal decision-makers (i.e., judges and jurors). I explained the methods of evaluating recidivism risk, and specifically the use of SRAIs for this purpose. I also presented previous research findings that indicate that cognitive bias resulting from factors external or internal to a forensic evaluator can affect the ratings they provide using an SRAI. In addition, I discussed research indicating there are external and internal factors that have the potential to bias jurors' interpretations of an SRAI risk estimate. Finally, I discussed a few strategies that are potentially promising to mitigate the potential for bias to affect forensic risk evaluations.

Forensic mental health practitioners often use SRAIs to estimate recidivism risk, as well as to identify potential treatment targets to reduce the risk of violent or sexual reoffending, and to develop plans to mitigate the risk of reoffending. For example, the findings reported in Chapter 2 of this dissertation indicate that Dutch forensic mental health practitioners use an SRAI always or most of the time when conducting a forensic risk evaluation and they rate SRAIs as very useful for this task. Despite the high rate at which SRAIs were used among the survey respondents, they nevertheless expressed concerns that cognitive bias may affect

a forensic risk evaluation they performed themselves, as well as those performed by their colleagues.

Regarding bias correction strategies, survey respondents endorsed using structured evaluation methods as the most effective strategy to minimize the potential for bias in a forensic risk evaluation. They also provided high ratings to several other debiasing strategies including continuing to learn after completing formal education and training, investigating all relevant data before forming an opinion and basing conclusions and opinions on sound data, and being an active consumer of scientific knowledge, among others. Although several scholars have written recently about the potential for bias to affect forensic evaluations, empirical testing of potential debiasing strategies with respect to forensic evaluations is notably limited.

Another significant gap in research regarding forensic evaluations is identifying how cognitive bias that results from factors external to the evaluator or individual evaluator characteristics affects evaluator ratings of violent recidivism risk using an SRAI. Therefore, I conducted the study presented in Chapter 3 to test the effects of media exposure and evaluator attitudes toward offenders on their ratings of an examinee's risk of violent recidivism using the HCR-20^{v3}. Contrary to expectations, negative media about a forensic examinee resulted in significantly fewer final risk judgments indicating high risk among evaluators exposed to the media as compared to a control group that based their ratings solely on case file information. Yet, I also found that evaluator attitudes toward offenders significantly predicted lower HCR-20 total scores and were significantly correlated with lower ratings of the examinee's clinical risk, risk management, and risk for imminent violence. These findings add to previous research that indicates SRAIs cannot be presumed to eliminate the effects of all sources of bias.

The second aim of this dissertation was to examine the extent to which a sexual recidivism risk estimate derived from an SRAI mitigated the biasing effects of risk-irrelevant contextual information on mock jurors' estimates of sexual recidivism risk in a Sexually Violent Predator (SVP) civil commitment case. One of the legal criteria that must be met to civilly commit a person under a so-called SVP statute is that the SVP *respondent*—the person whom the government seeks to commit—is more likely than not to engage in sexual reoffending. Jurors, who often decide these cases, hear expert testimony about the respondent's sexual recidivism risk based on an actuarial risk estimate, although they appear relatively insensitive to this estimate compared to other factors.

Because previous research has found that mock and actual jurors give little weight to expert testimony about sexual recidivism risk estimates, I also investigated whether some individual characteristics among mock jurors had a direct or indirect effect on their estimates of an SVP respondent's sexual recidivism risk. The individual characteristics I examined included mock jurors' level of need for cognition, asymmetrical skepticism toward an SRAI

risk estimate, and mock jurors' perceptions of people who have committed a sexual offense (PCSO).

I found that an SRAI sexual recidivism risk estimate was effective in mitigating a bias created by risk-irrelevant contextual information in the studies presented in Chapters 4 and 5. In the study in Chapter 4, I found that there was no moderating effect of mock jurors' level of need for cognition on risk-irrelevant contextual information about the SVP respondent and mock jurors' estimates of the respondent's risk of sexual reoffending. However, because the level of need for cognition was relatively high among the mock jurors in the sample, I could not exclude the possibility that mock jurors who are lower in need for cognition would have showed relatively less sensitivity to the SRAI risk estimate.

In the study in Chapter 5, I again found that an SRAI sexual recidivism risk estimate mitigated the effects of bias toward the SVP respondent based on risk-irrelevant contextual information. I examined whether there would be an asymmetrical skepticism effect toward an SRAI risk estimate, such that mock jurors would give lower ratings of the accuracy and reliability of the SRAI when it conflicted with their initial impression of the SVP respondent. However, I did not find evidence to support this; in fact, mock jurors overall rated the SRAI risk estimate as significantly less accurate and reliable when the risk estimate indicated below-average risk of sexual recidivism than when it indicated above-average risk. Yet, there is reason to suspect that in the studies in Chapters 4 and 5, mock jurors may have given relatively more weight to an SRAI risk estimate because they did not perceive the SVP respondent as a "stereotypical" sex offender.

In the final study presented in Chapter 6, using the same design as the study presented in Chapter 4, I found that the SRAI risk estimate did not eliminate the biasing effects of risk-irrelevant information on mock jurors' estimates of the SVP respondent's risk of sexual recidivism. However, the SRAI risk estimate did mitigate the bias with respect to civil commitment decisions in that there was no significant difference between the contextual information groups with respect to votes for civil commitment.

Nevertheless, there was a significant direct effect of mock jurors' preexisting perceptions of PCSOs on their estimates of the SVP respondent's sexual recidivism risk, such that more negative attitudes predicted higher risk estimates both before and after receiving the SRAI risk estimate. Furthermore, mock jurors' level of support for harsh sentencing and restrictive management of PCSOs in the community significantly predicted votes for civil commitment both before and after receiving the SRAI risk estimate. Taken together, these results indicate that mock jurors' estimates of the SVP respondent's risk of sexual recidivism and recommendations for civil commitment were significantly influenced by their preexisting perceptions of PCSOs.

The final chapter (Chapter 7) provided a summary of the major empirical findings from the preceding chapters and placed these findings in the context of previous research that indicates that variations in SRAI ratings can be attributed to factors that are unrelated to the examinee's risk of violent or sexual recidivism, thereby indicating potential errors related to bias. I also discussed the theoretical and practical implications of the findings from the studies in this dissertation for forensic mental health practice and legal decision-making. Finally, I discussed the limitations of the studies and directions for future research that examines the potential for bias to affect completion of SRAIs or interpretations of a risk estimate derived from an SRAI.

Overall, the findings from the studies presented in this dissertation provide further evidence that while SRAIs can be effective in mitigating the effects of cognitive bias with respect to some types of risk-irrelevant contextual information, they may be less effective in doing so with respect to internal sources of cognitive bias. The findings also demonstrate the importance of developing other reliable methods of debiasing in forensic risk evaluations. Although SRAIs can help organize risk-relevant information, ultimately, perceptions of recidivism risk appear to be in the eye of the beholder.

Sammanfattning

(Swedish Summary)

Sammanfattning (Swedish Summary)

Den forskning som har genomförts inom ramen för denna avhandling har haft två huvudsakliga syften. Det första syftet har varit att undersöka huruvida yttre eller inre källor till bias påverkar utbildade psykiatriska bedömarens skattningar och bedömningar av en undersökt persons risk att återfalla i våldsbrott, utförda av med hjälp av ett *strukturerat riskbedömningsinstrument* (Structured Risk Assessment Instrument, SRAI). De fynd som presenteras i kapitel 3 visade att bedömningar med instrumentet HCR-20 hade ett signifikant samband med bedömarens attityder gentemot lagöverträdare, men inte med exponering för mediaövervakning. Det andra syftet var att undersöka huruvida bias påverkar fingerade jurymedlemmars användning av SRAI-information vid uppskattning av en persons återfallsrisk och huruvida detta förhållande påverkas av den enskilda jurymedlemmens egenskaper. Såsom framgår av fynden från studierna i kapitel 4–6 gav risk-irrelevant kontextuell information upphov till bias i jurymedlemmarnas bedömningar av risk för återfall i sexualbrott och rekommendationer om tvångsvård. En riskuppskattning baserad på en SRAI kunde till stor del, men inte helt, motverka inverkan av bias på bedömningarna av risk för återfall i sexualbrott. Slutligen inhämtades skattningar av olika tänkbara strategier för korrigering av bias från rättspsykiatriskt verksamma kliniker, i syfte att ta reda på vilka strategier de ansåg vara effektiva för att undgå bias (kapitel 2). Dessa resultat kan vara till ledning för vidare forskningsinsatser, där strategier för att motverka bias i forensiska bedömningar prövas empiriskt.

Strukturerade riskbedömningsinstrument (SRAI) har under de senaste årtiondena fått allt större inflytande som en valid metod för att skilja personer med hög risk för återfall i vålds- eller sexualbrott från personer vars återfallsrisk är låg. I denna avhandlings inledande kapitel (kapitel 1) diskuterar jag den betydelse som en uppskattning av risken för återfall i sexualbrott kan ha för rättspsykiatiska kliniker och rättsliga beslutsfattare (exempelvis domare och juryledamöter). Jag förklarar metoderna för bedömning av återfallsrisk och särskilt hur SRAI:er används för detta syfte. Jag redogör också för tidigare forskningsfynd som tyder på att kognitiva bias, till följd av faktorer som är externa eller interna i förhållande till den forensiska bedömare, kan påverka de skattningar som denne gör med hjälp av en SRAI. Därtill tar jag upp forskning som tyder på att det finns externa och interna faktorer som kan leda till bias i juryledamöters tolkning av en SRAI-riskuppskattning. Slutligen diskuterar jag några potentiellt lovande strategier för att motverka risken att forensiska riskbedömningar påverkas av bias.

Rättspsykiatriska yrkesutövare använder ofta SRAI:er för att uppskatta återfallsrisken liksom för att identifiera personer som är mottagliga för behandling i syfte att minska risken för återfall i vålds- eller sexualbrott samt att ta fram planer för att minska risken för upprepad brottslighet. Exempelvis tyder de fynd som rapporteras i denna avhandlings kapitel 2 på att

nederländska rättspsykiatriska yrkesutövare använder sig av en SRAI varje eller nästan varje gång de utför en forensisk riskbedömning och att de skattar SRAI:er som mycket användbara för denna uppgift. Även om enkätens respondenter i stor utsträckning använder sig av SRAI:er, uttrycker de ändå farhågor om att kognitiva bias kan inverka på en forensisk riskbedömning som utförs av såväl dem själva som av deras kolleger.

När det gäller strategier för korrigering av bias förespråkade enkätens respondenter strukturerade bedömningsmetoder som den effektivaste strategin för att minimera möjliga bias vid en forensisk riskbedömning. De gav också höga skattningar av flera andra strategier för att motverka bias, däribland att fortsätta lära sig efter avslutad utbildning, att undersöka alla relevanta data innan man bildar sig en uppfattning, att basera slutsatser och omdömen på säkra data samt att aktivt tillägna sig vetenskaplig kunskap. Även om många vetenskapliga författare på senare tid har skrivit om möjligheten att forensiska bedömningar kan påverkas av bias, så finns en anmärkningsvärd brist på empirisk prövning av möjliga strategier för att motverka bias i samband med forensiska bedömningar.

En annan anmärkningsvärd lucka i forskningen kring forensiska bedömningar gäller identifieringen av hur kognitiva bias, till följd av faktorer utanför bedömaren eller bedömarens individuella egenskaper, påverkar bedömarnas skattningar av risk för återfall i våldsbrott med hjälp av en SRAI. Den experimentella studie som presenteras i kapitel 3 testade därför inverkan av medieexponering och bedömarnas attityder gentemot brottslingar på deras bedömningar av en undersökt persons risk för återfall i våldsbrott med hjälp av instrumentet HCR-20^{v3}. Tvärtemot vad som förväntades ledde exponering för negativ medierapportering om en forensisk undersökningsperson till signifikant färre slutliga riskbedömningar som indicerade hög risk, jämfört med en kontrollgrupp som baserat sina bedömningar enbart på informationen i fallbeskrivningen. Bedömarnas attityd gentemot brottslingar korrelerade dock med totalpoäng på HCR-20, skattningar av klinisk risk, riskhantering och omedelbar risk för våld hos den undersökta personen. Dessa fynd ansluter till tidigare forskning som tyder på att SRAI:er inte kan antas eliminera inverkan av samliga källor till bias.

Det andra syftet med denna avhandling var att undersöka i vilken utsträckning en bedömning baserad på en SRAI, av risken för återfall i sexualbrott, kan minska inverkan av bias på risk-irrelevant kontextuell information vid en fingerad jurys uppskattning av risken för återfall i sexualbrott i ett fall rörande tvångsvård av en *sexualbrottsförövare med hög återfallsrisk* (Sexually Violent Predator, SVP). Ett av den amerikanska SVP-lagens rekvisit för tvångsvård är att det ska vara mer troligt än icke troligt att den *svarande* SVP-personen – alltså den som staten vill tvångsvårda – kommer att föröva sexualbrott ånyo. Dessa fall avgörs ofta av juryledamöter som får höra vittnesmål från sakkunniga angående den svarandes risk för återfall i sexualbrott utifrån en matematisk riskuppskattning. De förefaller dock vara relativt okänsliga för denna riskuppskattning jämfört med andra faktorer.

Tidigare forskning tyder alltså på att ledamöter i fingerade och faktiska juryer fäster liten vikt vid sakkunnigas vittnesmål om uppskattad risk för återfall i sexualbrott. Därför undersöktes om vissa individuella egenskaper hos den fingerade juryns ledamöter hade en direkt eller indirekt inverkan på bedömningen av risken att en svarande SVP-person skulle återfalla i sexualbrott. De individuella egenskaperna hos den fingerade juryns ledamöter var bland annat deras kognitionsbehov, asymmetrisk skepsis mot en SRAI-riskuppskattning samt ledamöternas uppfattningar om sexualbrottsförövare.

En SRAI-baserad uppskattning av risken för återfall i sexualbrott visade sig, i de studier som presenteras i kapitel 4 och 5, verksam för att motverka bias till följd av risk-irrelevant kontextuell information. I studien som beskrivs i kapitel 4 fanns ingen modererande effekt av kognitionsbehovet hos ledamöterna i den fingerade juryn, vad gäller effekten av risk-irrelevant kontextuell information om den svarande SVP-personen, vid deras uppskattning av risken för återfall i sexualbrott. Kognitionsbehovet var dock relativt högt inom stickprovet och det går därför inte att utesluta möjligheten att deltagare med lägre kognitionsbehov skulle ha uppvisat en relativt sett lägre känslighet för SRAI-riskuppskattningen.

I den studie som presenteras i kapitel 5 motverkades åter effekterna av bias genom att risken för återfall i sexualbrott uppskattades med hjälp av en SRAI. Här rörde det sig om bias gentemot en svarande SVP-person bland ledamöterna av en fingerad jury till följd av risk-irrelevant kontextuell information. Det fanns emellertid inga belägg för att asymmetrisk skepsis skulle ha inverkat på SRAI-riskuppskattningen, på så sätt att den fingerade juryns ledamöter skulle ha gett lägre skattningar av SRAI-instrumentets exakthet och reliabilitet när det stod i strid med deras första intryck av den svarande SVP-personen. Överlag skattade den fingerade juryns medlemmar faktiskt SRAI-riskuppskattningen som signifikant mindre exakt och reliabel när riskuppskattningen visade på en lägre risk för återfall i sexualbrott än genomsnittet, än när den visade på en högre risk än genomsnittet. Det finns ändå skäl att misstänka att den fingerade juryns ledamöter, i de studier som beskrivits i kapitel 4 och 5, har fått relativt större vikt vid en SRAI-riskuppskattning eftersom de inte uppfattat den svarande SVP-personen som en ”stereotyp” sexualförbrytare.

I den sista studien, som presenteras i kapitel 6 och har samma utformning som studien i kapitel 4, utfördes uppskattningar av en svarande SVP-persons risk för återfall i sexualbrott av ledamöter i en fingerad jury. Här kunde SRAI-riskuppskattningen inte eliminera inverkan av bias, till följd av risk-irrelevant information. Skillnaden mellan de genomsnittliga uppskattningarna av risk för återfall i sexualbrott, som avgivits av de två fingerade jurygrupperna, var visserligen signifikant men differensens ringa storlek innebär att den sannolikt inte har någon större (om alls någon) praktisk betydelse. SRAI-riskuppskattningen minskade emellertid inte inverkan av bias på tvångsvårdsbeslutet, i och med att det inte

förelåg någon signifikant skillnad mellan de grupper som fått olika kontextuell information, vad gäller den andel som röstade för tvångsvård.

Deltagarnas förhandsuppfattningar om sexualbrottsförövare hade ändå en signifikant direkt effekt på deras skattningar på så sätt att mer negativa attityder predicerade högre riskuppskattningar både före och efter att ha tagit del av SRAI-riskuppskattningen. Dessutom predicerade stöd för stränga påföljder och restriktiv hantering av sexualbrottsförövare i samhället signifikant hur ledamöterna i de fingerade juryerna röstat om tvångsvård både före och efter att de tagit del av SRAI-riskuppskattningen. Sammantaget tyder dessa resultat på att ledamöterna i en fingerad jury, i sina uppskattningar av en svarande SVP-persons risk för återfall i sexualbrott och vid rekommendationer om tvångsvård, signifikant påverkats av sina tidigare uppfattningar om sexualbrottsförövare.

Det sista kapitlet (kapitel 7) ger en sammanfattning av de främsta empiriska fynden från föregående kapitel och sätter dessa fynd i ett sammanhang med tidigare forskning som tyder på att variationer i SRAI-bedömningar kan tillskrivas faktorer som inte är relaterade till den undersökta personens risk för återfall i vålds- eller sexualbrott, vilket alltså tyder på potentiella fel relaterade till bias. Teoretiska och praktiska implikationer för rättspsykiatrisk praxis och rättsligt beslutsfattande tas upp. Slutligen berör avhandlingen studiernas begränsningar och ger anvisningar för möjlig framtidiga forskning som undersöker en potentiell påverkan av bias vid ifyllandet av SRAI:er eller tolkningar av en riskuppskattning som härletts från en SRAI. Avhandlingen avslutas med sammanfattande kommentarer.

På det stora hela kan fynden från de studier som presenteras i denna avhandling ge ytterligare belägg för att SRAI:er visserligen kan vara effektiva för att motverka effekterna av kognitiva bias, till följd av vissa slags risk-irrelevant kontextuell information, men också mindre effektiva i detta avseende när det gäller interna källor till kognitiva bias. Fynden visar också på vikten av att utveckla andra tillförlitliga metoder för att motverka bias vid forensiska riskbedömningar. Även om SRAI:er kan underlätta sammanställningen av risk-relevant information tycks den upplevda risken för återfall, sist och slutligen, ligga i betraktarens öga.

Acknowledgments

Acknowledgments

The experience of conducting original research and turning the findings into this dissertation was one of the most challenging and rewarding opportunities I have had in life. I learned so much from my supervisors, who guided and supported me throughout this journey. Maartje, the hours you and I spent discussing the work and my writing was a source of comfort and consistency, particularly early on when I was struggling to adapt to the nuances of authoring a scientific manuscript. Marko, I appreciated your easygoing nature and your capacity for the attention to details in my writing. Karl, the time you spent with me talking through study design and statistical analysis was invaluable to my development as a researcher. Without your insights and guidance, I would not have achieved the quality of studies that comprise the main body of this dissertation. And finally, Corine, I would not have experienced the growth I have as a writer without your input. Your acumen and incisiveness as an editor of my work ultimately became what I think is one of the most important aspects of my growth as an academic. To all four of you, I am humbly grateful for your time, your patience, and your dedication in ensuring I was able to reach the greatest achievement of my life. I thank you for being part of my journey.

It's difficult for me to express how much my life changed for the better as a result of being part of the House of Legal Psychology and the people who made my time colorful, joyful, and filled with fun and laughter. To my paranympths, officemates, and "younger sisters": Carey and Bruna, I am indebted to you for being there to listen when I struggled and for pulling me out of my comfort zone and generally helping me be a better, more open version of me. To Tammy, I thank you for the talks and the unending support you have shown me as I have moved along my path. Your strength, determination, and sense of purpose inspired me and left an indelible impression on my heart. And to Enide, thank you for motivating me to run, especially on the days I didn't want to. When I felt like I wasn't making progress in my work, you helped me to feel like I accomplished something by moving my feet.

I am also thankful to all of the PhD researchers, faculty, and staff of the House of Legal Psychology who shared a part in this incredible experience. I am especially grateful to the Clinical Psychology section of Maastricht University and the many of you who I now consider friends.

No matter the road, or the distance, I know I have never traveled without the love and support of my family and friends. My Mom and Dad suffered through weekly calls with me, lending a sympathetic ear and words of encouragement through every struggle and victory I achieved. At every turn, you have encouraged me to forge my own path and to keep striving. Thank you for being with me no matter where I go. I love you. To my brother and sister, Jim and Jamie, thank you for cheering me on—I love you both! While I am so fortunate to have

support and encouragement from many family members and friends, there are a few that stand out. Sarah, without you, I probably wouldn't have even applied for the program! Thank you for pushing me to take a chance. Catey, we have seen each other through many trials and tribulations. My life is better for you being in it and I hope you know how much your friendship means to me. I hope we have a few more decades together!

Whereas scientific research taught me about what can be measured and quantified, the people and places that have graced my life have shown me what is immeasurable.

Research Dissemination

Research Dissemination

Peer-Reviewed Publications

Kamorowski, J., Ask, K., Schreuder, M., Jelicic, M., & de Ruiter, C. (2021). "He seems odd": The effects of risk-irrelevant information and actuarial risk estimates on mock jurors' perceptions of sexual recidivism risk. *Psychology, Crime and Law*.

Kamorowski, J., de Ruiter, C., Schreuder, M., Ask, K., & Jelicic, M. (2021). Forensic mental health practitioners' use of structured risk assessment instruments, views about bias in risk evaluations, and strategies to counteract it. *International Journal of Forensic Mental Health*.

Kamorowski, J., Schreuder, M., de Ruiter, C., Jelicic, M., & Ask, K. (2018). Risk assessment tools and criminal offending: Does bias determine who is "high risk"? *The Inquisitive Mind*, 38. <https://www.in-mind.org/article/risk-assessment-tools-and-criminal-reoffending-does-bias-determine-who-is-high-risk>

Conference Presentations

Kamorowski, J. (June 2019). Anatomy of risk: Cumulative disadvantage and risk assessment instruments. Presented orally at the International Association of Forensic Mental Health Services conference, Montreal, Canada.

Kamorowski, J. (November 2018). Perceptions of cognitive bias and the use of risk assessment instruments among Dutch mental health professionals. Presented orally at the European Association of Psychology and Law (EAPL) conference, Turku, Finland.

Curriculum Vitae

Curriculum Vitae

Jennifer Kamorowski was born on December 1, 1974 in New Hampshire (United States). She obtained a Bachelor of Science in Behavioral Science from Granite State College in New Hampshire in June 2005. She completed a Juris Doctor at the Massachusetts School of Law in June 2010. In December 2018, she received a Master of Arts in Criminology, Law and Society from George Mason University in Virginia. She began her PhD with the House of Legal Psychology in September 2017. Her PhD research was conducted at Maastricht University under the supervision of Corine de Ruiter, Marko Jelíćic and Maartje Schreuder and at Gothenburg University under the supervision of Karl Ask. She spent six months in Gothenburg for her mobility period. She submitted her thesis to Gothenburg University in June 2020, followed by Maastricht University. In August 2020, after completing her research and returning to the United States, she was appointed as an Assistant Professor of Criminal Justice at Plymouth State University in New Hampshire.

CV

Appendices

List of Appendices

Appendix A: Case vignette (Chapter 3)

Appendix B: HCR-20^{v3} Brief “Refresher” for Participants (Chapter 3)

Appendix C: Case vignette (Chapters 4–6)

Appendix D: Description of SVP civil commitment (Chapters 4–6)

Appendix E: Static-99R report (Chapters 4, 6)

Appendix F: Static-99R report (Chapter 5)

Appendix G: Supplemental materials (Chapter 4)

Appendix H: Supplemental materials (Chapter 5)

Appendix I: Supplemental materials (Chapter 6)

Appendix A

Case Vignette – Chapter 3: HCR-20 Study

Case Background of Mr. S.

Mr. S. was aged 19 in February 2014, when he drove 1800 miles from [City A] to the family's [City B] luxury home and opened fire while they slept.

In the summer of 2016, you are asked to perform a violence risk assessment on Mr. S., using the HCR-20 Version 3. You are asked to complete the HCR-20 coding for presence and relevance and arrive at a final judgment of Mr. S.'s risk of future violence, when he is released into the community after serving his sentence, without further supervision or intervention. Please read the case material provided for more information.

Index offense:

Mr. S. was 19 and a college student when he drove 18 hours straight from [City A] to his family's luxury home in [City B] on February 9, 2014. Around 2 am, Mr. S. shot and killed his sleeping parents, father B. (age 57) and mother C. (age 54), in their beds in the six-bedroom house before shooting his sleeping brother, D. (age 8), and firing at his also sleeping 17-year-old sister, E., who wasn't hit. D. is now paralyzed from the chest down. Another sister, 15-year-old F., who was sleeping in her room with three dogs, was spared an attack. Older brother G., then 21, was not home. Despite his supposed plan to kill himself, Mr. S. made no attempt to do so. He returned to his car out front, and drove back to the parking lot, where he called a cab to take him to the local airport. Police believe he'd bought an American Airlines ticket on his drive to the airport. He was back in Seattle by 10 a.m.

In the weeks following the attack, Mr. S. acted shocked and hysterical, and spent each day by his brother's side in hospital. "He would cry and say, 'D. is eight years old and he doesn't have a dad.'" S.V., a former girlfriend, observed. "He'd say, 'I can't believe somebody killed my parents.'" Two weeks after the shooting, Mr. S. and his brother G. filed for custody of their adopted siblings, D. and F.

Mr. S., who had no previous run-ins with the law, was arrested on March 6, 2014 after police determined he was the 'lone suspect' in the murder. Police found a semi-automatic weapon in his car and phone records placed him in the area of the killings around the time.

Motive for the offense:

Initially, Mr. S. told authorities: "I don't have a reason why, just a lot of problems." While Mr. S.'s motive initially focused on the teenager wanting to inherit his parents' estate, it eventually

settled into something more vague and incomprehensible. “I couldn’t even remember,” Mr. S. told interrogators about the night. “It was just a rush... I was not myself. I don’t know. I was something twisted.”

Mr. S. did not explain his motivations for the violence in court in 2015. When asked if he wanted to speak, he shook his head no. To those who knew the young man, his motive also is a mystery. “He never showed any signs of aggression or threatened to hurt someone,” said his former girlfriend, S.V. “He wasn’t that kid seen on TV crime documentaries that secludes himself and doesn’t talk to anyone. He never spoke badly about his family. He never talked about guns. Nothing.”

Biography:

Mr. S. is the second-born son in a family that had a life of financial privilege. Mother, C., did not come from a wealthy family but managed to accumulate wealth due to her savvy business deals. They say she was a real shark when it came to sales. Within six years, C. and B. had four children—first G. and Mr. S., and then the girls, E. and H.

On April 28, 1999, the family first began to fall apart. That day—with B. out of town on business—C. went to work and left the two youngest in the care of their 28-year-old nanny. About 10 a.m., the nanny was cooking in the kitchen when she saw 2-year-old H. outside on the deck of the spa. She grabbed 16-month-old E. from her highchair and went out to get the older sister. But at the deck, the nanny told investigators, H. “began fussing, kicking, and wiggle” and knocked E., whom the nanny had set on the deck, into the spa.

The panic-stricken nanny ran to get her husband, who was upstairs. He attempted made attempts at CPR as the nanny called C. at work. C., in hysterics, told her to call 911, but the emergency response was delayed because the dispatcher could not understand the nanny, who spoke only Spanish. Paramedics tried to revive the child, but she was pronounced dead. Police considered the death an accident, the coroner’s investigator reported.

The tragedy devastated the family. Within six months, B. filed for divorce, accusing his wife in a court declaration of being mentally unbalanced, of having “some psychological problem that results in manic sessions that are unpredictable, unfathomable, and incapable of being rationalized.” C., for her part, described B. as a neglectful parent interested mostly in golf, surfing, playing the drums, and their business. She also complained in a January 2000 declaration that “to this day, we have had no counseling as a family, and we have never dealt with the terrible tragedy of E.’s death and the irreparable effect it has had on all of our lives.”

The marriage further crumbled against the backdrop of a stressful business dispute. While he cited her “destructive raging,” she said in court papers that she “felt betrayed because ... [he] always told me that he and I would be together in the business.”

In February 2000, a judge issued a divorce judgment after the couple reached a property settlement. With a custody trial pending, C. applied for a temporary restraining order against B. in March 2000, alleging he had assaulted her in front of their children. But within two weeks, they reconciled.

Some nine years after those days of tumult and heartbreak, B. and C. celebrated their new San Juan Capistrano home with family members and friends. The couple wanted to keep the family together because “their children meant everything to them.”

According to some collateral informants, B. and C. were absent a lot, and too busy to discipline their children. Relatives described Mr. S. as a funny, caring, and happy child. They could not believe this sweet and special kid was the perpetrator of this crime. In 2007, his parents adopted two children from a Russian orphanage.

In the summer of 2013, Mr. S. attempted suicide (taking an overdose of pills) after a bad break-up with his girlfriend. His mother was concerned about the amount of time he spent playing League of Legends and World of Warcraft video games. Mr. S. graduated with honors from high school.

During the fall semester of college, Mr. S. had some heated exchanges with his mother. Two days before the murders, Mr. S.’s mother complained that Mr. S. had forgotten his father’s birthday the previous day. Subsequently, Mr. S. texted her: “I forgot his birthday just as much as he forgot he has a son.”

Mr. S. would later tell police that his parents’ reaction to his suicide attempt was surprisingly casual: “They really didn’t take it serious or acknowledge [it].” Instead of getting him some help, his parents sent him off to Washington State to attend college. Mr. S. returned home for the Christmas holidays. On January 3, he and his mother met her good friend X. for lunch. “He was really positive about his future,” C’s childhood friend says. “It didn’t seem like he was up against a wall, that he was desperate.” Still, there are other informants, who report that Mr. S. was not attending his college classes, playing a lot of videogames and smoking weed.

Appendix B

HCR-20^{v3} “Refresher” – Chapter 3: HCR-20 Study

QUICK GUIDELINE: HCR-20 V.3 ADMINISTRATION PROCESS

This brief guideline is meant for those participants that may need to refresh their knowledge about the administration of the HCR-20. Although we recommend reading the guidelines, you can skip it if you are confident that you are familiar enough with the scale.

PRESENCE OF A RISK FACTOR

As you revise the case information your first task is to carefully rate the PRESENCE of each risk factor in each subscale of the HCR-20. Based on the information you have about the case, you will select one of the 3 available boxes (*No, Possibly or Partially, and Yes*) for the presence of each factor. Try to think of evidence that confirms and disconfirms a risk factor to ensure your evaluation is balanced and to make it easier for you to later justify your decision.

In case you don't have information available for a risk factor you can select the box under the *Omit* category. Please, note that this is not the same as the *No* category. To omit a risk factor, you must conclude you don't have the necessary information to make relevant observations, whereas to mark a risk factor with *No*, you simply judge that the absence of information as reliable (i.e., the case material shows no history of drug abuse and the subject consistently denies substance misuse). As a general rule, it is recommended to use the *Omit* option sparingly because omitting risk factors can negatively affect the effectiveness of risk assessment.

The HISTORICAL SUBSCALE refers to past behaviours and experiences. Risk factors in this subscale can be grouped into 3 categories: problems in adjustment or living (H3, H4, H8), problems with mental health (H5, H6, H7), and past antisocial issues (H1, H2, H9, H10). You will consider the presence of each of these factors if they are present at any moment in an individual's life.

In the CLINICAL SUBSCALE you will find factors that relate to the current psychosocial, mental health, and behavioural functioning of the evaluatee. You shall rate those areas in accordance with the information you will be provided with.

The RISK MANAGEMENT SUBSCALE is focused on future psychosocial adjustment, living situation, and use of risk management strategies.

RELEVANCE OF RISK FACTORS

After you have rated the presence of each risk factor it's time for scoring their relevance. Risk factors can be more or less relevant depending on the examinee because they don't necessarily contribute equally to general risk in every person. For this task, you should consider the PRESENCE of a risk factor as a feature that potentially could elevate the risk for THIS PARTICULAR CASE. For instance, let's say you rate the factor "violent ideation" as definitely present (you marked the box "yes"), then for the relevance rating you want to consider to which extent this particular factor, in your opinion, increases the risk for this particular individual. To determine if this factor can elevate the level of risk, you might think about the role it has played in past violence or may play in future violence.

For the relevance, you will consider only those factors that you scored as present or partially present. You will decide to which extent a factor is relevant by asking yourself these questions:

Did it contribute to the evaluatee's past violence?

Is it likely to make the person act violently in the future?

Is it likely to impair the evaluatee's non-violent problem-solving skills in the future?

Is it crucial to manage through professional intervention to reduce general risk?

You shall rate the relevance of a PRESENT risk factor by marking it as *Low*, *Moderate*, or *High*.

Appendix C

Case Vignette –Chapters 4–6: SVP Studies

Positive and Negative Profile Information for John Smith

Negative condition background information: John Smith, aged 45, worked as a janitor in an office building for two months. Mr. Smith's co-workers reported that he mostly kept to himself and they did not know him well. Some co-workers also reported that they sometimes had to do extra work because Mr. Smith did not complete his duties. Mr. Smith was not known by many people in his community and was described by his neighbors as "not very sociable." He rarely spoke to his neighbors, and some described Mr. Smith as "odd." In his free time, Mr. Smith liked to watch television or go to online chat rooms. He also enjoyed playing games online. Mr. Smith has been married and divorced twice. He has two children, who are aged 15 and 10. Mr. Smith is not very involved in the lives of his children due to a very bitter divorce from their mother (his first wife).

Negative condition offense information: John Smith, was convicted for raping a 35-year-old woman with whom he was acquainted. He is now eligible for release from prison after serving a 5-year prison sentence for the offense. Mr. Smith also has one previous criminal conviction – at the age of 18 he got into a fight at a bar and pleaded guilty to the charge of assault. He was sentenced to perform 240 hours of community service as a result, which he successfully completed. Most of the local news media outlets reported about the rape. Many community members were shocked and outraged by the crime, and most people believed Mr. Smith's prison sentence of 5 years was insufficient. In fact, the victim of the crime was quoted by the local media as stating, "I was very hurt by what Mr. Smith did to me, and I am not satisfied that the prison sentence he was given is fair. I don't know how I can move on with my life."

Positive condition background information: John Smith, aged 45, worked as the manager of a local bank, where he was employed for 20 years. Mr. Smith was friendly with everyone and was known as a good worker. Mr. Smith was known by many people in his community and was described by his neighbors as "very sociable." He always said hello to his neighbors, and most described Mr. Smith as "a nice guy." In his free time, Mr. Smith coached his children's baseball teams and regularly volunteered at a local homeless shelter. Mr. Smith is also a talented musician, and often played the organ at the church where he was a member. Mr. Smith and his ex-wife met in high school and eventually married and lived in the same community in which they both grew up. Mr. Smith has two children, who are aged 15 and 10. Mr. Smith was married for 15 years before his wife filed for divorce.

Positive condition offense information: John Smith, aged 45, was convicted for raping a 35-year-old woman with whom he was acquainted. He is now eligible for release from prison after serving a 5-year prison sentence for the offense. Mr. Smith also has one previous criminal conviction – at the age of 18 he got into a fight at a bar and pleaded guilty to the charge of assault. He was sentenced to perform 240 hours of community service as a result, which he successfully completed. Most of the local news media outlets reported about the rape. Many community members were shocked by the crime, but most people believed that Mr. Smith's prison sentence of 5 years was just. In fact, the victim of the crime was quoted by the local media as stating, "Although I was very hurt by what Mr. Smith did to me, I am satisfied that the prison sentence he was given is fair. I can now move on with my life."

Appendix D

Explanation of Civil Commitment – Chapters 4–6: SVP Studies

Background Information Related to Sexual Offender Civil Commitment

NOTICE: You have been selected to be a juror in a civil commitment hearing. Mr. Smith is eligible to be released from prison because he has served his 5-year sentence. However, Mr. Smith must have a hearing to decide whether he should be ordered (by a judge) to be confined to a locked facility for sexual offender treatment (civil commitment). Civil commitment will be explained in further detail on the next three pages. You as the juror must decide two things:

- 1) how likely you think it is that Mr. Smith will commit another sexual offense and
- 2) whether Mr. Smith should be ordered to civil commitment.

You should base your decision on all of the information you think is relevant to these questions.

What is sex offender civil commitment?

Twenty American states and the federal government have enacted civil commitment laws for individuals who meet the legal criteria of a “sexually violent predator” (SVP) or “sexually dangerous person” (SDP).

Civil commitment under SVP or SDP laws permits individuals convicted of a sexual offense to be confined indefinitely in a secure (locked) hospital or treatment facility.

During his or her time in civil commitment, an individual is expected to undergo treatment until it is determined that s/he can safely be released to the community..

Who is recommended for sex offender civil commitment?

Confinement in a secure facility under SVP laws most often occurs after a person has served his or her criminal (prison) sentence.

Several months before the person is scheduled to be released from prison, an evaluation is conducted by a forensic mental health specialist (for example a psychologist or psychiatrist). A recommendation is made about whether the person should be considered a candidate for civil commitment.

Because confinement in civil commitment involves taking away a person’s freedom, he or she is entitled to a court hearing.

At the hearing, the district attorney (prosecutor) must present evidence that is sufficient to meet certain legal requirements before a judge can order that a person be held under civil commitment laws.

The legal process of civil commitment

A court hearing - often referred to as a civil commitment hearing - is therefore necessary to determine if the person meets the legal requirements for civil commitment.

A judge or jury decides whether the person meets the legal criteria for sex offender civil commitment.

Generally, four criteria must be met for civil commitment as an SVP, as follows:

- 1) a history of sexual offending
- 2) a mental abnormality (sometimes referred to as a mental disorder, personality disorder, or “behavioral abnormality”)
- 3) an impairment in his ability to control his sexual behavior, and
- 4) likelihood of future sexual offending.

Appendix E

Static-99R Report – Chapter 4: NFC Study and Chapter 6: PSO Study

Background Information on Static-99R Development and Report for John Smith

Please read the following background information about the use of risk assessment tools to estimate the likelihood of sexual reoffending. Forensic psychologists or psychiatrists frequently use a risk assessment tool(s) in their evaluation to estimate the likelihood that the offender will commit another sexual offense. Risk assessment tools commonly used by evaluators include a variety of risk factors. The Static-99R is the most widely used instrument to estimate a sexual offender's risk for committing a new sex offense. The Static-99R was developed by studying thousands of sexual offenders from Canada, Europe, the United States and Australia.

The researchers who studied these offenders coded their scores on the Static-99R and observed the rates of re-offense among sex offenders with each score. Now, evaluators who use the Static-99R can examine a particular sex offender's score and then check to see how common it was for offenders with that same score to commit a new sex offense based on previous studies.

How is the score calculated using the Static-99R?

The Static-99R is a checklist of 10 risk factors that evaluators code based on a sex offender's official records. Each item is assigned a number (which is usually 0 to 3, although offenders who are older than 40 can be scored as a -1 or a -3). A table of the items in the Static-99R is pictured below. The risk factors contained in the Static-99R are associated with a numerical score, and the total score is obtained by adding up the ratings for each risk factor. This final score is associated with an estimated likelihood of sexual reoffending, which the evaluator can then use to estimate an individual's likelihood of committing another sexual offense.

You will now be provided with additional information regarding Mr. Smith. The following information was derived from Mr. Smith's score on the Static-99R.

Reminder: The Static-99R is a risk assessment tool that estimates how likely each offender is to commit a new sex offense.

Static-99R results for Mr. Smith

In routine samples of sexual offenders, the average 5-year sexual reoffending rate is between 5% and 15%. This means that out of 100 sexual offenders of varied risk levels, between 5 and 15 would be charged or convicted for a new sexual offense after 5 years in the community. Conversely, between 85 and 95 would not be charged or convicted for a new sexual offense during that time period. Mr. Smith's Static-99R score was **2**. This score falls in the “**average risk**” category. In routine samples of sex offenders with the same score, the 5-year sexual reoffending rate is between 7% and 8.8%. This means that out of 100 sexual offenders with the same risk score between 7 and 9 would be charged or convicted for a new sexual offense after 5 years in the community. Conversely, between 91 and 93 would not be charged or convicted for a new sexual offense during that time period. The above values are based on the table entitled “Static-99R Routine Sample: Estimated 5-year Sexual Recidivism Rates” in Phenix, Helmus & Hanson (October 19, 2016) Static-99R & Static-2002R Evaluators' Workbook. Available from www.static99.org.

**Please note, the Static-99R does not measure all potentially relevant risk factors and Mr. Smith's reoffending risk may be higher or lower than that indicated by the Static-99R based on factors not included in this risk assessment tool.*

Appendix F

Static-99R Report – Chapter 5: Asymmetrical Skepticism Study

Background Information on Static-99R Development and Report for John Smith for “Low Risk” and “High Risk” Conditions

Please read the following background information about the use of risk assessment tools to estimate the likelihood of sexual reoffending. Forensic psychologists or psychiatrists frequently use a risk assessment tool(s) in their evaluation to estimate the likelihood that the offender will commit another sexual offense. Risk assessment tools commonly used by evaluators include a variety of risk factors. The Static-99R is the most widely used instrument to estimate a sexual offender’s risk for committing a new sex offense. The Static-99R was developed by studying thousands of sexual offenders from Canada, Europe, the United States and Australia.

The researchers who studied these offenders coded their scores on the Static-99R and observed the rates of re-offense among sex offenders with each score. Now, evaluators who use the Static-99R can examine a particular sex offender’s score and then check to see how common it was for offenders with that same score to commit a new sex offense based on previous studies.

How is the score calculated using the Static-99R?

The Static-99R is a checklist of 10 risk factors that evaluators code based on a sex offender’s official records. Each item is assigned a number (which is usually 0 to 3, although offenders who are older than 40 can be scored as a -1 or a -3). A table of the items in the Static-99R is pictured below.

Risk Factors Assessed with the Static-99R

Items of the Static-99R

-
1. Age at release from index offense
 2. Ever lived with a lover (at least 2 years)
 3. Index non-sexual violence
 4. Prior non-sexual violence
 5. Prior sex offenses
 6. Four or more prior sentencing dates (excluding index)
 7. Any convictions for non-contact sex offenses
 8. Any unrelated victims
 9. Any stranger victims
 10. Any male victims
-

Note. Adapted from (Phenix, Helmus, et al., 2016). Used with permission.

The risk factors contained in the Static-99R are associated with a numerical score, and the total score is obtained by adding up the ratings for each risk factor. This final score is associated with an estimated likelihood of sexual reoffending, which the evaluator can then use to estimate an individual's likelihood of committing another sexual offense.

You will now be provided with additional information regarding Mr. Smith. The following information was derived from Mr. Smith's score on the Static-99R.

Reminder: The Static-99R is a risk assessment tool that estimates how likely each offender is to commit a new sex offense.

Static-99R results for Mr. Smith¹³

In routine samples of sexual offenders, the average 5-year sexual reoffending rate is between 5% and 15%. This means that out of 100 sexual offenders of varied risk levels, between 5 and 15 would be charged or convicted for a new sexual offense after 5 years in the community. Conversely, between 85 and 95 would not be charged or convicted for a new sexual offense during that time period.

Mr. Smith's Static-99R score was **0 / 4**. This score falls in the “**below average**” / “**above average risk**” category. In routine samples of sex offenders with the same score, the 5-year sexual reoffending rate is between **3.3% and 4.7% / 10% and 12.1%**. This means that out of 100 sexual offenders with the same risk score between **3 and 5 / 10 and 12** would be charged or convicted for a new sexual offense after 5 years in the community. Conversely, between **95 and 97 / 88 and 90** would not be charged or convicted for a new sexual offense during that time period. The above values are based on the table entitled “Static-99R Routine Sample: Estimated 5-year Sexual Recidivism Rates” in Phenix, Helmus & Hanson (October 19, 2016) Static-99R & Static-2002R Evaluators’ Workbook. Available from www.static99.org.

**Please note, the Static-99R does not measure all potentially relevant risk factors and Mr. Smith's reoffending risk may be higher or lower than that indicated by the Static-99R based on factors not included in this risk assessment tool.*

¹³ Note. The factors that were varied between the “low risk” and “high risk” Static-99R conditions appear in bold, with the “low risk” variant preceding the slash and the “high risk” variant presented after.

Appendix G

Supplemental Materials - Chapter 4: Need for Cognition Study

H1 Analysis Including Participants Who Failed One or Both Attention Checks

As a check on the effect of excluding participants who failed one or both attention checks, here we report the results of our analyses that includes the 427 participants included in the main analyses and the additional 171 participants who were excluded for failing one or both attention checks ($N = 598$). Overall, the general risk estimates ($M = 66.6\%$, $SD = 20.3$) were much higher than actual sexual recidivism rates of 5-15% (Alper & Durose, 2019; Hanson et al., 2014; Hanson & Morton-Bourgon, 2005, 2009; A. J. Harris & Hanson, 2004; Helmus et al., 2012; Phenix, Helmus, et al., 2016). An independent samples t -test confirmed that there was no significant difference between conditions with respect to participants' general risk estimates, $t(596) = -1.11$, $p = .268$, Hedges' $g = 0.09$, 95% CI [-0.07, 0.25]. Nevertheless, to isolate the effects of the contextual information, we controlled for participants' general risk estimates in the analysis of T1 risk estimates.

A one-way ANCOVA was conducted on participants' ratings of the respondent's recidivism risk at T1, with contextual information (positive vs. negative) as the independent variable and participants' general estimates of sexual recidivism risk as covariate. There was a significant, medium-sized effect of contextual information, $F(1, 595) = 50.08$, $p < .001$, $\eta^2_p = .078$, 90% CI [.046,.114]. In support of H1, participants in the positive contextual information condition rated the recidivism risk as significantly lower ($M = 44.8$, $SD = 26.1$) than did participants in the negative contextual information condition ($M = 57.9$, $SD = 22.6$). Moreover, the analysis indicated that participants' general recidivism risk estimates accounted for a large and significant portion of the variance in T1 risk estimates, $F(1, 595) = 240.2$, $p < .001$, $\eta^2_p = .288$, 90% CI [.239,.334]. As expected, participants who reported higher general risk estimates reported higher case-specific risk estimates at T1, bivariate $r = .53$, $p < .001$.

Results of H2 and H3 Hypothesis Testing with T2 Raw Data Bootstrapped Confidence Intervals (Main Sample)

We are aware of the problems that can be introduced into hypothesis testing when adding a constant to the original values and/or using a log-transformation of the data (Feng et al., 2014). Therefore, to exclude the possibility that adding a constant of one and log-transforming the T2 risk estimates substantially altered our conclusions in the regression analysis for H2

and H₃, we conducted a moderation analysis on the raw T₂ data to compute bootstrapped confidence intervals using the PROCESS v. 3.5 macro in SPSS (Hayes, 2018). This analysis did not change the nature of the results. Bootstrapped results of the full model are presented in Table S1.

Table S1 Bootstrap Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' Raw T₂ Risk Estimates

Predictor	b	95% CI	SE _b	p
Step 1 ($R^2 = .122^{***}$)				
General risk estimate	0.406	[0.287, 0.522]	0.060	< .001
NFC	-0.016	[-0.189, 0.153]	0.088	.869
Contextual information	3.697	[-0.543, 8.219]	2.236	.098
NFC × Contextual Information	-0.163	[-0.346, 0.155]	0.127	.513

Note. N = 427. Contextual information condition was dummy-coded (0 = positive, 1 = negative). NFC = Need for Cognition scale. Bootstrapped results are based on 5,000 samples and a 95% confidence interval. Final model: $F(4, 422) = 14.68, p < .001, R^2_{\text{Adj}} = .122$.

Results of H₂ and H₃ Hypothesis Testing with T₂ Raw Data Bootstrapped Confidence Intervals (Total Sample)

We conducted and reported an exploratory analysis of H₂ and H₃ in the main paper utilizing the entire sample that completed the study and were not excluded for geocoordinate issues (N = 598). To address concerns that adding a constant and log-transforming the T₂ risk estimates introduced problems in the regression analysis (Feng et al., 2014), we again conducted a moderation analysis on the raw T₂ data to compute bootstrapped confidence intervals using the PROCESS v. 3.5 macro in SPSS (Hayes, 2018). This analysis did not change the nature of the results. Bootstrapped results of the full model are presented in Table S2.

Table S2 Bootstrap Results of Hierarchical Multiple Linear Regression Analysis Predicting Participants' Raw T₂ Risk Estimates (Full Sample)

Predictor	b	95% CI	SE _b	p
Step 1 ($R^2 = .122^{***}$)				
General risk estimate	0.441	[0.332, 0.555]	0.057	< .001
NFC	-0.221	[-1.356, -0.047]	0.095	.025
Contextual information	2.832	[-0.543, 6.950]	2.110	.189
NFC × Contextual Information	0.010	[-0.253, 0.277]	0.134	.943

Note. N = 598. Contextual information condition was dummy-coded (0 = positive, 1 = negative). NFC = Need for Cognition scale. Bootstrapped results are based on 5,000 samples and a 95% confidence interval. Final model: $F(4, 594) = 20.54, p < .001, R^2_{\text{Adj}} = .122$.

Exploratory Analyses: Contextual Information Effects

Contextual Information and Static-99R Effects on Recidivism Risk Estimates

Participants' risk estimates based on the case vignette decreased significantly from T1 to T2 ($M = 48.0\%$, $SD = 25.0$ and $M = 25.6\%$, $SD = 24.4$, respectively), $t(426) = 20.72$, $p < .001$, Hedges' $g_{rm} = 0.91$, 95% CI [0.80, 1.01], suggesting a large overall effect of the Static-99R information in reducing recidivism risk estimates. Interestingly, a Welch's t -test indicates a significantly greater reduction in risk estimates from T1 to T2 between the positive ($M_{raw} = -17.8$, $SD = 20.3$) and negative ($M_{raw} = -27.2$, $SD = 23.4$) contextual information condition, $t(417.24) = 4.44$, $p < .001$, Hedges' $g = 0.43$, 95% CI [0.24, 0.62]. However, the 5-year recidivism risk estimate indicated in the Static-99R information for offenders whose risk score is a 2 (as was the case for Mr. Smith) is estimated to lie between 7% and 8.8%. Nevertheless, many participants estimated the risk as substantially higher than the Static-99R information and only about 23.0% provided a risk estimate that was within the range indicated by the Static-99R.

Notably, participants in the negative contextual information condition were particularly likely to report risk estimates radically higher than the Static-99R. A chi-square test of independence revealed a significant difference in the likelihood between contextual information conditions of providing a T2 risk estimate that was higher than that indicated by the Static-99R, $\chi^2(1) = 6.99$, $p = .008$, Cramer's $V = .13$ (a small-to-medium effect). Post-hoc analyses (Bonferroni-corrected) of adjusted residuals showed that participants who reviewed negative information about the respondent were significantly more likely at T2 to overestimate the respondent's risk for sexual recidivism (56.8%) than were participants who reviewed positive information about the respondent (69.2%), $p < .05$. This finding suggests that despite having access to risk-relevant Static-99R information, participants who were initially exposed to negative information about the SVP respondent evidenced the effects of this bias in their estimates of the respondent's sexual recidivism risk at T2.

Contextual Information and Importance Ratings of Factors to T2 Risk Estimate

Interestingly, we found no significant differences between the negative and the positive contextual information groups in the mean ratings they assigned to the importance of either the risk-irrelevant information (i.e., the SVP respondent's relationships, employment history, prison sentence, or the impact of the crime on the victim), the respondent's criminal history, or the Static-99R with respect to the risk estimates they provided at T2 (Table S3). This finding

is suggestive of the operation of unconscious bias and humans' lack of awareness about what factors influence their decision-making (Wilson & Brekke, 1994).

Table S3 Positive and Negative Information Mean Group Importance Ratings to T2 Risk Estimate

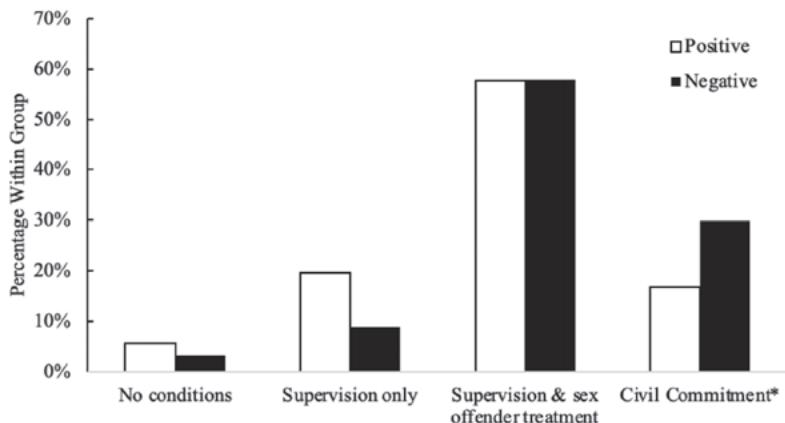
Factor	Positive		Negative		<i>t(425)</i>	<i>p</i>	Hedges'g 95% CI
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Relationships with others	3.26	1.13	3.16	1.12	0.913	.362	0.09 [-0.10, 0.28]
Employment history	2.53	1.15	2.41	1.27	1.02	.310	0.10 [-0.09, 0.29]
Criminal history	3.92	1.01	4.07	0.92	-1.66	.098	0.16 [-0.03, 0.35]
Impact on victim	3.57	1.24	3.39	1.31	1.46	.146	0.14 [-0.05, 0.33]
Length prison sentence	3.36	1.14	3.38	1.23	-0.148	.882	0.02 [-0.17, 0.21]
Static-99R information	3.64	0.96	3.79	0.99	-1.50	.135	0.15 [-0.04, 0.34]

Note. Positive *n* = 213. Negative *n* = 214.

Contextual Information Effects on Disposition Recommendations

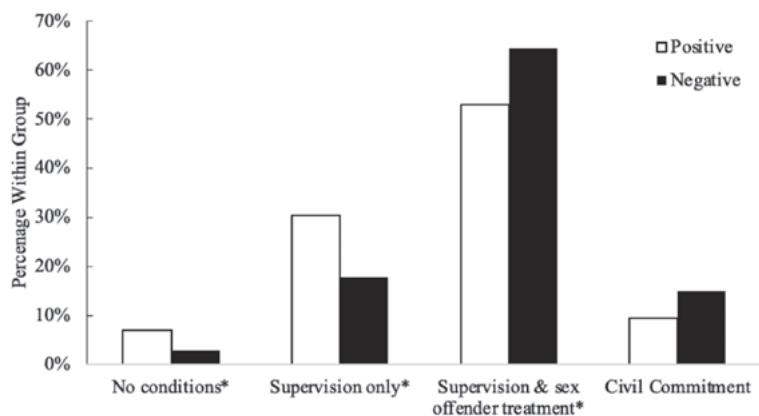
Participants' disposition recommendations at T1 and T2 are shown in Figure 1 and Figure 2, respectively. At T1, nearly a quarter of all participants (*n* = 100; 23.4%) recommended civil commitment for Mr. Smith, whereas over half of participants (*n* = 247; 57.8%) recommended release to the community with supervision and sexual offender treatment. At T2, the number of participants who recommended civil commitment was reduced by nearly half (*n* = 52; 12.2%), whereas the number of participants recommending supervision and sex offender treatment was relatively constant (*n* = 251; 58.8%). The number of participants recommending supervision only increased slightly from T1 (*n* = 61; 14.3%) to T2 (*n* = 103; 24.1%).

Figure S1 Disposition Recommendations by Contextual Information Condition at T1



Note. An asterisk denotes a significant difference between groups at the $p < .05$ level.

Figure S2 Disposition Recommendations by Contextual Information Condition at T2



Note. An asterisk denotes a significant difference between groups at the $p < .05$ level.

To examine whether the contextual information manipulation influenced disposition recommendations at T1 and T2, two chi-square tests of independence were conducted. The distribution of disposition recommendations at T1 differed significantly between the contextual information conditions, $\chi^2(3) = 17.83, p < .001$, Cramer's $V = .20$ (a medium effect). Post-hoc analyses (Bonferroni-corrected) of adjusted residuals showed that participants in the positive contextual information condition were significantly more likely to recommend probation only, and significantly less likely to recommend civil commitment at T1, than were participants in the negative contextual information condition, both $ps < .05$.

The distribution of disposition recommendations at T2 also differed significantly as a function of contextual information, $\chi^2(3) = 16.19, p < .001$, Cramer's $V = .20$ (a medium effect). Post-hoc analyses (Bonferroni-corrected) of adjusted residuals showed that participants in the positive (vs. negative) condition significantly more often recommended release to the community with no conditions or with probation only, all $p < .05$.

Furthermore, participants in the negative condition were significantly more likely to recommend supervision and sexual offender treatment than participants in the positive condition. However, there was no significant difference in the likelihood that participants in the negative condition would recommend civil commitment at T2 compared to the positive condition. This finding suggests that the Static-99R information did mitigate the effects of the risk-irrelevant contextual information with respect to disposition recommendations, but that participants in the negative condition were nevertheless significantly more inclined to choose the more restrictive option of supervision and sexual offender treatment than were participants in the positive condition.

We did not make a prediction about how risk-irrelevant contextual information would affect mock jurors' disposition recommendations, but the results of our exploratory analysis indicated that before reviewing the risk-relevant Static-99R information, participants in the negative condition were significantly more likely to recommend civil commitment than were participants in the positive information condition.

Contextual Information Effects on Importance of Factors to T2 Disposition Recommendation

Participants rated how important the following factors were to their T2 disposition recommendation: the SVP respondent's relationships, employment history, criminal history, and prison sentence; the impact of the crime on the victim; the participant's own estimate of the likelihood of respondent sexually reoffending; and the Static-99R. An independent samples t -test revealed that participants in the positive (vs. negative) information condition rated the respondent's relationships as significantly more important to their T2 disposition recommendation ($M = 3.31, SD = 1.19$ and $M = 2.98, SD = 1.26$, respectively), $t(425) = 2.73, p = .007$, Hedges' $g = 0.27$, 95% CI [0.079, 0.460]. Participants in the positive condition also rated the respondent's employment history as significantly more important to their decision than did participants in the negative condition ($M = 2.74, SD = 1.26$ and $M = 2.48, SD = 1.31$, respectively), $t(425) = 2.06, p = .040$, Hedges' $g = 0.20$, 95% CI [0.012, 0.393]. One possible explanation for these findings is that participants in the positive condition, who were also more likely to recommend a less restrictive disposition option of no conditions or probation only, viewed previous employment and relationships with others as protective factors against

the risk of sexual reoffending. There were no significant differences between contextual information conditions with respect to their importance ratings of criminal history, impact of the crime on the victim, their own T2 estimate of the respondent's sexual recidivism risk, the length of the prison sentence, nor the Static-99R on their T2 disposition recommendation.

Exploratory Analyses: Responders Versus Overstimators

We conducted additional analyses to examine factors that might differentiate mock jurors who overestimated reoffending risk at T2 and voted to commit the respondent after reviewing Static-99R information from mock jurors who did not show this pattern of overestimation. We defined two groups for these analyses: the *responder* group comprises participants whose T2 risk estimate was within or below the Static-99R range ($n = 158$; 37.0%) and the *overestimator group*, (i.e., participants whose T2 risk estimate was higher than the Static-99R range; $n = 269$; 63.0%). A Welch's t -test revealed a large, significant difference in the mean T2 risk estimates between the groups, $t(272.9) = -19.9$, $p < .001$, Hedges' $g_{rm} = 1.53$, 95% CI [1.31, 1.75], with the overall T2 risk estimates of the overestimator group being significantly higher than those of the responder group ($M_{raw} = 36.7$, $SD = 24.7$ and $M_{raw} = 6.6$, $SD = 1.8$, respectively).

Importance of Risk-Irrelevant and Risk-Relevant Factors to Overestimating Risk at T2

A separate independent samples t -test revealed no significant difference between the responder and overestimator groups with respect to their importance ratings of the SVP respondent's criminal history. Furthermore, although the mean rating of the importance of the Static-99R was higher for the responder group than the overestimator group, this difference was not statistically significant (Table S4). These findings indicate that the risk-relevant information was rated as approximately equally important, regardless of whether participants overestimated the risk or not.

However, the overestimator group rated the SVP respondent's relationships with others, his employment history, the impact of the crime on the victim, and the prison sentence as significantly more important to their T2 risk estimate than did the responder group. The effect sizes for each of these differences is small to medium. These findings indicate that participants who overestimated risk at T2 rated several risk-irrelevant factors as significantly more important to their risk estimate compared to participants who did not overestimate the recidivism risk.

Table S4 Responder and Overestimator Mean Importance Ratings of Factors to T2 Risk Estimate

Factor	Responders		Overestimators		t	p	Hedges'g
	M	SD	M	SD			95% CI
Relationships with others	3.06	1.14	3.30	1.11	t(425) 2.14	.033	0.21 [0.02, 0.41]
Employment history	2.26	1.04	2.59	1.29	t(386.07) 2.94	.003	0.27 [0.08, 0.47]
Criminal history	4.01	0.86	3.99	1.02	t(373.59) -0.229	.819	0.02 [-0.18, 0.22]
Impact on victim	3.12	1.34	3.69	1.20	t(300.50) 4.40	< .001	0.45 [0.26, 0.65]
Length prison sentence	3.13	1.10	3.51	1.21	t(354.50) 3.38	.001	0.32 [0.13, 0.52]
Static-99R information	3.79	0.89	3.67	1.03	t(367.74) -1.29	.198	0.12 [-0.07, 0.32]

Note. Responders $n = 158$. Overestimators $n = 269$.

Effects of Overestimating Risk on T2 Disposition Recommendations

To examine the effect of overestimating risk at T2 on disposition recommendations, we conducted a chi-square test of independence to test whether participants whose T2 risk estimate fell within or below the Static-99R range and the overestimator group differed with respect to their disposition recommendations. The analysis revealed a significant difference in T2 disposition recommendations as a function of group membership, $\chi^2(3) = 49.4$, $p < .001$, Cramer's $V = .34$ (a large effect). Adjusted residuals showed that participants in the overestimator group were significantly more likely to recommend civil commitment or probation and treatment than participants whose T2 risk estimate was within or below the Static-99R range, all $ps < .05$. Furthermore, participants whose T2 risk estimate was within or below the range indicated by the Static-99R were significantly more likely to recommend probation only or no conditions than the overestimator group, all $ps < .05$.

Responders Versus Overestimators Importance Ratings of Factors to T2 Disposition Recommendation

Following our previous findings that indicated differences between participants who overestimated risk at T2 and those who did not, we conducted an independent samples t-test to compare the groups' mean importance ratings of the various factors to their T2 disposition recommendations. Overall, the overestimator group rated the impact of the crime on the victim as significantly more important to their disposition recommendation than did the group

that did not overestimate risk at T2 ($M = 3.78$, $SD = 1.14$; $M = 3.23$, $SD = 1.29$, respectively), $t(425) = 4.55$, $p < .001$, 95% CI [0.310, 0.783], Hedges' $g = 0.46$, 95% CI [0.260, 0.657]. Furthermore, the only other significant difference was that the overestimator group rated the prison sentence as significantly more important to their T2 disposition recommendation than did the group that did not overestimate risk at ($M = 3.48$, $SD = 1.16$ and $M = 3.2$, $SD = 1.11$, respectively), $t(425) = 2.44$, $p = .015$, Hedges' $g = 0.24$, 95% CI [0.048, 0.442]. These results, in combination with our previous finding that participants who overestimated risk at T2 also rated the impact of the crime on the victim and the prison sentence of greater importance to their T2 risk estimate than those who did not overestimate risk, indicate a potentially meaningful relationship between these two factors and SVP decision-making.

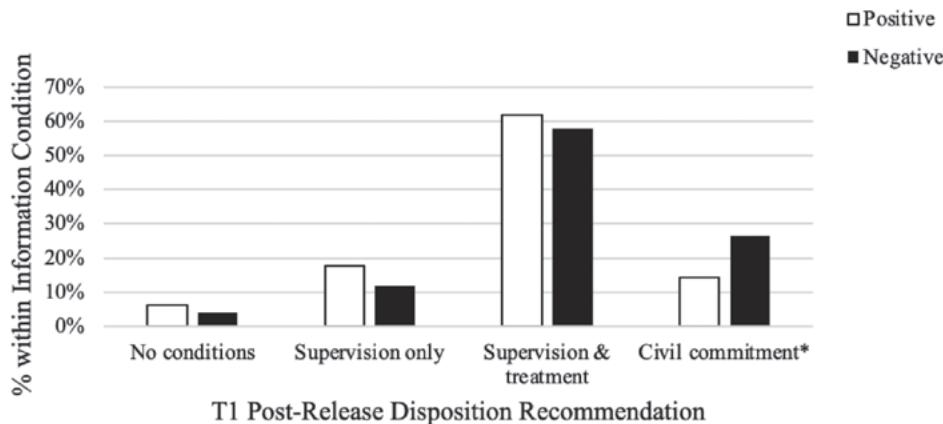
Appendix H

Supplemental Materials - Chapter 5: Asymmetrical Skepticism Study

Exploratory Analysis: Post-Release Disposition Recommendations

A chi-square test of independence showed a significant difference in the distribution of post-release disposition recommendations between the positive and negative contextual information condition at T1, $\chi^2(3) = 9.95, p = .019$, Cramer's $V = .16$ (small to medium effect). Results are presented in Figure 2.

Figure S1 T1 Post-Release Disposition Recommendations by Contextual Information Condition



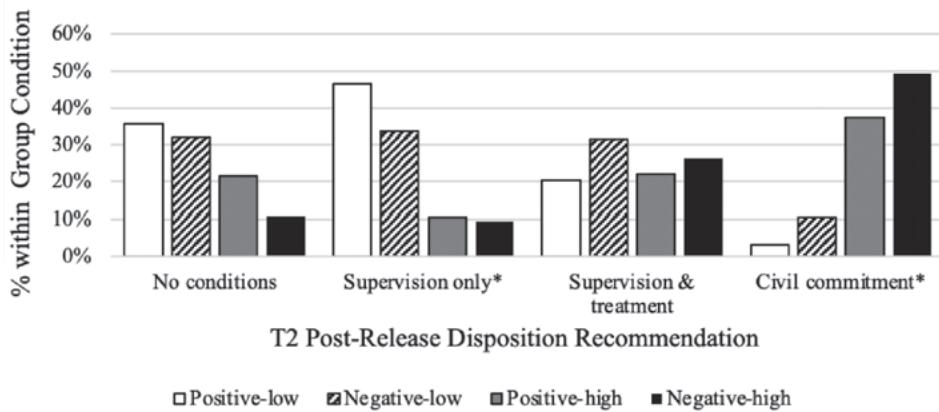
Note. An asterisk denotes a significant difference between groups at the $p < .05$ level.

Examination of the post-hoc analyses of adjusted residuals (Bonferroni-corrected $\alpha = .006$) revealed a significant difference in the proportion of participants in the negative contextual information condition who recommended civil commitment at T1 as compared to the proportion of participants in the positive information condition who recommended civil commitment, $z = 2.88, p = .004$. There were no significant differences between the groups with respect to the proportion recommending other dispositions.

We did not have a preregistered hypothesis about how contextual information and a Static-99R risk estimate taken together would affect participants' ultimate post-prison recommendations at T2. Therefore, we conducted an exploratory examination of this effect.

A chi-square test of independence showed a significant difference in the distribution of post-release disposition recommendations between the groups, $\chi^2(9) = 80.29, p < .001$, Cramer's $V = .27$ (large effect; see Figure S2).

Figure S2 T2 Post-Release Disposition Recommendations by Group Condition



Note. An asterisk denotes a significant difference between groups at the $p < .05$ level.

Post-hoc analyses of adjusted residuals (Bonferroni-corrected $\alpha = .003$) showed that a significantly greater proportion of participants in the positive contextual information and low Static-99R risk estimate group recommended supervision only ($z = 5.4, p < .001$) than the proportion of participants in the positive/high group, ($z = -3.0, p = .003$.) and negative/high group ($z = -3.9, p < .001$). In addition, a significantly greater proportion of participants in both the positive ($z = 3.3, p = .001$) and negative ($z = 5.0, p < .001$) contextual information/high Static-99R risk estimate groups recommended civil commitment than did participants in either contextual information group that received a low Static-99R risk estimate, (positive/low: $z = -4.5, p < .001$; negative/low: $z = -3.6, p < .001$). There was no significant difference between the group proportions that recommended either no conditions or supervision and mandated sex offender treatment.

Exploratory Analyses: Importance of Factors to T2 Civil Commitment Recommendations

Participants were asked to rate how important various factors were to their T2 post-release disposition recommendation on a 5-point Likert scale (1 = *not at all important*, 5 = *extremely important*). The factors were: (a) SVP respondent's relationships with others, (b)

employment history, (c) criminal history, (d) impact of the crime on the victim, I participants' own estimate of SVP respondent's risk of sexual recidivism, (f) length of prison sentence, and (g) Static-99R estimate. Using these ratings as predictors, we conducted a binary logistic regression to examine whether the importance ratings of any of the factors significantly predicted a vote for civil commitment. Results are presented in Table S1.

The sole factor that significantly predicted a vote for civil commitment was the participants' rating of the importance of the impact of the crime on the victim. Each 1-point increase in the participants' rating of importance of this factor increased the likelihood of recommending civil commitment by 1.7 times.

Table S1 Importance Ratings of Factors Predicting T2 Civil Commitment Recommendations

Factor	b	OR	95% CI	p
Relationships with others	-0.08	0.92	[0.70, 1.22]	.572
Employment history	-0.18	0.84	[0.63, 1.11]	.214
Criminal history	0.38	1.47	[0.99, 2.18]	.059
Impact of crime on victim	0.54	1.71	[1.22, 2.41]	.002
Own estimate of sexual recidivism risk	0.21	1.23	[0.90, 1.67]	.190
Length prison sentence	0.26	1.30	[0.93, 1.81]	.121
Static-99R information	0.18	1.19	[0.89, 1.60]	.244

Note. OR = Odds ratio.

$\chi^2(7) = 49.17, p < .001$, Nagelkerke $R^2 = .201$.

Parallel Analyses of T2 Risk Estimates

To investigate the robustness of the results reported in the main analysis, parallel analyses were performed on participants' T2 risk estimates. Table S2 reports the results of generalized linear model (GLM) regression analyses using the gaussian (normal) distribution and identity link function for the raw (untransformed), rank-transformed (as reported in the main analysis), and log-transformed T2 risk estimates. The table further reports GLM regressions using the gamma distribution (log link) and negative binomial distribution (logit link). As shown by the table, reported *p*-values for the model coefficients differ only slightly and do not lead to qualitatively different interpretations of the results.

Table S2 Comparison of Alternative Generalized Linear Model Regression Analyses of Participants' T2 Risk Estimates

Model	b	95% CI	z	p
Raw values ($R^2 = .299$, AIC = 3,407)				
Contextual information	4.50	[-0.02, 9.02]	1.95	.052
Static-99R risk level	23.00	[18.48, 27.51]	9.98	< .001
General risk estimate	0.39	[0.28, 0.50]	7.15	< .001
Contextual information × Static-99R	2.54	[-6.51, 11.58]	0.55	.583
Rank-transformed values ($R^2 = .368$, AIC = 4,430)				
Contextual information	16.80	[-0.80, 34.40]	1.87	.062
Static-99R risk level	111.16	[93.56, 128.75]	12.38	< .001
General risk estimate	1.61	[1.19, 2.03]	7.53	< .001
Contextual information × Static-99R	-9.12	[-44.36, 26.12]	-0.51	.612
Log-transformed values ($R^2 = .369$, AIC = 933)				
Contextual information	0.17	[-0.00, 0.33]	1.94	.054
Static-99R risk level	1.07	[0.90, 1.24]	12.50	< .001
General risk estimate	0.01	[0.01, 0.02]	7.34	< .001
Contextual information × Static-99R	-0.07	[-0.40, 0.27]	-0.38	.701
Gamma regression with log link ($R^2 = .343$, AIC = 3,044)				
Contextual information	0.15	[-0.02, 0.33]	1.69	.092
Static-99R risk level	0.94	[0.77, 1.12]	10.42	< .001
General risk estimate	0.02	[0.01, 0.02]	8.39	< .001
Contextual information × Static-99R	-0.10	[-0.46, 0.25]	-0.57	.567
Negative binomial regression with logit link ($R^2 = .338$, AIC = 3,036)				
Contextual information	0.16	[-0.01, 0.33]	1.86	.062
Static-99R risk level	0.99	[0.82, 1.16]	11.49	< .001
General risk estimate	0.02	[0.01, 0.02]	9.15	< .001
Contextual information × Static-99R	-0.12	[-0.46, 0.22]	-0.68	.499

Note. The analyses of raw, rank-transformed, and log-transformed values used a gaussian (normal) distribution with the identity link function, which produces identical results as ordinary least squares (OLS) regression. A constant of 1 was added to the T2 Risk estimates before log-transformation and before use in the gamma regression, as neither method accepts zero values. AIC = Akaike information criterion. Contextual information (0 = *positive*, 1 = *negative*) and Static-99R risk level (0 = *low risk*, 1 = *high risk*) were dummy-coded. N = 376.

Appendix I

Supplemental Materials – Chapter 6: Perceptions of Sex Offenders Study

Parallel Analyses of T2 Risk Estimates

To investigate the robustness of the results reported in the main analysis, parallel analyses were performed on participants' T2 risk estimates. Table S1 reports the results of generalized linear model (GLM) regression analyses using the gaussian (normal) distribution and identity link function for the raw (untransformed) and log-transformed T2 risk estimates. The table further reports GLM regressions using the gamma distribution (log link) and negative binomial distribution (logit link). As shown by the table, reported *p*-values for the model coefficients differ only slightly and do not lead to qualitatively different interpretations of the results.

Table S1 Comparison of Alternative Generalized Linear Model Regression Analyses of Participants' T2 Risk Estimates

Model	<i>b</i>	95% CI	<i>z</i>	<i>p</i>
Raw values ($R^2 = .229$, AIC = 3,659)				
General risk estimate	0.30	[0.17, 0.43]	4.47	< .001
PSO score	0.64	[0.45, 0.83]	6.66	< .001
Contextual information	4.97	[0.52, 9.42]	2.19	.029
PSO × Contextual Information	-0.19	[-0.52, 0.15]	-1.07	.284
Log-transformed values ($R^2 = .224$, AIC = 3,402)				
General risk estimate	0.01	[0.01, 0.02]	4.36	< .001
PSO score	0.02	[0.02, 0.03]	6.54	< .001
Contextual information	0.21	[0.05, 0.37]	2.55	.011
PSO × Contextual Information	-0.01	[-0.02, 0.01]	-0.81	.421
Gamma regression with log link ($R^2 = .242$, AIC = 3,405)				
General risk estimate	0.01	[0.01, 0.02]	5.13	< .001
PSO score	0.02	[0.02, 0.03]	7.09	< .001
Contextual information	0.22	[0.08, 0.38]	2.97	.003
PSO × Contextual Information	-0.00	[-0.02, 0.01]	-0.83	.407
Negative binomial regression with logit link ($R^2 = .239$, AIC = 3,400)				
General risk estimate	0.01	[0.01, 0.02]	5.18	< .001
PSO score	0.02	[0.02, 0.03]	7.21	< .001
Contextual information	0.24	[0.08, 0.39]	3.03	.002
PSO × Contextual Information	-0.01	[-0.02, 0.01]	-0.89	.371

Note. The analyses of raw and log-transformed values used a gaussian (normal) distribution with the identity link function, which produces identical results as ordinary least squares (OLS) regression. A constant of 1 was added to the T2 Risk estimates before log-transformation and before use in the gamma regression, as neither method accepts zero values. AIC = Akaike information criterion. Contextual information (0 = positive, 1 = negative) and Static-99R risk level (0 = low risk, 1 = high risk) were dummy-coded. $N = 402$.

Factors Predicting T2 Risk Estimates

A multiple linear regression analysis revealed that participants' ratings of the importance placed on the SVP respondent's criminal history and on the impact of the crime on the victim were significant positive predictors of participants' T2 risk estimates, whereas the importance placed on the Static-99R score was a significant negative predictor. Jointly, the predictors in the regression model accounted for approximately 21% of the variance in T2 risk estimates. These results are presented in Table S2.

Table S2 Self-Reported Factors Predicting T2 Risk Estimates

Predictor	b	95% CI	SE _b	β	p
Relationships with others	-0.019	[-0.109, 0.071]	0.046	-0.023	.673
Employment history	0.017	[-0.067, 0.100]	0.042	0.022	.695
Criminal history	0.123	[0.036, 0.211]	0.044	0.133	.006
Impact of the crime on the victim	0.134	[0.065, 0.204]	0.035	0.195	< .001
Length of prison sentence	0.090	[-0.001, 0.180]	0.046	0.101	.051
Static-99R information	-0.346	[-0.423, -0.269]	0.039	-0.412	< .001

Note. N = 402. Model F(6, 395) = 18.71, R²_{Adj.} = .209 p < .001.

These supplemental analyses reveal that at T2, after reviewing the Static-99R report, participants who rated the SVP respondent's criminal history and impact of the crime on the victim as more important to their estimate of the SVP respondent's risk of sexual reoffending provided significantly higher estimates of that risk. Previous studies have also found that mock and actual jurors rate criminal history as important in SVP civil commitment decision-making (Krauss et al., 2012; Lieberman et al., 2007; Scott et al., 2010; D. B. Turner et al., 2015). In fact, criminal history is predictive of sexual recidivism risk (Craig et al., 2003), but it is already accounted for in sexual recidivism SRAIs, including the Static-99R (Phenix, Fernandez, et al., 2016). Criminal history, therefore, may carry disproportionate weight in SVP cases because of what is effectively "double counting."

As for the impact of the crime on the victim, this factor is not empirically predictive of sexual recidivism, but was nevertheless associated with an increased perception of the SVP respondent's likelihood of reoffending. This finding is in line with previous research indicating that victim impact statements tend to increase mental health professionals' perceptions of violent reoffending risk as well as the tendency to find that the offender meets the legal criteria to qualify as a dangerous offender (Lynett & Rogers, 2000). In fact, victim impact statements have been shown to increase mental health professionals' perceptions of an offender's future recidivism risk and decisions to indefinitely confine the offender (Lynett & Rogers, 2000). Our

findings suggest victim impact statements may produce a similar effect on jurors' perceptions of an offender's risk of recidivism.

However, we also found that as participants' ratings of the importance of the Static-99R increased, their estimate of the SVP respondent's risk of sexual reoffending significantly decreased. This finding suggests that when jurors perceive that SRAI information is important, their perceptions of the respondent's sexual recidivism risk become more accurately informed by factors that are relevant to risk, and less so by risk-irrelevant information.

Self-Reported Factors Predicting Recommendation for Civil Commitment at T2

We conducted a binary logistic regression analysis to evaluate whether participants' decision to choose civil commitment at T2 could be predicted by ratings of the importance of risk-irrelevant and risk-relevant (i.e., Static-99R report) factors to their disposition recommendation. We included as predictors: participants' importance ratings of the SVP respondent's relationship with others, employment history, and criminal history; impact of the crime on the victim; participant's own estimate of the respondent's sexual reoffending risk; length of the prison sentence; and Static-99R information. The results are presented in Table S3.

Participants' importance ratings with respect to two factors significantly predicted civil commitment at T2: first, the impact of the crime on the victim and, second, the Static-99R information. As the ratings of the importance of the impact of the crime on the victim increased, the likelihood of recommending civil commitment significantly increased. In contrast, as the importance rating of the Static-99R increased, the odds of recommending civil commitment significantly decreased.

Table S3 Results of Binary Logistic Regression Analysis of Factors Predicting Civil Commitment Recommendations at T2

Predictor	b	OR	95% CI	p
Relationships with others	-0.194	0.82	[0.61, 1.11]	.197
Employment history	0.081	1.09	[0.80, 1.47]	.600
Criminal history	-0.047	0.95	[0.70, 1.31]	.770
Impact of crime on victim	0.308	1.36	[1.02, 1.81]	.035
Own estimate of Smith's reoffending risk	0.248	1.28	[0.95, 1.73]	.104
Length of prison sentence	0.206	1.23	[0.89, 1.71]	.218
Static-99R information	-0.575	0.56	[0.43, 0.73]	< .001

Note. OR = Odds Ratio. Disposition recommendation is dummy coded (0 = recommendation other than civil commitment and 1 = civil commitment).

Model $\chi^2(7) = 35.18$, $p < .001$, Nagelkerke $R^2 = .147$.

Like the factors participants rated as important to their T2 risk estimates, participants' ratings of the effect of the crime on the victim to their T2 disposition recommendation significantly predicted civil commitment. Our findings are in line with previous research indicating victim impact statements can lead to increased perceptions of risk (Lynett & Rogers, 2000) and recommendations for harsher penalties or more restrictive recommendations (Butler, 2008; Corteen & Steele, 2018; Wevodau et al., 2014). In previous studies, victim impact statements have exhibited a biasing effect on jurors in that they tend to increase affective responses of anger and punitiveness (Wevodau et al., 2014), as indicated by jurors' tendency to recommend a longer or more severe criminal sentence (Butler, 2008; Corteen & Steele, 2018; Lynett & Rogers, 2000; Paternoster & Deise, 2011; Wevodau et al., 2014). Similarly, our findings suggest that a victim impact statement may be perceived as important by some jurors in SVP hearings (especially those with pre-existing punitive beliefs). Thereby, a victim impact statement may increase the juror's perceptions of the respondent's risk of sexual recidivism and increase the likelihood that the juror will advise civil commitment, even though the impact of the crime on the victim is irrelevant to recidivism risk (and to other legal criteria for SVP civil commitment).

Nevertheless, we also note that the importance rating that mock jurors assigned to the Static-99R with respect to their disposition recommendation significantly negatively predicted a recommendation for civil commitment at T2. In other words, as the importance rating for the Static-99R increased, the likelihood of recommending civil commitment significantly decreased. This finding, combined with our finding that higher importance ratings of the Static-99R predicted significantly lower sexual recidivism risk estimates, suggests perhaps that *when* jurors perceive results of an SRAI as important, this information may be effective in mitigating some of the effects of risk-irrelevant information.

