

UC Irvine

UC Irvine Previously Published Works

Title

The Effect of Adjusted Actuarial Risk Assessment on Mock-Jurors' Decisions in a Sexual Predator Commitment Proceeding

Permalink

<https://escholarship.org/uc/item/7w62q39r>

Journal

Jurimetrics, 53(3)

Authors

Scurich, Nicholas

Krauss, Daniel A

Publication Date

2013-11-07

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

**THE EFFECT OF ADJUSTED
ACTUARIAL RISK ASSESSMENT
ON MOCK-JURORS' DECISIONS
IN A SEXUAL PREDATOR
COMMITMENT PROCEEDING**

**Nicholas Scurich
Daniel A. Krauss***

ABSTRACT: Twenty states and the federal government have adopted statutes that authorize the post-incarceration commitment of sexually violent predators. Actuarial risk assessment is commonly used, and in some states statutorily required, to assess the risk of sexual recidivism in these proceedings. Professionals sometimes modify actuarial risk estimates with their own clinical judgment, the so-called *adjusted actuarial approach*. Although this approach is controversial and courts almost uniformly permit it, the effect of this practice on fact finders is unknown. This experiment found that adjusting actuarial risk estimates affected mock-jurors' decisions to commit a respondent, but only when the adjustment increased the risk estimate. Adjusting the risk estimate downwards did not decrease the commitment rate. Notably, this effect occurred without the expert providing any rationale for the adjustment. Further analyses suggest that participants engaged in *motivated reasoning*, which refers to the tendency to selectively credit or discredit information depending on whether it is congenial to the desired outcome. Participants who chose to commit the respondent deemed the assessment highly acceptable when it indicated high risk, and relatively unacceptable when it indicated low risk, even though the substance of the assessments was identical. Implications for the adjusted actuarial approach are discussed in conjunction with existing legal admissibility standards for expert testimony.

CITATION: Nicholas Scurich and Daniel A. Krauss, The Effect of Adjusted Actuarial Risk Assessment on Mock-Jurors' Decisions in a Sexual Predator Commitment Proceeding, 53 *Jurimetrics J.* 395–413 (2013).

Sexual violent predator (SVP) laws allow for the indeterminate post-incarceration civil confinement of defendants previously charged with or con-

*Nicholas Scurich, Ph.D. is an Assistant Professor in the Department of Psychology & Social Behavior, and the Department of Criminology, Law & Society at the University of California—Irvine. Correspondence concerning this article may be addressed to Nicholas Scurich, 4201 Social & Behavioral Sciences Gateway, Irvine, CA 92697-7085. Electronic correspondence should be addressed to nscurich@uci.edu. Daniel A. Krauss, J.D., Ph.D., ABPP (forensic psychology) is a Full Professor in the Department of Psychology at Claremont McKenna College. The authors thank William Grove for a cogent review of this article.

victed of a sexual offense. Washington State enacted the nation's first SVP law in 1990. The move was largely in response to widespread public perception that sexual offenders recidivated at a high rate, a sentiment fueled by the high-profile crimes of Earl Shriver and Wesley Dodd.¹ Since the adoption of Washington State's law, 19 other states and the federal government have passed similar laws.² Ironically, the public perception that led to the promulgation of SVP laws—that sexual offenders have higher recidivism rate compared

1. In 1989, Earl Shriver committed attempted murder, rape, and assault on 7-year-old Ryan Hade in Tacoma, Washington. At the time of his assault on Ryan, Shriver had recently been released from prison for two earlier sexual offenses. Also in 1989, Wesley Dodd, a sex offender with a longer history of previous offenses, admitted to killing 3 boys from Vancouver, Washington, and molesting over 30 additional children. See CHARLES PATRICK EWING, JUSTICE PERVERTED: SEX OFFENSE LAW, PSYCHOLOGY, AND PUBLIC POLICY 10 (2011) ("Public outrage over these rare but horrible crimes, committed within such a limited geographic area and short time span, led the Washington legislature to pass what would become the first. . . laws aimed at allowing states . . . to confine sex offenders . . . even after they have served their full criminal sentences.").

2. KATHY GOOKIN, WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY, COMPARISON OF STATE LAWS AUTHORIZING INVOLUNTARY COMMITMENT OF SEXUALLY VIOLENT PREDATORS: 2006 UPDATE, REVISED, Doc. No. 07-08-1101 (2007); Daniel A. Krauss et al., *Dangerously Misunderstood: Representative Jurors' Reactions to Expert Testimony on Future Dangerousness in a Sexually Violent Predator Trial*, 18 PSYCHOL. PUB. POL'Y L. 18 (2012); Daniel A. Krauss & Nicholas Scurich, *Risk Assessment in the Law: Legal Admissibility, Scientific Validity, and Some Disparities Between Research and Practice*, 30 BEHAV. SCI. L. 215–16 (2013).

States with SVP laws are Arizona (ARIZ. REV. STAT. ANN. §§ 36-3701–36-3717 (2012)), California (CAL. WELF. & INST. CODE §§ 6600–6609.3 (West, Westlaw through ch. 20 of 2013 Reg. Sess. also including chs. 27, 29, & 41)), Florida (FLA. STAT. ANN. §§ 394.910–394.932 (West, Westlaw from the 2013 1st Reg. Sess. of the 23rd Legislature through June 28, 2013)), Illinois (725 ILL. COMP. STAT. ANN. 207/1–207/99 (West, Westlaw through P.A. 98-21 of the 2013 Reg. Sess.)), Iowa (IOWA CODE ANN. §§ 229.A1–229.A.16 (West, Westlaw current with legislation from the 2013 Reg. Sess.)), Kansas (KAN. STAT. ANN. §§ 59-29a01–59-29a24 (West, Westlaw through 2012 Reg. Sess.)), Massachusetts (MASS. GEN. LAWS ANN. ch. 123A, §§ 1–16 (West, Westlaw through Chapter 22 of the 2013 1st Annual Sess.)), Minnesota (Sexually Dangerous Persons—Civil Commitments, 2013 Minn. Sess. Law Serv. Ch. 49 (H.F. 947)(West)), Missouri (MO. REV. STAT. §§ 632.480–632.513 (West, Westlaw through July 1, 2013, of the 2013 1st Reg. Sess. of the 97th General Assembly)), Nebraska (NEB. REV. STAT. ANN. §§ 71-1201–71-1226 (West, Westlaw through the 102nd Legislature Second Regular Session (2012))), New Hampshire (N.H. REV. STAT. ANN. § 135-E (West, Westlaw current through ch. 82 of the 2013 Reg. Sess., not including changes and corrections made by the State of New Hampshire, Off. Legis. Serv.)), New Jersey (N.J. STAT. ANN. §§ 30.4–27.24-30.4-35 (West, Westlaw through L. 2013, c. 68 and J.R. No. 9.)), New York (N.Y. MENTAL HYG. LAW §§ 10.01–10.17) (West, Westlaw through L. 2013, chs. 1 to 56 and 60 to 66), North Dakota (N.D. CENT. CODE ANN. §§ 25-03.3-01 to 25-03.3-24 (West, Westlaw through the 2011 Reg. & Spec. Sess. of the 62nd Leg. Assemb.)) Pennsylvania (42 PA. CONS. STAT. ANN. §§ 6401–6409 (West, Westlaw through Reg. Sess. Act 2013-11) (juveniles only)), South Carolina (S.C. CODE ANN. §§ 44-48-10 to -170 (West, Westlaw through end of 2012 Reg. Sess.)), Texas (TEX. HEALTH & SAFETY CODE ANN. §§ 841.001–841.151 (West, Westlaw through ch. 36 of the 2013 Reg. Sess. of the 83rd Leg.)), Virginia (VA. CODE ANN. §§ 37.2-900–37.2-921 (West, Westlaw through the end of the 2013 Reg. Sess. and the end of the 2013 Sp. Sess. I.)), Washington (WASH. REV. CODE ANN. § 71.09.010–71.09.903 (West, Westlaw 2013 Legislation effective through July 1, 2013)), Wisconsin (WIS. STAT. ANN. §§ 980.01–980.14 (West, Westlaw through 2013 Wisconsin Act 18, published June 22, 2013.)); and the individual provision of the *United States Code* is 18 U.S.C. § 4248 (2006).

to other types of offenders—has not been empirically supported.³ Further, a substantial drop in sexual offending rates actually predates these laws' enactment.⁴

Although jurisdictional differences exist in the statutory requirements of SVP laws, they all include that the respondent: (1) has prior charge(s) or conviction(s) for a specified sexual offense(s); (2) can be diagnosed with a mental abnormality or personality disorder; and (3) as a result of that diagnosis, has an increased likelihood of committing future sexual violence.⁵ While each of these components can be contested, the crux of these civil trials most often involves whether the defendant is at increased risk for sexual reoffending with the primary evidentiary source for this information coming from the expert testimony of mental health professionals.⁶

Mental health professionals base their risk assessment expert testimony on a variety of different methodologies, which range from less scientifically to more scientifically supported.⁷ On the former side are *unstructured clinical judgments* in which experts idiosyncratically and intuitively combine their judgments and professional experiences to develop estimates of risk. On the more scientific side are *actuarial approaches* in which experts use an assessment instrument derived from research that has found empirical links between certain risk factors and recidivism for a specific population and outcome. This approach offers a mathematical formula for the practitioner to use in combining these factors to reach a recidivism estimate.⁸ An intermediary approach, known as *structured professional judgment (SPJ)*, provides an assessment list that allows the practitioner to arrive at a recidivism estimate based on the presence or absence of risk factors derived from the empirical literature. The SPJ

3. See JOHN Q. LA FOND, PREVENTING SEXUAL VIOLENCE: HOW SOCIETY SHOULD COPE WITH SEX OFFENDERS 46 (2005) ("Surprisingly, official law enforcement records indicate that sex offenders, as a group, are not especially dangerous. In fact they commit fewer new crimes than many other types of criminals.").

4. See EWING, *supra* note 1, at xv. See also Tamara Lave, *Throwing Away the Key: Has the Adam Walsh Act Lowered the Threshold for Sexually Violent Predator Commitment Too Far?*, 14 U. PA. J. CONST. L. 391 (2011) (reporting that sex offenses in the United States have been decreasing for two decades).

5. Holly A. Miller et al., *Sexually Violent Predator Evaluations: Empirical Evidence, Strategies for Professionals, and Research Directions*, 29 LAW & HUM. BEHAV. 29, 31 (2005). For example, under the Kansas SVP law, a sexually violent predator is defined as "any person who has been convicted of or charged with a sexually violent offense and who suffers from a mental abnormality or personality disorder which makes the person likely to engage in repeat acts of sexual violence." KAN. STAT. ANN. § 59-29a02(a) (1996). A mental abnormality is defined as "a congenital or acquired condition affecting the emotional or volitional capacity which predisposes the person to commit sexually violent offenses in a degree constituting such a person a menace to the health and safety of others." KAN. STAT. ANN. § 59-29a02(b) (1996).

6. See generally Eric S. Janus Robert Prentky, *Sexual Predator Laws: A Two-Decade Retrospective*, 21 FED. SENT'G REP. 90 (2008).

7. See Jennifer Skeem & John Monahan, *Current Directions in Violence Risk Assessment*, 20 CURRENT DIRECTIONS PSYCHOL. SCI. 38, 38–42 (2011).

8. See, e.g., VERNON L. QUINSEY ET AL., VIOLENT OFFENDERS APPRAISING AND MANAGING RISK (2006); R. KARL HANSON & DAVID THORNTON, STATIC 99: IMPROVING ACTUARIAL RISK ESTIMATES FOR SEXUAL OFFENDERS (1999), available at http://www.publicsafety.gc.ca/res/cor/rep/_fl/1999-02-st99-imp-rsk-ssmnt-eng.pdf.

approach, however, does not base its risk factors on one specific population or outcome, and it allows practitioners to use their own judgments in weighing and combining each enumerated factor in their final risk estimates.⁹

Hanson and Morton-Bourgon's 2009 meta-analysis of 118 studies clearly demonstrates that unstructured clinical judgments were substantially less accurate in predicting recidivism than extant actuarial measures or SPJ instruments for all outcome measures of interest, including general recidivism, sexual recidivism, and sexual violence.¹⁰ The demonstrated superiority of these latter two approaches to risk assessment has led some jurisdictions to require that expert practitioners use specific risk assessment instruments in their sex offender evaluations.¹¹

More scientifically supported assessment instruments, however, are not without their faults. In particular, criticisms of actuarial risk assessment instruments include: (a) their lack of generalizability beyond the population sample on which they were developed; (b) their failure to incorporate rare risk factors (that is, an unusual risk factor specific to the individual) or protective factors (for example, a supportive spouse); and (c) their reliance on static predictive factors (that is, factors, such as age at first offense, that are not subject to change over time or intervention).¹² Most recently, they have been criticized for a large margin of error when applied to the individual case.¹³

9. See, e.g., STEPHEN D. HART, ET AL., *THE RISK FOR SEXUAL VIOLENCE PROTOCOL (RSVP): STRUCTURED PROFESSIONAL GUIDELINES FOR ASSESSING RISK OF SEXUAL VIOLENCE* (2003).

10. See generally R. Karl Hanson & Kelly E. Morton-Bourgon, *The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies*, 21 *PSYCHOL. ASSESSMENT* 1 (2009) ("On the basis of a meta-analysis of 536 findings drawn from 118 distinct samples (45,398 sexual offenders, 16 countries), empirically derived actuarial measures were more accurate than unstructured professional judgment for all outcomes (sexual, violent, or any recidivism,").

11. See Marcus T., Boccaccini et al., *Field Validity of the STATIC-99 and MnSOST-R Among Sex Offenders Evaluated for Commitment as Sexually Violent Predators*, 15 *PSYCOL. PUB. POL'Y L.* 280 (2009). See, e.g. VA. CODE ANN. § 37.2-903 (Virginia requires use of the STATIC-99 system for all evaluations of sexually violent predators and requires those scoring above a certain threshold receive further clinical evaluation).

12. Krauss et al., *supra* note 2, at 20.

13. This criticism is now considered a "core controversy" within the risk assessment field. For arguments against the application of actuarial risk estimates to an individual case, see Stephen Hart et al., *Precision of Actuarial Risk Assessment Instruments: Evaluating the "Margins of Error" of Group v. Individual Predictions of Violence*, 190 *BRIT. J. PSYCHIATRY* s60 (2007) (presenting a mathematical demonstration that the 95% confidence intervals for an individual's risk estimate "were so high as to render risk estimates virtually meaningless."). A follow-up analysis, using a different mathematical procedure, lead the authors to reach a "surprising—perhaps even controversial [conclusion]." See generally David J. Cooke & Christine Michie, *Limitations of Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice*, 34 *LAW & HUM. BEHAV.* 259, 270-72 (2010) (finding that "on the basis of empirical findings, statistical theory, and logic it is clear that predictions of future offending cannot be achieved, with any degree of confidence, in the individual case.").

These conclusions, including the mathematical procedures that supposedly lead to them, have been disputed. For a comprehensive review, see Nicholas Scurich & Richard S. John, *A Bayesian Approach to the Group Versus Individual Prediction Controversy in Actuarial Risk Assessment*, 37 *LAW & HUM. BEHAV.* 237 (2012) (disabusing the aforementioned analyses and

A proposed remedy to these deficiencies is a mixed approach known as “adjusted actuarial assessment.”¹⁴ As the name implies, the approach entails adjusting the actuarial estimate to account for unconsidered risk factors and to tailor the estimate to the particular individual. One testifying expert “characterized the process as utilizing the [] actuarial instrument and adjusting [it] with ‘[a] dash of clinical judgment.’”¹⁵ Adjustments are typically precipitated by concerns about generalizability, or the presence of rare or protective risk factors.¹⁶ For example, an expert might adjust the risk estimate if the appropriate base rate differs from the one on which the instrument was developed, or for instance if the respondent voiced an apparently serious intention to harm a specified person.¹⁷

There is some empirical support for the practice of adjusting actuarial risk estimates in other contexts. For example, research indicates that meteorological forecasts are more reliable with the addition of clinical judgment.¹⁸ For this and other reasons, some risk assessment researchers strongly support—indeed would require—the practice of adjusting actuarial risk estimates,¹⁹ while others have reservations about the practice.²⁰ Written in 2001, the defini-

demonstrating a coherent approach to applying group-derived risk estimates to an individual case with Bayesian statistics).

14. R. Karl Hanson, *What Do we Know About Sex Offender Risk Assessment?*, 4 PSYCOL. PUB. POL’Y & L. 50, 53 (1998) (“The adjusted actuarial approach begins with an actuarial prediction, but expert evaluators can then adjust (or not) the actuarial prediction after considering potentially important factors that were not included in the actuarial measure.”)

15. George G. Woodworth & Joseph B. Kadane, *Expert Testimony Supporting Post-Sentence Civil Incarceration of Violent Sexual Offenders*, 3 LAW, PROBABILITY & RISK 221, 234 (2004).

16. JOHN MONAHAN ET AL., *RETHINKING RISK ASSESSMENT: THE MACARTHUR STUDY OF MENTAL DISORDER AND VIOLENCE* 130 (2001). (“[T]wo primary reasons are given in support of allowing clinicians the option to use their judgment to revise actuarial violence risk assessment estimates. The first reason can be termed *questionable validity generalization* and the second, *rare risk or protective factors*.”)

17. Shoba Sreenivasan et al., *Alice in Actuarial Land: Through The Looking Glass of Changing Static-99 Norms*, 38 J. AM. ACAD. PSYCHIATRY & L. 400, 403 (2010) (presenting a hypothetical example in which a structured risk assessment tool is administered but requires clinical adjustment in order to render the assessment sensible).

18. See MONAHAN ET AL., *supra* note 16, at 134 (“Clinical involvement actually increases, rather than decreases, predictive accuracy in the meteorological context.”)

19. Hanson, *supra* note 14, at 53 (“[I]t would be imprudent for a clinica[n] to automatically defer to an actuarial risk assessment.”) Stephen D. Hart, *The Role of Psychopathy in Assessing Risk for Violence: Conceptual and Methodological Issues*, 3 LEGAL CRIMINOLOGICAL PSYCHOL. 121, 126 (1998) (“Reliance—at least complete reliance—on actuarial decision making by professionals is unacceptable.”)

20. QUINSEY ET AL., *supra* note 8, at 197 (arguing that “actuarial assessment is too good and clinical judgment too poor to risk contaminating the former with the latter”). See Brian R. Abbott, *Throwing the Baby Out With the Bath Water: Is It Time for Clinical Judgment to Supplement Actuarial Risk Assessment?*, 39 J. AM. ACAD. PSYCHIATRY L. 222, 226–27 (2011) (noting the lack of empirical evidence supporting arguments in favor of adjusted actuarial risk assessment, and noting that injecting clinical judgment into actuarial risk assessments obscures transparency and consistency in the assessment). See generally William M. Grove & Paul E. Meehl, *Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy*, 2 PSYCOL. PUB. POL’Y L. 293, 299 (1996) (reviewing the lack of empirical support for adjusting actuarial risk estimates with clinical judgment).

tive text on risk assessment endorses the practice of adjusting actuarial risk estimates, but with the qualification that this recommendation is subject to further research demonstrating the validity of such adjustments.²¹

To date, almost no empirical research on this topic in the context of risk assessment has been conducted. Our search yielded an unpublished doctoral dissertation²² and only one published, peer-reviewed empirical study.²³ That study systematically examined how one common actuarial risk assessment instrument was used by clinicians in actual practice. It found that clinical adjustments were common, occurring in nearly 1/3 of all 100 cases.²⁴ It also found that such adjustments *decreased* predictive validity.²⁵ Indeed, the adjustments were commonly in the wrong direction; that is, upward adjustments of risk were in fact associated with lower risks of recidivism and vice versa.²⁶ The authors also noted that less than half of the adjustments were made in accord with the instructions of the instrument being used.²⁷

These findings call into question the propriety of adjusting actuarial risk estimates. Nevertheless, the practice of adjusting actuarial risk assessments appears common in legal proceedings. In our semi-systematic review of appellate cases involving SVP proceedings, we found that adjusting actuarial estimates almost always occurred,²⁸ and it was always in the direction of an upwards adjustment of risk.²⁹ Courts appear extremely receptive to this prac-

21. MONAHAN ET AL., *supra* note 16, at 134–35.

22. Kathleen Spencer Gore, *Adjusted Actuarial Assessment of Sex Offenders: The Impact of Clinical Overrides on Predictive Accuracy* (2007) (unpublished Ph.D. dissertation, Iowa State University) (UMI No. 3274898; on file with Parks Library Special Collections Department, Iowa State University).

23. *See, e.g.*, Jennifer E. Storey et al., *Utilization and Implications of the Static-99 in Practice*, 24 *SEXUAL ABUSE* 289 (2012). *But see* Hanson & Morton-Bourgon, *supra* note 10, at 7 (In their meta-analysis of 118 prediction studies, they found three studies comparing actuarial to the adjusted actuarial approach, but two were completed by probation officers and one used correctional staff or psychologists. Nevertheless they concluded across sexual, violent and other recidivism measures that “[f]or all three measures, for all types of raters, and for all outcomes, the adjusted scores showed lower predictive accuracy than did the unadjusted actuarial scores.”).

24. Storey et al., *supra* note 23, at 294.

25. *Id.* at 297.

26. *Id.* at 296.

27. *Id.* at 296–97. Unscrupulous adjustments seem to occur in legal proceedings as well. For instance, after reviewing the proffered risk assessment testimony in an actual sexual predator commitment proceeding, two commentators characterized the expert’s adjustment to the actuarial estimate as “sheer seat-of-the-pants extrapolation.” *See* Woodworth & Kadane, *supra* note 15, at 234.

28. *See* Krauss & Scurich *supra* note 2. Our review of the limited number of publicly available sex offender commitment proceedings is by no means comprehensive. Indeed, many if not most sex offender commitment proceedings are not published or accessible by the public. Nevertheless, we have no reason to believe that the practices reflected in the published proceedings we reviewed are systematically unrepresentative of the many cases that are not publically available.

29. *See, e.g.*, *In re Detention of Shaw*, No. 40723-0-II, 2011 WL 69706601, at *5 (Wash. App. Dec. 29, 2011) (allowing a psychologist to adjust his actuarial risk estimates based on various outside factors, including “antisocial orientation, substance abuse, intimacy deficits, hostility, negative emotionality, and inadequate self-assessment of risk”); *In re Palmer*, 265 P.3d 565 (Kan. Ct. App. 2011); *In re Williams*, 253 P.3d 327, 336 (Kan. 2011) (“[L]ow scores on the actuarial tests weigh against finding the State has met its burden. However, other evidence can convince a ra-

tice, and blithely assume that it adds incremental validity.³⁰ Yet, the prevalence and reception of this practice raises the question: what effect, if any, does the adjusted actuarial approach have on jurors?

In general, empirical research examining the effect of risk assessment on mock jurors in SVP proceedings is quite limited. One noteworthy study found that a voir dire sample of jurors was more influenced by unstructured clinical judgment testimony in their decisions in a simulated, videotaped SVP hearing than they were by testimony based on an actuarial risk instrument.³¹ Other research shows that legal decision makers have difficulty applying actuarial risk estimates,³² and generally disfavor the actuarial approach.³³ The adjusted actuarial approach is unique because it combines both clinical and actuarial risk assessments. Research examining the effect of the adjusted actuarial ap-

tional factfinder that the State has met its burden . . . especially when . . . both experts based their opinions on factors other than the tests.”); *see also* Woodworth & Kadane, *supra* note 15, at 233–35 (describing two other instances in which courts preferred the clinically adjusted-augmented risk estimates to the raw actuarial risk estimates).

30. Consider, for example, the recent opinion of the North Dakota Supreme Court, *In re Rubey*, 818 N.W.2d 731, 735 (N.D. 2012):

Dr. Benson [Rubey's expert] based her opinions entirely on the actuarial instruments. She did not find a diagnosis of personality disorder, which diagnosis Dr. Lisota found to be a significant factor both as to risk of re-offending and difficulty in controlling behavior. Dr. Benson did not do a PCL–R, but agreed with Dr. Lisota's scoring and interpretation of the score. The Court finds Dr. Lisota's opinions more persuasive than Dr. Benson's because he looked at and applied *more than just the actuarial instruments in forming his opinions and drawing his conclusions* [emphases added].

Survey research supports the idea that judges disfavor actuarial risk assessment compared to clinical testimony. *See, e.g.*, Richard E. Redding et al., *What Judges and Lawyers Think About the Testimony of Mental Health Experts: A Survey of the Courts and Bar*, 19 BEHAV. SCI. & L. 583, 590 (2001).

31. Krauss et al., *supra* note 2, at 29.

32. *See, e.g.*, Paul Slovic et al., *Violence Risk Assessment and Risk Communication: The Effects of Using Actual Cases, Providing Instruction, and Employing Probability Versus Frequency Formats*, 24 LAW & HUM. BEHAV. 271 (2000) (finding that the format of the actuarial risk estimate (that is, probability versus frequency) influenced decision makers' perceptions of risk); Nicholas Scurich & Richard S. John, *The Effect of Framing Actuarial Risk Estimates on Involuntary Civil Commitment Decisions*, 35 LAW & HUM. BEHAV. 83, 83 (2011) (finding that the framing of an actuarial risk estimate (for example, $p(\text{violence}) = .24$ versus $p(\text{no violence}) = .76$) had a disparate effect on decisions to involuntarily confine a mentally disordered individual). A recent study suggests that these effects could be the result of innumeracy, which refers to the capacity to understand and reason with quantitative information. *See* Nicholas Scurich et al., *Innumeracy and Unpacking: Bridging the Nomothetic/Idiographic Divide in Violence Risk Assessment*, 36 LAW & HUM. BEHAV. 548 (finding that numerate participants, as measured by self-reported numerical abilities, were less likely to be influenced by the degree to which an actuarial risk estimate was unpacked; that is, the degree to which the risk factors on which the estimate was based were articulated).

33. *See* Redding et al., *supra* note 30; *see also* Phylissa P. Kwartner et al., *Judges' Risk Communication Preferences in Risk for Future Violence Cases*, 5 INT'L J. FORENSIC MENTAL HEALTH 185 (2006) (finding that judges prefer nominal risk estimates (for example, high risk) to probabilistic risk estimates, which are typically associated with actuarial risk assessment).

One possible explanation of this finding is that nominal risk estimates come closer to answering the *ultimate issue*, insofar as one assumes concordance between certain nominal categories (for example, “high risk”) and statutory language (for example, “likely to reoffend”). *See* Nicholas Scurich & Richard S. John, *Prescriptive Approaches to Communicating the Risk of Violence in Actuarial Risk Assessment*, 18 PSYCOL. PUB. POL'Y & L. 50, 68 (2012).

proach on fact finders could have implications for its legal admissibility, and inform the debate regarding the professional practice of adjusting actuarial risk estimates. The following study furnishes evidence that bears on this question.

I. METHODS

A. Participants

Three hundred sixty-two jury-eligible U.S. citizens participated in this experiment. One hundred eighty-two (51%) men and 180 (49%) women served as mock jurors. Their ages ranged from 18–66 years, with a mean of 33.9 (standard deviation; $SD = 12.48$) and median of 30 (interquartile range; $IQR = 18$). Participants were recruited through Amazon Mechanical Turk (AMT), which provides an online platform through which “requesters” can post human information tasks (HITs) that “workers” can complete.³⁴ Common HITs include surveys, questionnaires, and market research questions about products and websites. Our HIT required workers to be at least 18 years old and a citizen of the United States. Workers were paid for their participation.³⁵

B. Procedure and Design

Participants read a synopsis of a sexual predator commitment proceeding, the facts of which were adapted from an actual appellate case in Minnesota.³⁶ The case described a thirty-seven-year-old respondent who was nearing the end of a ten-year incarceration for molesting his three-year-old daughter and four-year-old stepdaughter. The State petitioned to have the respondent committed at the termination of his sentence. Participants were told that the respondent did not contest two of the three commitment criteria (he did not contest that he previously engaged in harmful sexual conduct, and he suffers from a personality disorder). However, the respondent claimed that he is not a risk to the community and that the evidence is insufficient to conclude that he is likely to engage in future acts of harmful sexual conduct. To this end, a court-appointed psychologist provided an assessment of the respondent’s risk.

To ensure that participants paid attention to the materials, one question asked participants to select a specific response from a variety of options. Consistent with current practice,³⁷ participants who did not select the appropriate response were removed from all analyses. Less than six percent ($n = 21$) of participants were removed for failing this question.

Participants were then given judicial instructions,³⁸ and asked whether they would commit the respondent as a sexually violent person. Participants

34. See Winter Mason & Siddarth Suri, *Conducting Behavioral Research on Amazon’s Mechanical Turk*, 44 BEHAV. RES. 1, 1–2 (2012).

35. See *id.* at 15.

36. *In re* Civil Commitment of Lueck, No. 31-PR-08-3446, 2010 WL 3744394, at *1–4 (Minn. Ct. App. Sept. 28, 2010).

37. See Daniel M. Oppenheimer et al., *Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power*, 45 J. EXPERIMENTAL SOC. PSYCHOL. 867, 868 (2009).

38. The participants were instructed:

Effect of Adjusted Actuarial Risk Assessment on Mock-Jurors' Decisions

subsequently responded to 10 items probing the acceptability of the risk assessment. Some items pertained specifically to the quality of the risk assessment. For example, participants were asked, "To what extent do you think the risk assessment is accurate?" and "How scientific is the expert's risk assessment?" Other items probed broader questions pertaining to the risk assessment and its persuasiveness. For example, participants were asked "How satisfied are you with the manner in which the risk assessment was conducted?" and "How persuasive would the risk assessment be to the average juror?" All ratings were made on an 11-point Likert scale, where zero indicated low values (for example, "not at all satisfied") and 10 indicated high values (for example, "extremely satisfied"). Five questions were presented on two webpages each. The order of the webpages and the order of questions within each page were fully randomized.

The key manipulations of the experiment were the type (actuarial or clinical) and outcome (low or high risk) of the risk assessment. The design was a 3 x 3 between-participants, fully crossed factorial design (*actuarial risk assessment*: none; low risk; high risk x *clinical risk assessment*: none; low risk; high risk). Participants were randomly assigned to one of the nine possible experimental conditions. When applicable, the actuarial assessment always preceded the clinical assessment, which either agreed or disagreed with the outcome of the actuarial assessment. For example, participants in one condition were told:

. . . Dr. Nichols testified that the Static-99 placed the respondent in the *high risk* group. . . Dr. Nichols also conducted a clinical assessment of the respondent's risk. . . Dr. Nichols testified that, in his expert opinion, the actuarial risk estimate is accurate. He believes that the respondent poses a *high risk* of reoffending.

It was emphasized that the clinical risk assessment considered precisely the same, and no other risk variables, than those considered by the actuarial risk assessment. Importantly, the expert provided no reasoned elaboration for ei-

You have two major duties as a juror: A, determine the facts; and B, reach a verdict by applying the law to the facts. In fulfilling these duties, you must not be concerned with any opinion you may feel I have about the facts. You are the sole judges of the facts.

Expert witnesses: a witness qualified as an expert by education or experience may state opinions on matters in that witness's field of expertise and may also state reasons for those opinions. Expert opinion testimony should be judged just as any other testimony. You are not bound by it. You may accept it or reject it in whole or in part and you should give it as much credibility and weight as you think it deserves considering the witness's qualifications and experience, the reasons given for the opinions, and all the other evidence in the case.

California law provides that a person may be involuntarily committed to the custody of the California Department of Health Services if that person is found beyond a reasonable doubt to be a sexually violent person. A sexually violent person means a person to whom all of the following apply: 1.) the person must have engaged in a course of harmful sexual conduct; 2.) has manifested a sexual, personality, or other mental disorder or dysfunction; and 3.) as a result, is likely to engage in future acts of harmful sexual conduct. The State must prove both of these elements beyond a reasonable doubt or you must find that he is not a sexually violent person.

The State alleges that the respondent is a sexually violent person. The State has the burden of proving this beyond a reasonable doubt. In some cases it is only necessary to prove that a fact is more likely true than not or that its truth is highly probable. In cases such as this the State's proof must be more powerful than that. It must be beyond a reasonable doubt. Proof beyond a reasonable doubt is proof that leaves you firmly convinced that the respondent is a sexually violent person.

ther agreeing or disagreeing with the outcome of the actuarial risk assessment. He simply asserted his opinion. An example is contained in the appendix.

Participants in the (actuarial assessment none, clinical assessment none) condition received no risk assessment information. These participants only learned about the respondent’s background, including his conviction for molestation.

C. Results

A substantial majority of participants (84%, $n = 304$) voted to commit the respondent. Eighty-nine percent (95% CI = .78, 1.0) of participants who received no risk assessment voted to commit the respondent. Collapsing across the outcome of the assessments (that is, high and low risk combined), 90% (95% CI = .82, .97) of participants who received an actuarial risk assessment voted to commit the respondent, compared to 81% (95% CI = .73, .89) of participants who received a clinical risk assessment. A logistic regression indicated that the type of risk assessment had no effect on participants’ decision to commit the respondent ($\chi^2 = 2.792, d.f. = 2, p = .248$).

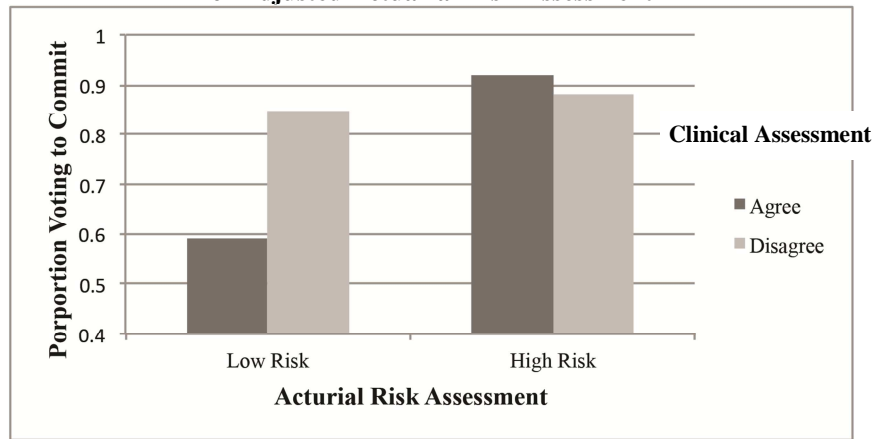
Table 1 reports the proportion (with 95% CIs) of participants voting to commit the respondent decomposed by the type of risk assessment. The left-most column reports the proportion of participants voting to commit the respondent with no actuarial assessment; that is, based on clinical testimony alone (“low risk” or “high risk”), or no risk assessment testimony at all (the upper left cell). The top row reports the proportion of participants voting to commit the respondent when no clinical testimony is presented. The diagonal values reflect the experimental conditions in which the actuarial and clinical assessments agreed. For instance, the actuarial assessment indicated low risk and the clinical assessment agreed, or the actuarial assessment indicted high risk and the clinical assessment disagreed.

Table 1. Proportion of Participants Voting for Commitment

		<u>Actuarial Risk Assessment</u>		
		None	Low Risk	High Risk
Clinical Risk Assessment	None	0.89 [.77, 1.0]	0.78 [.67, .89]	1.0 [.90, 1.0]
	Low Risk	0.76 [.64, .87]	0.59 [.48, .70]	0.88 [.78, .99]
	High Risk	0.87 [.76, .98]	0.85 [.75, .96]	0.92 [.80, 1.0]

Figure 1 presents the results of the experimental conditions that included both types of risk assessments. The rightmost cluster of bars represents the participants who received a high risk actuarial estimate. As is apparent, clinical agreement with the high-risk actuarial estimate produced a meager increase in the proportion of participants voting to commit compared to when the clinical assessment disagreed with the high-risk actuarial estimate and stated that the respondent posed a low risk. This difference is not statistically significant ($\chi^2 = 7.02, d.f. = 2, p = .21$).

Figure 1. Commitment Verdicts as a Function of Adjusted Actuarial Risk Assessment



The bars on the left display the proportion of participants who voted to commit the respondent when the actuarial risk assessment indicated low risk, decomposed by whether the clinical assessment agreed or disagreed with that estimate. When the clinical assessment agreed with the actuarial estimate of low risk, 59% of participants voted to commit; however, when the clinical assessment disagreed with the low risk actuarial estimate, and instead asserted that the respondent posed a high risk, the proportion of participants voting to commit increased to 85%. A logistic regression ($\chi^2 = 7.613, d.f. = 2, p < .05$) indicated that, compared to when the clinical assessment agrees with the low risk actuarial estimate, the odds increase 4.06 (95% CI = 1.38, 11.89; Wald = 6.51; $p < .05$) that participants will vote to commit the respondent when the clinical assessment disagrees with the low actuarial risk estimate and states that he poses a high risk. Bear in mind that the clinical assessment purportedly relied on the exact same risk variables, and it never provided any reasoned elaboration for why it reached a different outcome. In sum, when the expert disagreed with a low actuarial risk estimate and moved his risk estimate upward, the commitment rate increased, but when the expert disagreed with a high risk estimate and moved his risk estimate downward, the commitment rate did not decrease.

The ten questions probing the acceptability of the risk assessment were combined into a composite measure yielding a Cronbach's alpha = .921.³⁹ The high inter-item correlation provides good evidence of a latent construct,⁴⁰ which will be referred to as "acceptability."

A one-way analysis of variance (ANOVA) was conducted to determine if one type of risk assessment was deemed more acceptable than another. It failed to detect any significant differences $F(2, 202) < 1$. The clinical ($m = 7.24$, $SD = 2.21$) and actuarial assessments ($m = 7.48$, $SD = 2.32$) were not more acceptable than no assessment whatsoever ($m = 7.24$, $SD = 7.24$).

The acceptability of the assessment, however, did depend on the outcome of the assessment. For the actuarial risk assessment, the favorability ratings were highest when the assessment indicated high risk ($m = 8.47$, $SD = 1.54$), and lowest when the assessment indicated low risk ($m = 6.43$, $SD = 2.23$); the mean favorability rating for no actuarial assessment was 7.33 ($SD = 2.23$). These differences are statistically significant: $F(2, 362) = 31.27$, $p < .001$, $\eta_p^2 = .148$. Similarly for the clinical assessment, the mean favorability rating was 8.28 ($SD = .189$) when the assessment indicated high risk, and 6.64 ($SD = 2.06$) when the assessment indicated low risk; the mean favorability for no clinical assessment was 7.33 ($SD = 2.23$). These differences were also statistically significant $F(2, 362) = 18.41$, $p < .001$, $\eta_p^2 = .093$.

It seems clear that the acceptability of the risk assessment largely depended on the outcome of the assessment. When the risk assessment indicated low risk, the assessment was deemed relatively unacceptable, yet when the exact same risk assessment indicated high risk, it was deemed relatively acceptable. Because acceptability is contingent on the outcome of the risk assessment, one might surmise that acceptability could also depend on how the risk estimate relates to the participant's verdict.⁴¹

To examine this possibility, we collapsed across the clinical assessment and compared it to the actuarial low and high risk categories. A two-way ANOVA with participant's verdict (release or commit) and risk level (low or high) as the independent variables, and risk assessment acceptability as the dependent variable, detected a significant main effect for risk level $F(1, 245) = 13.62$, $p < .001$, $\eta_p^2 = .053$, and a significant interaction $F(2, 245) = 3.94$, $p < .05$, $\eta_p^2 = .016$. The main effect for participant's verdict was not significant $F(1, 245) < 1$. These findings are illustrated in Figure 2, which depicts the

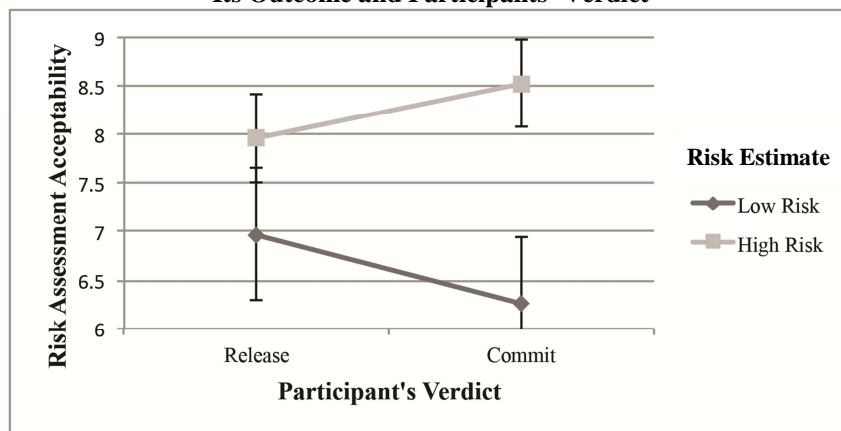
39. This approach was used by Dan Simon and Nicholas Scurich. See Dan Simon & Nicholas Scurich, *Lay Judgments of Judicial Decision Making*, 8 J. EMPIRICAL LEGAL STUD. 709, 715 (2011).

40. A high interitem correlation indicates the presence of an underlying latent construct. In nonpsychometric parlance, the items are highly correlated because they share a "common cause"; that is, the items are measuring the same "thing." Administering multiple items to estimate this "thing" (that is, latent construct) increases the reliability of the scale by decreasing the statistical noise associated with any particular item. For an accessible discussion of Psychometric Theory, see ROBERT F. DEVELLIS, *SCALE DEVELOPMENT: THEORY AND APPLICATIONS* 28 (2003).

41. The findings of Simon & Scurich, *supra* note 39, motivated this hypothesis. For an elaboration of the findings, see *infra* note 68.

acceptability ratings (+/- 2 S.E.) as a function of participants' verdict and the outcome of the risk assessment.

Figure 2. Evaluation of the Risk Assessment as a Function of Its Outcome and Participants' Verdict



The acceptability ratings were not related to the outcome of the risk assessment for the participants who voted to release the respondent; these participants did not think that the assessment was more acceptable if it indicated low risk or high risk $t(39) = .79, p = .435$. However, for participants who voted to commit the respondent, the acceptability of the assessment was contingent on the outcome it yielded. For these participants, the assessment was highly acceptable when it indicated high risk, and relatively unacceptable when it indicated low risk. The difference is statistically significant: $t(207) = 2.27, p < .001$.

II. DISCUSSION

The primary purpose of this study was to examine what effect adjusted actuarial risk assessment had on mock-jurors' decisions in a sexual predator commitment proceeding. A ceiling effect was found in that a majority of participants voted to commit the respondent. Indeed, nearly 90% of the participants who received no risk assessment voted to commit the respondent based solely on information related to the previous crimes for which he was imprisoned. This finding is particularly interesting given that our synopsis of the case omitted details that were likely to evoke even greater contempt for the respondent.⁴² For instance, the experimental stimuli omitted the fact that the offender continued to have sexual intercourse with minors despite his aware-

42. *In re* Civil Commitment of Robert Lueck, No. 31-PR-08-3446, 2010 WL 3744394, at *1 (Minn. Ct. App. Sept. 28, 2010) (stating respondent had two prior convictions for third-degree sexual assault at the age of 19, when he had sexual intercourse with a 15-year-old and a 14-year-old).

ness that he was HIV positive.⁴³ The fact that nearly 90% of participants would commit the respondent on the basis of his previous crimes alone suggests that their decisions could have had punitive and retributive aspirations.⁴⁴ Of course, one could argue that the lack of any risk assessment required these participants to rely on their intuition about the respondent's risk.⁴⁵

The apparent ceiling effect might have actually dampened the observed effect that adjusting the risk estimate upwards has on verdicts. It is possible that the upward adjustment effect is understated because it was constrained by the fact that the nonadjusted commitment rate was quite high. Conversely, the ceiling effect should present an optimal condition to find that adjusting downward decreases the commitment rate (because the range below the estimate is relatively larger). However, adjusting the risk estimate downward had no effect when the actuarial risk estimate indicated high risk (right side of Figure 1), which buttresses the finding that only adjusting upward matters.

One possible explanation of why the downward adjustment had no effect is that participants engaged in *motivated reasoning*. Motivated reasoning refers to the biased search and processing of information to reinforce some goal extrinsic to decision accuracy.⁴⁶ In short, people seek out information that is congenial to their preferred outcome, and when presented with information, they selectively credit or discredit that information depending on whether it

43. *Id.* at *1–2.

44. See generally Kevin M. Carlsmith et al., *The Function of Punishment in the "Civil" Commitment of Sexually Violent Predators*, 25 BEHAV. SCI. L. 437 (2007) The study found that, contrary to professed and legally mandated intentions, retributive aspirations suffuse sex predator commitment decisions. In particular, the study found that the decision to commit a sexual predator turned not only on the likelihood of recidivism, but also on whether the respondent had been subjected to sufficient punishment for his crime. Insufficiently punished respondents were committed at a relatively higher rate, even when some participants, who were jury-eligible citizens, were told the respondent posed virtually no risk of recidivism.

45. It is possible that the reliance on "intuition" could also explain the pattern of results of the participants who were furnished with a risk assessment, either clinical or actuarial. In particular, the finding that adjusting the actuarial estimate had an effect only when it was in the upward direction could be the result of a predisposition that precluded participants from accepting the possibility that the respondent posed a "low" risk despite what the risk assessment indicated. This phenomenon is consistent with the posited explanation of motivated reasoning. See Ziva Kunda, *The Case for Motivated Reasoning*, 108 PSYCHOL. BULL. 480, 480 (1990). A facet of this phenomenon is known as *selective scrutiny*. Selective scrutiny refers to the tendency to scrutinize noncongenial information but to accept information at face value when it is consistent with a predisposition. See generally Charles G. Lord et al., *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*, 37 J. PERSONALITY SOC. PSYCHOL. 2098 (1979) (finding that people with strong opinions on social issues are likely to consider empirical evidence in a biased manner). Hence, it is possible that because participants initially believed the respondent posed a high risk, they accepted the risk assessment when it confirmed this belief, and disregarded the assessment when it frustrated that belief.

46. See generally Kunda, *supra* note 45 (finding that subjects' decisions about whether to accept a study's findings at face value depend less on the methodology employed than on whether the study's results coincided with their existing beliefs).

supports the preferred outcome. Empirical research suggests this phenomenon is ubiquitous.⁴⁷

Evidence of motivated reasoning is provided by the analyses examining the reported acceptability of the risk assessment. For the participants who voted to commit, the risk assessment was considered highly acceptable when it indicated high risk, but relatively unacceptable when it indicated low risk.⁴⁸ This effect was obtained despite the fact the assessments were identical except for their outcome. There is no logical reason that the acceptability of the risk assessment should be contingent on the outcome it furnishes. Indeed, such reasoning effaces the need for a risk assessment at all. While reasonable people can and will disagree about the quality of a given risk assessment, selectively crediting or discrediting the assessment based on its outcome can only reinforce preexisting desires and goals, which does little to promote informed decisions.

Motivated reasoning can be especially pernicious because it has the potential to carry over and infect other, seemingly unrelated aspects relevant to the decision making process. Simon's work on *coherence-based reasoning* demonstrates that people tend to view unrelated information as interdependent, which in turn reinforces and strengthens preexisting beliefs.⁴⁹ Thus, not only do people selectively search and credit consistent evidence, they tend to view all the evidence as correlated, and in the direction of providing overwhelming support for the desired conclusion.⁵⁰ As this possibility relates to the current study, keep in mind that the respondent stipulated to two of the three commitment criteria. Coherence-based reasoning would predict that information bearing on one criterion would affect the perception of information pertaining to the other two.⁵¹ Although there clearly is a dependency between the com-

47. See Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 176–84 (1998).

48. This finding is consistent with one previous study examining judicial preferences of risk communication. Specifically, that study found that actual judges ascribed more “probative value” to a risk assessment when it indicated high risk than when the assessment indicated low risk. Kwartner et al., *supra* note 33, at 191. In the high-risk condition, three risk factors were present, and in the low risk condition, the three risk factors were absent. The measure of probative value was a rating made on a 10-point scale ranging from “not at all valuable” to “extremely valuable.” *Id.* at 188. However, it is not clear that the presence of the three risk factors is commensurate with the absence of the three risk factors in terms of probity. Thus, the study does not furnish particularly strong evidence of motivated reasoning because the judges might have (correctly) treated the presence of risk factors as more probative than the absence. In contrast, the present study did not manipulate the substance of assessment; it only manipulated the outcome, which clarifies the cause of the differences—that is, agreement with the outcome—in the acceptability ratings.

49. See Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511, 583–86 (2004).

50. See *id.* See also Dan Simon et al., *The Redux of Cognitive Consistency Theories: Evidence Judgments by Constraint Satisfaction*, 86 J. PERSONALITY & SOC. PSYCHOL. 814 (2004) (presenting several empirical studies demonstrating that complex, substantively independent and under-determined evidence will morph into a perception of the evidence which strongly supports the preferred outcome and strongly disagrees with the alternative outcome).

51. See generally Lisa E. Hasel & Saul M. Kassin, *On the Presumption of Evidentiary Independence: Can Confessions Corrupt Eyewitness Identifications?*, 20 PSYCHOL. SCI. 291 (2009) (presenting a study in which participants who identified a particular suspect as the perpe-

mitment criteria—for example, evidence of a previous conviction for a sexual offense is relevant to the likelihood of future recidivism—the assessment of the latter should not be unduly influenced by the former, which is precisely what coherence-based reasoning would predict to occur. This conjecture requires empirical testing, but it could have significant implications for legal policy.

Expert testimony is subject to greater scrutiny by the courts than other forms of evidence because it is thought to have a greater influence on jurors, an assumption that is generally supported by empirical research.⁵² A trilogy of United States Supreme Court cases (*Daubert v. Merrell Dow*⁵³; *General Electric v. Joiner*⁵⁴; and *Kumho v. Carmichael*⁵⁵) make clear that federal judges are responsible for evaluating the scientific validity of expert testimony before admitting it. The majority of states have followed the lead of the federal courts, implementing potentially more exacting standards for the evaluation of expert testimony. Even states that have retained a general acceptance standard for evidentiary admissibility (*Frye v. United States*⁵⁶) have moved generally towards greater scientific scrutiny of experts.⁵⁷ Yet, while some courts are required to closely evaluate the scientific validity of expert testimony, they have shown little willingness to do so when that testimony concerns the risk of sexual recidivism. To date, almost no expert testimony on future sexual risk has been excluded by the courts.⁵⁸ Instead, courts uniformly hold that their evidentiary admissibility standards do not apply to expert testimony based on actuarial assessments of risk, or, if they do apply, that most of such assessments meet their standard.⁵⁹ Further, in the rare instances when such expert

trator learned that a different person confessed to the crime and subsequently changed their identification to the confessor. Coherence-based reasoning is the posited explanation of why the confession affected the identification).

52. See generally Margaret Kovera & Brian Cutler, *Expert Psychological Testimony*, 20 CURRENT DIRECTIONS PSYCHOL. SCI. 54 (2011) (highlighting that given the specialized nature of expert testimony and its risk of abuse, lawmakers have created safeguards to prevent “junk science” from influencing jurors).

53. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 113 S. Ct. 2786 (1993).

54. *General Elec. Co. v. Joiner*, 522 U.S. 136, 118 S. Ct. 512 (1997).

55. *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 119 S. Ct. 1167 (1999).

56. *Frye v. U.S.*, 54 App.D.C. 46, 293 F. 1013 (D.C. Cir. 1923).

57. See generally Edward K. Cheng & Albert H. Yoon, *Does Frye or Daubert Matter? A Study of Scientific Admissibility Standards*, 91 VA. L. REV. 472, 503 (2005) (After compiling and comparing removal rates among state jurisdictions with *Frye* or *Daubert* admissibility standards, the authors conclude: “The results of this study are consistent with the theory that the power of the Supreme Court’s *Daubert* decision was not so much in its formal doctrinal test, but rather in its ability to create greater awareness of the problems of junk science. This suggests that courts apply some generalized level of scrutiny when considering the reliability of scientific evidence, regardless of the governing standard. If accepted, this thesis suggests that debates about the practical merits and drawbacks of adopting a *Frye* versus a *Daubert* standard are largely superfluous.”).

58. See generally Randy Otto & John Petrila, *Admissibility of Expert Testimony Regarding Recidivism Risk in Sexually Violent Predator Proceedings*, in 3 THE SEXUAL PREDATOR: LAW AND PUBLIC POLICY/CLINICAL PRACTICE (Anita Schlank ed., 2006).

59. But see CHRISTOPHER SLOBOGIN, PROVING THE UNPROVABLE: THE ROLE OF LAW, SCIENCE, AND SPECULATION IN ADJUDICATING CULPABILITY AND DANGEROUSNESS 122–25 (2007) (offering a different interpretation of why courts may (and should) not be applying eviden-

testimony has been scrutinized, the courts have often made errors in their reasoning about the testimony.⁶⁰

This legal reluctance is especially troubling because approximately 75% of states with SVP laws allow for jury trials, and a substantial body of research demonstrates that jurors generally have difficulty evaluating complex expert testimony.⁶¹ Other empirical research suggests that this difficulty generalizes to expert testimony on risk assessment in SVP hearings.⁶² The present finding that jurors are influenced by unsupported and unelaborated expert pronouncements further illustrates the limited ability of jurors to appropriately evaluate expert testimony in general, and risk assessments in particular, and it impels courts to be cognizant of these potential limitations when making decisions about the admissibility of such evidence.

The usual caveats associated with trial simulations apply to the current experiment.⁶³ Notably, some research has found differences between college students and more representative samples of jurors in SVP trial simulations (for example, representative samples are more likely to commit the individual), and without further study and replication, it is not clear if the present AMT mock jurors might also show differences from more representative juror samples.⁶⁴ Additionally, our stimuli were intentionally austere, and thus potentially unrepresentative of the way in which such adjustments occur in legal settings. In particular, adjustments in legal proceedings are usually precipitated by the putative neglect of unconsidered yet relevant risk variables,⁶⁵ and clinicians tend to explicate the rationale for adjustment.⁶⁶ However, the austere approach was used to isolate the effect of adjustment from any confounds that exist in ecology, such as the influence of the substantive (but potentially spe-

tiary admissibility standards strictly to expert testimony on risk in SVP hearings). Slobogin argues that rather than *Daubert* or *Frye* controlling the admissibility of this evidence that Federal Rule of Evidence 403 (Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons) should control the circumstances in which this testimony is admissible, with the side (state versus the offender) that seeks to admit the evidence being important in the court's decision to admit or not admit the evidence. Further discussion of this issue is beyond the scope of this article.

60. See Krauss & Scurich, *supra* note 2, at 224.

61. Kovera & Cutler, *supra* note 52, at 55–56.

62. See generally Krauss et al., *supra* note 2.

63. See generally Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 LAW & HUM. BEHAV. 75 (1999).

64. Krauss et al., *supra* note 2.

65. Two assumptions are built into the “unconsidered yet relevant” argument: (1) the risk variables are unconsidered in that they are not contemplated by the actuarial instrument; and (2) the risk variables are relevant in that they add incremental validity to the prediction. The voluminous research examining clinical versus actuarial predictions generally concludes that both assumptions are untenable. See, e.g., REID HASTIE & ROBYN M. DAWES, RATIONAL CHOICE IN AN UNCERTAIN WORLD (2nd ed. 2009). For a discussion within the context of SVP assessment, see Krauss & Scurich, *supra* note 2. Nevertheless, the ‘unconsidered yet relevant’ argument appears to be highly persuasive to courts when evaluating the scientific validity of expert testimony. See *id.*

66. Indeed, in some states psychologists are required to articulate the bases of their assessment, including any modifications to an actuarial risk assessment, as a matter of professional ethics. See, e.g., Minnesota Board of Psychology Code of Conduct 7200.5000 Subpart 2.

cious⁶⁷) reasons that are provided to justify the adjustment. It is possible that more ecologically valid materials could yield different results.⁶⁸ The current findings should be regarded as preliminary unless replication occurs. Other future research might examine whether cultural values influence the tendency to engage in motivated reasoning in this context,⁶⁹ and what can be done to mitigate its occurrence.

III. FINAL THOUGHTS

Adjusting actuarial risk estimates appears to be commonplace in legal proceedings, but fundamental issues have yet to be carefully considered by the courts when evaluating the quality of the expert testimony that this practice yields. Empirical validity is obviously one important dimension of the practice; evidence lacking validity cannot logically assist in determining the fact at issue. A second dimension concerns what effect this practice has on jurors. The findings of the present study speak to this latter issue, and suggest that the practice has an asymmetric influence on jurors' decisions. Although the present findings are to be considered preliminary until replicated, at a minimum they support the position that judges ought to cogently appraise risk assessment testimony in general, and adjusted actuarial risk assessment in particular, before admitting it into evidence.

67. See Woodworth & Kadane, *supra* note 15, at 233–35.

68. If the motivated reasoning explanation of the present results is accurate, however, furnishing reasons for the adjustments will not remediate the tendency to consume actuarial risk assessments in a biased manner. In a different context, researchers have found that providing reasoned elaboration increased the legitimacy of a legal decision rendered by an appellate court, but only for participants who disagreed with the decision, and, more significantly, its effect was meager. Agreement with the decision overwhelmingly affected the legitimacy ratings of the decision. See Simon & Scurich, *supra* note 39, at 719.

69. See generally Dan Kahan et al., *Cultural Cognition and Public Policy: The Case of Outpatient Commitment Laws*, 34 LAW & HUM. BEHAV. 118 (2010) (finding that people's cultural values shape their views about the outpatient civil commitment laws).

APPENDIX

The court appointed a forensic psychologist, Dr. Nichols, who has a Ph.D. in clinical psychology, and has conducted over 200 sex offender evaluations for the court. He spent approximately 6 hours over 2 days interviewing the respondent.

To assess the risk that the respondent will reoffend, Dr. Nichols used the Static-99, an actuarial tool that is widely used in SVP commitment proceedings. In the same way that insurance companies use actuarial tools to assess driving risk, researchers have developed actuarial tools to assess the risk of sexual reoffense. These tools rely on risk factors with an empirically established relation to sexual offending, such as age at release, relationship to prior victims, and prior involvement with the criminal justice system. The Static-99 uses 10 risk variables, which are scored by the examiner based on whether the variable is present. The total score for all 10 variables is then summed. The summed scores lead to classification in one of three possible risk groups: low; moderate; or high.

(Actuarial low, clinical high)

Dr. Nichols testified that the Static-99 placed the respondent in the *low risk* group.

Dr. Nichols also conducted a clinical assessment of the respondent's risk. He explained that he used his expert judgment to generate an estimate of the risk of reoffense based on the same 10 variables considered by the Static-99. He emphasized that his assessment considered all 10 of the variables.

Dr. Nichols testified that, in his expert opinion, the actuarial risk estimate is inaccurate. He believes that the respondent poses a *high risk* of reoffending.

(Actuarial high, clinical low)

Dr. Nichols testified that the Static-99 placed the respondent in the *high risk* group.

Dr. Nichols also conducted a clinical assessment of the respondent's risk. He explained that he used his expert judgment to generate an estimate of the risk of reoffense based on the same 10 variables considered by the Static-99. He emphasized that his assessment considered all 10 of the variables.

Dr. Nichols testified that, in his expert opinion, the actuarial risk estimate is inaccurate. He believes that the respondent poses a *low risk* of reoffending.