DANGEROUSLY MISUNDERSTOOD: Representative Jurors' Reactions to Expert Testimony on Future Dangerousness in a Sexually Violent Predator Trial
Daniel A. Krauss, John G. McCabe, and Joel D. Lieberman
Online First Publication, July 25, 2011. doi: 10.1037/a0024550

CITATION
DANGEROUSLY MISUNDERSTOOD
Representative Jurors’ Reactions to Expert Testimony on Future Dangerousness in a Sexually Violent Predator Trial

Daniel A. Krauss  
Claremont McKenna College

John G. McCabe  
Claremont Graduate University

Joel D. Lieberman  
University of Nevada–Las Vegas

Past research examining the effects of psychological expert testimony concerning future dangerousness in sexual violent predator commitment trials has produced equivocal results on whether mock jurors are more influenced by intuitive clinical expert testimony than empirically based actuarial testimony. This study advances this line of research by examining these effects using a more ecologically valid sample of 156 venire jurors who watched a simulated, videotaped SVP trial based on a case transcript. As predicted, jurors were more influenced by the less scientific, clinical expert testimony. The data provided only limited support for a theoretical explanation based on Cognitive-Experiential Self-Theory (CEST). The policy implications of these results are discussed.

Keywords: expert testimony, jury decision-making, sexually violent predators, clinical expert testimony, actuarial expert testimony

In the United States legal system, predictions of an individual’s future likelihood of reoffense or violence have a long history. These predictions are employed in many legal determinations that have significant liberty implications for the individuals involved, from bail decisions to capital sentencing. In the last 20 years, a significant minority of states have adopted Sexually Violent Predator or Person laws (hereinafter SVP laws). These modified civil commitment laws allow for postincarceration civil confinement of sexual offenders who are found likely to commit future acts of sexual violence. In these legal adjudications, mental health experts proffer expert testimony on whether or not a specific individual poses a future danger. Yet, almost since the inception of dangerousness-based legal standards, the expertise of mental health professionals on these matters has been questioned (e.g., Krauss & Sales, 2001; LaFond, 2005; see Slobogin, 2006 for a detailed discussion of these issues).

Historically, courts have been largely deaf to the cries of critics in these important legal decisions. Beginning with the landmark U.S. Supreme Court decision in Bare-
foot v. Estelle (1983), courts have uniformly admitted mental health practitioners’ expert testimony on an individual’s future risk of dangerousness or violence in a wide variety of matters. They have done so even when such expert testimony has been challenged under evidentiary admissibility standards rather than constitutional ones (e.g., see Otto & Petrila, 2006, for a review of these latter legal challenges with regard to future sexual violence testimony in SVP cases).

Yet, this trend may be changing. Most recently, the Texas Court of Criminal Appeals has found specific expert testimony on future dangerousness to fail its evidentiary admissibility standard in a death penalty case (Coble v. State, 2010), and this case is currently being appealed to the United States Supreme Court on grounds that it violates the Constitution’s 8th amendment. Further, the American Psychological Association has submitted an amicus brief in support of the appeal (APA, 2011) based on a burgeoning body of psycholegal literature that suggests that jurors generally lack the ability to distinguish good science from bad (e.g., Cutler & Kovera, 2010) and an additional smaller body of empirical literature that suggests jurors may be specifically unable to accurately differentiate more scientifically valid expert testimony on dangerousness from less valid expert testimony and to appropriately weigh each in their decision-making (e.g., see Krauss, McCabe, & McFadden, 2009).

Regardless of what the United States Supreme Court decides in Coble, the APA position in its brief and the Texas appellate court’s ruling prohibiting expert testimony on future dangerousness on evidentiary grounds, suggest courts will be more likely to closely scrutinize expert testimony on future dangerousness in a number of different legal contexts, including SVP cases. In light of this trend, the present research examines how a sample of venire jurors process risk prediction expert testimony of varying scientific quality in a realistically simulated SVP trial.

Sexual Violent Predator Laws

Beginning in 1990 and continuing to 2008, 20 states and the federal government adopted SVP laws (McCabe, Krauss, & Lieberman, 2010). As of 2006, when only 17 states possessed SVP laws, over 4,500 individuals had been committed under their provisions with only 494 offenders having been released once they were confined (Gookin, 2007). The number of committed sexual offenders is likely to increase substantially in the next several years as a result of the federal government’s SVP law (18 U.S.C. 4248) being found not to violate the U.S. Constitution (United States v. Comstock, 2010). This decision allows for the country’s largest prison system to begin civilly committing sex offenders after they have served their federal prison sentences.

At the heart of the vast majority of SVP adjudications, is whether there is a likelihood that the individual will reoffend (Janus & Prentky, 2008). In fact, some states mandate that mental health professionals provide expert testimony on this issue to the court, and in many SVP cases psychological expert testimony on the respondent’s risk of future violence is the only piece of testimony admitted at trial (Guy & Edens, 2003; Miller, Amenta, & Conroy, 2005).

---

1 Since it is a civil trial, the individual facing SVP commitment is referred to as a respondent instead of as a defendant.
Of the 20 states that have SVP laws, the vast majority allow either the respondent or the petitioner to choose a jury trial (Gookin, 2007). In the end, it is often left to the jurors in SVP cases to determine how much weight to grant to mental health expert testimony on future dangerousness with the respondent’s liberty hanging in the balance (LaFond, 2005).

**Expert Testimony on Future Risk**

In the past several decades, there have been vast improvements in the quality of expert testimony mental health practitioners could offer concerning an individual’s risk of future violence or reoffense. Expert testimony on future risk can vary from less reliable, intuition-based clinical judgments to more scientifically valid, empirically based assessments. Clinical judgments are typically based on an expert’s knowledge combined idiosyncratically with their prior experiences. In contrast, scientific evaluations primarily involve the utilization of specially designed assessment instruments. These instruments either rely on empirically derived risk factors that are optimally combined to produce recidivism probability estimates based upon a specific population (i.e., actuarial instruments), or on a clinician combining, based on his or her judgment, a series of relevant risk factors that have been derived for the empirical literature but are not based on a particular sample population (i.e., structured or guided professional judgment instruments).

A substantial research base demonstrates that with regard to dangerousness predictions intuitive clinically based predictions are significantly inferior to these other two prediction methods (Hanson & Morton-Bourgon, 2009; Quinsey, Harris, Rice, & Cormier, 2006). Hanson and Morton-Bourgon’s (2009) meta-analysis found that unstructured clinical judgments were substantially less accurate than existing actuarial measures or guided professional judgment instruments for all outcome measures of interest: recidivism, sexual recidivism, and sexual violence.

There is still controversy, however, surrounding the use of actuarial methods in SVP cases. These instruments have been criticized for: (a) their lack of generalizability beyond the population sample on which they were developed; (b) their failure to incorporate rare risk factors (i.e., an unusual risk factor specific to the individual) or protective factors (e.g., a supportive spouse); (c) their reliance on static predictive factors (i.e., factors like age at first offense, that are not subject to change over time or intervention); and (d) their reliance on relatively small development samples for their probability estimates, which could lead to substantial error rates when they are applied to individual cases (Douglas & Skeem, 2005; Monahan, 2003; Hart, Michie, & Cooke, 2008). Others scholars have questioned whether these instruments’ predictive accuracy is sufficiently high and their error rates sufficiently low that they should be used in SVP cases where significant individual liberties are at stake (Wollert, 2006).

---

2 Pennsylvania, MN, NJ, and North Dakota are the four states that do not allow the option of a jury trial in these cases (Gookin, 2007).
Even with these weaknesses, there is little doubt that actuarial measures are superior to unstructured clinical judgment for the assessment of sexual recidivism. The Association for the Treatment of Sexual Abusers (ATSA), the primary scientific organization for the treatment and evaluation of sex offenders, has expressly endorsed the use of actuarial measures for the evaluation of sex offenders in SVP hearings (ATSA, 2001), and the state of Virginia expressly requires the use of one specific actuarial instrument in their SVP statute (Boccaccini, Murrie, Hawes, & Caperton, 2009). It is not clear, however, how often available actuarial measures are employed by evaluators in SVP hearings across the U.S. One recent survey across 12 jurisdictions of 41 experienced practitioners that perform SVP evaluations determined that a sizable minority of evaluators (26%) believed that actuarial measures (which included guided or structured professional judgment instruments in this survey) were simply recommended or optional in evaluations (Jackson & Hess, 2007). A reasonable inference from this finding is that even some experienced evaluators are still using less valid unstructured clinical judgment in their assessments. As a consequence, jurors in SVP hearings are likely faced with expert testimony of differing scientific quality and are tasked with appropriately evaluating and weighing this expert testimony.

**Juror Decision-Making on Future Risk**

Relatively few studies exist directly examining whether jurors are influenced by expert testimony on future risk, and even fewer explore the mechanisms by which this influence may occur. Additionally, with the exception of a handful of juror interview studies and one experimental study, all the research in this area has utilized college students as mock jurors. Although experimental jury decision-making research has generally demonstrated few differences between college student and more representative samples (Bornstein, 1999), a number of psychology and legal scholars have warned against using nonrepresentative populations (e.g., Diamond, 1997; Weiten & Diamond, 1979; Wiener, Krauss, & Lieberman, in press). The demand for more ecologically valid sample populations and realistic simulations has also been echoed by courts, which have been noticeably reluctant to consider experimental research lacking these elements (e.g., see Lockhart v. McCree, 1986). Moreover, recently researchers found a number of significant differences between college student mock jurors and a marketing firm community sample across an identical SVP hearing (McCabe, Krauss, & Lieberman, 2010).

**Juror interview studies.** One research team has examined the effects of different types of experts testimony in juror decisions in Texas SVP hearings (Meeks, Boccaccini, & Turner, 2009; Turner, Boccaccini, Meeks, & Murrie, 2009). In their survey of 161 jurors across 14 different cases, interviewed jurors did not place great weight on expert testimony on future risk in comparison to other factors, such as: the respondent’s prior crimes, the respondent’s failure to change, and expert testimony about the offender possessing a behavioral abnormality. When specifically asked about future risk expert testimony, the jurors stated that they weighted actuarial expert testimony more heavily than clinical expert testimony, but that they also rated experts that presented actuarial expert testimony as less credible. While this initial study is informative, it suffers from
a relatively small sample size and may not generalize to other jurisdictions. In addition, all juror interview studies are subject to biases associated with: (a) the nonrandom selection of participants, (b) the vagaries of memory, and (c) individuals’ general lack of awareness of the mental processes that lead to their decision (Nisbett & Wilson, 1977).

**Mock jury decision-making studies.** Experimental studies examining the effects of different types of expert testimony in SVP hearings have been more equivocal. Some research has shown that mock jurors exhibit a preference for unstructured clinical testimony, but that tendency can be overcome through an inducement to adopt a rational mindset. Yet, other research has found that mock jurors do not favor one type of testimony relative to the other.

For example, Lieberman and colleagues (2007) explored whether individuals could be motivated to be more receptive to actuarial testimony (and less influenced by clinical testimony) using Cognitive-Experiential Self-Theory (CEST; Epstein, 1994; Epstein & Pacini, 1999) as a theoretical framework. CEST maintains that information is processed by two partially independent systems—an experiential and a rational mode. The experiential mode is an affect-oriented information processing system. This system relies on intuition to effortlessly and efficiently make judgments. The second system, known as a rational mode, is a more mentally effortful system that is deliberate and analytic. Individuals base judgments on evidence and logic when in a rational mindset. CEST posits that behavior is guided by the relative contributions of these two systems, but that situational demands can motivate individuals to be more likely to adopt one system over another at any given time.

Lieberman et al. (2007; see also Krauss, Lieberman, & Olson, 2004) explored the effects of rational and experiential information processing on college student mock jurors reactions to clinical and actuarial expert testimony, by embedding directives to adopt one mindset or another in jury instructions. The researchers were able to induce male (but not female) jurors to more favorably weight expert testimony consistent with their prompt (i.e., rationally instructed jurors favored actuarial expert testimony; whereas the other jurors were more influenced by clinical testimony). The researchers surmised that their inducements may have been ineffective for women because they possessed a stronger bias toward commitment or conviction in the simulated case presented. More generally, research has found that women jurors are more favorable toward conviction than males across a variety of hearings involving sexual offenders and sexual abuse allegations (Brekke & Borgida, 1988; Bottoms & Goodman, 1994; Guy & Edens, 2003, 2006; Kovera, McAuliff, & Hebert, 1999; Schutte & Hosch, 1997).

However, several other studies of college student mock jurors using simplistic trial transcripts found few differences between mock juror decisions based on expert testimony of different scientific quality (Guy & Edens, 2003, 2006). In both of these studies jurors placed equal weight on actuarial and unstructured clinical expert testimony. Yet, in these studies, few mock jurors voted against commitment regardless of their condition, making it more difficult to easily discern the impact of the different types and quality of expert testimony in these decisions. In addition, the use of a potentially less sensitive dependent measure may have
contributed to the lack of consistency between these findings and those previously discussed.³

The only study employing a noncollege student sample of jurors in a SVP hearing revealed a significant juror bias in favor of unstructured clinical expert testimony in comparison to actuarial expert testimony (McCabe, Krauss, & Lieberman, 2010). This study directly compared college student mock jurors to a community sample recruited by a marketing firm across a 1-hr video reenactment of a SVP trial based on a trial transcript. Several substantial differences between the two samples were found including: (a) jurors in the community sample evidenced greater confidence in commitment decisions, (b) jurors in the community sample demonstrated a greater bias in favor of unstructured clinical expert testimony over actuarial expert testimony in their confidence of their commitment decisions, (c) jurors in the community sample exhibited a greater gender bias in favor of commitment (i.e., more women in the community sample were strongly in favor of commitment as compared to the men), and (d) the college student sample evidenced a greater likelihood of processing information in a more rational manner. This last finding was based on the administration of a CEST-derived individual differences processing style measure—Rational versus Experiential Inventory (RVEI; Pacini & Epstein, 1999). The RVEI measures an individual’s intrinsic motivation to process information in the rational or experiential mode. The researchers suggested that this cognitive processing style difference (i.e., college students scored higher on the RVEI’s rational processing scale than did recruited jurors) may be responsible for the finding that clinical expert testimony was less influential relative to actuarial testimony for the college student sample compared to the community sample of jurors. The researchers also found that the student’s two RVEI subscales were associated with verdict preference, but this was not the case for the community sample. Among the students, being high on the rational scale and low on the experiential scale (i.e., a more rational processing) was related to fewer and less confident commitment decisions, while being lower on the rational scale and higher on the experiential scale (i.e., more experiential processing) was associated with greater and more confident commitment decisions.

Although McCabe et al. (2010) used a sample recruited from the community, it is unclear whether this sample was truly representative of individuals actually summoned to court for jury service. For example, participants were recruited by a marketing firm, willing to voluntarily travel to a university campus to participate, and paid for their time.

The Present Research and Hypotheses

More research is clearly needed in SVP adjudications given the important implications of future dangerousness testimony, and because jurors in these cases may be placing more weight on less scientifically valid expert testimony. Con-

³ In both the Guy & Edens (2003/2006) experiments, the researchers simply used percentage of convictions as their dependent measure. It is likely that this binary dependent variable is less sensitive to effects of expert testimony than the continuous verdict composite variable used by Lieberman et al., 2007.
sequently, the current study was designed to extend our knowledge in this area using more realistic materials and methodology than has typically been used, and to test the findings of McCabe et al. (2010) that cognitive processing style differences are responsible for dissimilarities between more representative and college student mock juror samples.

This study examines three central hypotheses generated from the existing literature on juror decision-making in SVP cases. First, it is hypothesized that, consistent with McCabe et al., (2010), venire jurors will be more influenced in their eventual decision by less scientifically valid unstructured clinical expert testimony than by more scientifically valid actuarial expert testimony. Second, it is predicted that the current study’s juror participants will be more similar to the community sample in McCabe et al. (2010) and score lower on the rational portion of the RVEI than did that study’s student sample. Further, those jurors who score high on the RVEI’s (Pacini & Epstein, 1999) rational scale and low on the experiential scale will be examined to determine if they mimic previous college student mock jurors samples or McCabe et al.’s community samples. In the college student sample, participants high on rational and low on experiential processing (i.e., rationals) demonstrated lower verdict-confidence (a leniency effect) relative to participants who were low on the rational scale and high on the experiential scale (i.e., experientials), while McCabe et al.’s community sample did not evidence this relationship. Finally, it is expected that female jurors will be more likely to commit the respondent than male jurors. Since previous research performed in this context has been limited to college student samples that lacked demographic diversity, additional analyses of other demographic characteristics that might affect verdicts was also undertaken.

**Method**

**Participants**

One-hundred and fifty six jury eligible citizens from Southern California who were called for and arrived for jury duty, but who were not seated on a jury, participated in the study. These jurors were not called for any cases and were eventually excused for the day by the court. Members of this unused jury venire were asked if they wanted to participate in a research study on jury decision making. Participants were offered nominal compensation ($25) for their participation and were treated in accordance with American Psychological Association guidelines for the treatment of human subjects. The participants’ demographic information is displayed in Table 1.

**Procedure and Stimulus Materials**

Testing sessions were randomly assigned to condition and participants were tested in groups of between three and 26.\(^4\) Participants were shown a 1-hr

---

\(^4\) Since participants for each day were shown the same stimuli, each day was randomly assigned to condition, and then days were assigned to conditions to equate groups within the experiment. There is no reason to believe that participants coming to jury duty on different days are not randomly distributed.
Table 1

Juror Participants’ Demographic Information (N = 156)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (SD)</td>
<td>40.9 (15.5)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>74</td>
<td>47.4</td>
</tr>
<tr>
<td>Female</td>
<td>82</td>
<td>52.6</td>
</tr>
<tr>
<td>Marital Status¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Married</td>
<td>66</td>
<td>42.3</td>
</tr>
<tr>
<td>Married</td>
<td>62</td>
<td>39.7</td>
</tr>
<tr>
<td>Divorced</td>
<td>24</td>
<td>15.4</td>
</tr>
<tr>
<td>Widowed</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>Have Children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>80</td>
<td>51.3</td>
</tr>
<tr>
<td>No</td>
<td>76</td>
<td>48.7</td>
</tr>
<tr>
<td>Number of Children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>75</td>
<td>48.1</td>
</tr>
<tr>
<td>One</td>
<td>20</td>
<td>12.8</td>
</tr>
<tr>
<td>Two</td>
<td>34</td>
<td>21.8</td>
</tr>
<tr>
<td>Three</td>
<td>17</td>
<td>10.9</td>
</tr>
<tr>
<td>Four</td>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>Five</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>Number of Children under 18 Years¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>85</td>
<td>55.2</td>
</tr>
<tr>
<td>One</td>
<td>17</td>
<td>11.0</td>
</tr>
<tr>
<td>Two</td>
<td>30</td>
<td>19.5</td>
</tr>
<tr>
<td>Three</td>
<td>12</td>
<td>7.8</td>
</tr>
<tr>
<td>Four</td>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>Five</td>
<td>2</td>
<td>1.3</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>79</td>
<td>50.6</td>
</tr>
<tr>
<td>Asian</td>
<td>32</td>
<td>20.5</td>
</tr>
<tr>
<td>Hispanic/Latino/Latina</td>
<td>30</td>
<td>19.2</td>
</tr>
<tr>
<td>Black/African-American</td>
<td>7</td>
<td>4.5</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some High School</td>
<td>3</td>
<td>1.9</td>
</tr>
<tr>
<td>High School or GED</td>
<td>10</td>
<td>6.4</td>
</tr>
<tr>
<td>Trade School</td>
<td>9</td>
<td>5.8</td>
</tr>
<tr>
<td>Some College</td>
<td>60</td>
<td>38.5</td>
</tr>
<tr>
<td>College Graduate</td>
<td>48</td>
<td>30.8</td>
</tr>
<tr>
<td>Post Graduate</td>
<td>26</td>
<td>16.7</td>
</tr>
<tr>
<td>Political Affiliation²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Republican</td>
<td>56</td>
<td>37.1</td>
</tr>
<tr>
<td>Democrat</td>
<td>56</td>
<td>37.1</td>
</tr>
<tr>
<td>Independent</td>
<td>28</td>
<td>17.9</td>
</tr>
<tr>
<td>Libertarian</td>
<td>6</td>
<td>4.0</td>
</tr>
<tr>
<td>Green Party</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>2.6</td>
</tr>
<tr>
<td>Yearly Income¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0–$24,999</td>
<td>22</td>
<td>14.3</td>
</tr>
<tr>
<td>$25,000–$49,999</td>
<td>40</td>
<td>26.0</td>
</tr>
<tr>
<td>$50,000–$74,999</td>
<td>22</td>
<td>14.3</td>
</tr>
<tr>
<td>$75,000–$99,999</td>
<td>28</td>
<td>18.2</td>
</tr>
<tr>
<td>$100,000–$124,999</td>
<td>15</td>
<td>9.7</td>
</tr>
<tr>
<td>Over $125,000</td>
<td>27</td>
<td>17.5</td>
</tr>
</tbody>
</table>

¹Two participants did not provide complete data on this variable. ²Five participants did not provide complete data on this variable.
videotaped reenactment of an actual Arizona SVP civil commitment trial. Although the transcript was abbreviated by necessity, every effort was made to keep the simulation as faithful as possible to the original case. In addition, the actors portraying the attorney representing the state (the petitioner) and the respondent’s attorneys were both defense attorneys who formerly worked as prosecuting attorneys in Arizona. The actor portraying the expert psychological witness, who depending on condition provided either clinically based or actuarially based future risk expert testimony, was a licensed psychologist in Arizona who frequently provided expert testimony to the courts in his jurisdiction. Roughly equal numbers of participants heard either clinically based (n = 80) or actuarially based testimony (n = 76).

Rational versus Experiential Inventory (RVEI). Participants were administered the Rational versus Experiential Inventory (RVEI; Pacini & Epstein, 1999; Epstein, Pacini, Denes-Raj & Heier, 1996). The RVEI is a 31 item scale that combines a slightly abbreviated version of Need for Cognition scale (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984) to measure propensity for rational processing and a Faith in Intuition scale. The Faith in Intuition scale was designed to assess individuals’ propensity to use experiential processing by measuring the extent to which they use and trust their intuition. The RVEI has been validated in several previous research studies (Epstein et al., 1996; Pacini & Epstein, 1999). This research has shown that heuristic processing tends to be primarily determined by the Faith in Intuition scale, that the main subscales are independent and inversely related to one another, and that the scales contribute to prediction beyond the Big Five personality scales on a variety of measures (e.g., categorical thinking, conservative ideology, punitive attitudes toward criminals).

RVEI scores, for both the rational and experiential scales, were calculated for each participant. The mean score on the rational scale for the juror sample was 71.60 (SD = 9.70; minimum = 33; maximum = 95), while the mean score on the experiential scale was 41.05 (SD = 5.80; minimum = 26; maximum = 59). Both scales’ distributions were approximately normal. The scales, both together and separately, were found to be reliable (RVEI scale overall, α = .77; for the rational scale, α = .81, for the experiential scale, α = .73).

After completing the RVEI, participants viewed the stimulus video which consisted of, in order, (a) the judge’s pretrial jury instructions, (b) opening statements from the petitioner’s and then the respondent’s attorney, (c) expert testimony (either clinical or actuarially based), (d) cross-examination by the respondent’s attorney, (e) closing arguments from the petitioner’s and respondent’s attorney, and (e) final instructions from the judge. The presentation contained information about the respondent including that he had served a sentence for charges relating to exposing himself, compelling two minor girls to masturbate him, compelling one minor girl to have oral sex with him and showing her hardcore pornography. In addition, it included information regarding a previous conviction for molesting two neighbor minor girls, for which the offender previously served five years of probation. Finally, the presentation included information about the respondent having completed a 2-year-long sex offender treatment program while incarcerated. Identical film footage and content was used for all aspects of the trial (e.g., judicial instructions, opening statements,
etc.) except for small segments of the expert testimony and cross-examination detailed below.

**Expert testimony.** The expert witness testified for the petitioner. In both the clinical and actuarial conditions the expert testified to the following: (a) the results of a two to three hour interview with the respondent, a review of the respondent’s sexual history, whether the respondent suffered from any paraphilias, his past convictions for child molestation, prior treatment, and several other factors; (b) his opinion that the respondent was a pedophile; (c) that the respondent had unspecified deficiencies in completing a sex offender treatment program in prison; and (d) his belief that the respondent was likely to recidivate. Identical film footage was used for this portion of expert testimony regardless of condition. Only the rationale for the expert’s opinion (experience as a clinician or use of actuarial instruments) was varied across conditions.

The clinical expert testified that his opinion was based on his interview with the respondent, while the actuarial expert testified that his opinion was based on his use of two actuarial instruments designed to predict recidivism in sex offenders. The first of these was the Rapid Risk Assessment of Sexual Offender Recidivism (RRASOR; Hanson, 1997) and the second was the Sexual Offender Risk Appraisal Guide (SORAG; Rice & Harris, 1995). The language used by the expert to describe the two instruments was based on the actual Arizona hearing on which the video was based. The clinical expert testimony was based on clinical expert testimony found in another available Arizona SVP trial transcript. Notably, the conclusion drawn by the expert in both conditions did not vary. The actuarial expert testimony only differed from the clinical expert testimony for approximately 350 words of the 1,500 word videotape segment in which the actuarial instruments were described. Identical film footage showed the expert in both conditions conclude by saying, “The law asks us to make a determination as to whether a person is likely or not likely [to recidivate] and my opinion is that he is likely to.”

The two direct examination conditions are excerpted in Appendix A.

**Cross-examination.** Following the expert testimony, the respondent’s attorney cross-examined the witness. In both expert testimony conditions identical

---

5 Based on the limited available data, it is not unusual for only the petitioner to present expert testimony. In only one of the fourteen trials in the Meeks et al. (2009) juror interview study did both sides present expert testimony on risk.

6 The RRASOR is a brief, 4-item instrument used to assess the risk of recidivism for male sex offenders with at least one prior offense. These items were later incorporated in the Static-99 (Hanson & Thornton, 2000). Although currently the Static-99 is more widely used than the RRASOR (Hanson, 2006), in order to maintain ecological validity and stay faithful to the transcript upon which the stimulus was based, the RRASOR was used.

7 The SORAG is a 14-item instrument. Ten of these 14 items are identical to the items in the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 1998), an instrument use to predict recidivism for general, violent, and sexual offenses. The additional items of the SORAG pertain to factors shown to be important to recidivism for sex offenders. Both actuarial instruments have evidenced comparable predictive accuracy to other available actuarial measures (Hanson & Morton-Bourgon, 2009).

8 In the actual case, the expert offered some percentages of the likelihood of recidivism for the respondent based on the actuarial instruments used, but this information was removed so that the two expert conditions would provide identical as possible information relating to their opinion.
footage showed the attorney questioning whether the respondent’s participation in a sexual offender treatment program had any impact on his diagnosis and whether just a two to three hour interview was sufficient to draw conclusions. The respondent’s attorney also individually pointed out limitations associated with the techniques used by the expert in the two expert testimony conditions. In the actuarial condition, the expert admitted that the instruments could only provide estimates for offenders in general and his recidivism interview was based on a relatively short interview. In the clinical condition, the expert admitted that the expert’s opinion was not based on any information other than a relatively short interview, and that his clinical judgment might be flawed. Three counterarguments to the experts’ testimony were advanced by the respondent’s attorney in both expert testimony conditions, and each argument was based on cross-examinations from Arizona SVP civil commitment cases which had used the two types of expert testimony. The two cross-examination conditions are excerpted in Appendix A.

Dependent measures. Participants provided two assessments of the dangerousness of the respondent, after both opening statements (Time 1) and after the judge’s final instructions (Time 2). Participants were asked to answer the question of whether beyond a reasonable doubt “the respondent was a sexually violent person, and therefore should be committed.” Participants responded either yes (coded +1) or no (coded −1) and then indicated their confidence on a scale ranging from 1 (not very confident) to 9 (extremely confident). A composite variable was created by multiplying the not commit/commit assessment (−1 or 1) by the confidence measure. The resulting variable ranged from −9 (extremely confident the respondent should not be committed) to +9 (extremely confident the respondent should be committed).

Jurors were next asked to assign percentages to three factors that may have influenced their decision on whether or not the respondent presented a future risk. These were his past offenses, the offense for which he is currently incarcerated, and the expert testimony. Participants were instructed that these three percentages must total 100%. In addition, participants provided ratings of the expert testimony. Participants rated the expert testimony on a 9-point Likert-type scale (1 = not at all, 5 = moderately, and 9 = highly) across four dimensions: influential, scientific, confident, and credible. Finally, participants completed a series of multiple choice questions concerning the hearing including a question regarding the expert witness’ rationale for his opinion.

Results

Manipulation Checks

Expert testimony. Examination of posttrial questions revealed that most participants paid close attention to the video. A relatively small number of participants were unable to correctly identify which type of testimony they had heard (30 out of 156), either based on the expert’s training and experience (clinical) or on a statistical instrument (actuarial), while two participants failed to respond to the question. Given that jurors in an actual case may misremember important details of evidence or testimony presented and sensitivity analyses
indicated that, with one exception,\(^9\) excluding jurors who failed these checks did not materially affect the results, no jurors were excluded on the basis of the manipulation checks.

**Initial Verdicts, Demographic Data, Confidence Ratings, and RVEI Scores**

After opening statements, jurors made their first assessment of their verdict and their confidence in that verdict. Initially, 57% of jurors voted for commitment of the respondent, but after the presentation of expert testimony, cross-examination, and final judicial instructions this number increased to 82%. Greater verdict change occurred among clinical group members (34% increase) than in the actuarial group (15%).

**Demographic data and initial verdicts.** A hierarchical multiple regression analysis was performed using all demographic and individual difference variables to predict verdict-confidence at Time 1. These variables included: gender, age, number of children cared for, education level, and yearly income.\(^10\) Because gender has been shown to affect verdict decisions in cases involving sex offenders (Brekke & Borgida, 1988; Bottoms & Goodman, 1994; Guy & Edens, 2003, 2006; Kovera, McAuliff, & Hebert, 1999; Schutte & Hosch, 1997), gender was entered on step 1. All other variables were entered on step 2. Thus, all other variables were assessed statistically controlling for the impact of gender. Gender approached significance as a predictor of verdict-confidence at Time 1 with female’s verdict-confidence ratings trending higher than males (\(\beta = 1.89, t = 1.79, p = .08\)). While controlling for gender, the only other variable to reach significance was the dichotomous race variable with white/Caucasian participants evidencing greater leniency in verdict-confidence at Time 1 (\(\beta = 3.07, t = 2.64, p = .009, R^2\) Change = 4.4%).

Prior to the presentation of the manipulated expert testimony the two groups differed in their verdict-confidence assessment, with a one-way (actuarial vs. clinical expert testimony) Analysis of Variance (ANOVA) finding significant differences, \(F(1, 154) = 8.06, p < .05\), \(\eta^2 = .05\) (Actuarial \(M = 3.19, SD = 6.18\); Clinical \(M = 2.28, SD = 6.65\)). A Pearson’s chi-square analysis revealed no significant differences in the distribution of female participants between testimony conditions (\(\chi^2(1) = .11, p > .73\)). A Pearson chi-square analysis found no significant differences in the distribution of participant race across testimony conditions (\(\chi^2(1) = .63, p < .43\)).

\(^9\)In one multiple regression analysis, a dichotomous race variable (white or non-white) was a significant predictor of initial verdict confidence at Time 1 (\(p = .009\)), with whites being significantly more lenient. After excluding those who could not correctly identify the type of testimony they heard, this predictor only trended toward significance (\(p = .07\)). For all other tests, significant and non-significant results remained similar with the exclusion of those who could not correctly identify the type of testimony they heard.

\(^{10}\)In order to reduce potential redundancy (or multicollinearity) that could threaten the validity of the results, only the number of children variable was used, instead of the two highly related variables: whether the participant had children and if those children were under eighteen years of age. In addition, three categorical variables were recoded into dichotomous variables due to very low frequencies in some categories. These were marital status (unmarried or married/divorced/widowed), political affiliation (liberal, Democrat/Green, or not liberal, Republican/Independent/Libertarian), and race (white/Caucasian or non-white/Caucasian).
conditions \( \chi^2(1) = .027, p > .87 \). Because the main variable of interest is the effect of expert testimony, consistent with previous research jurors’ initial verdict assessments were used as a covariate in subsequent analyses.\(^{11}\)

**RVEI scores and initial verdicts.** RVEI scores also differed initially between the two testimony conditions as well. An independent samples \( t \) test revealed significantly lower scores on the experiential scale in the clinical testimony condition relative to the actuarial condition, \( t(151) = -2.50, p = .01, d = .40 \). There were no statistically significant differences between groups on the rational scale \( (p = .80) \). Median splits were then performed on both subscales’ scores. Thirty-six participants were found to be high on the rational scale and low on the experiential scale \( (rationals) \), while 28 were low on the rational scale and high on the experiential scale \( (experientials) \). Of these 64 total participants, 33 heard actuarial testimony \( (18 \quarex{experientials} and 16 \quarex{rationals}) \); however, of the 31 who heard clinical testimony, 10 were experientials, but 21 were rationals. This overrepresentation of rationals in the clinical testimony condition approached traditional significance levels \( \chi^2(1) = 3.23, p = .07 \).

Jurors’ rational and experiential scores were also compared to those of undergraduate mock jurors in McCabe, Krauss, and Lieberman (2010) in an independent samples \( t \) test. Contrary to prediction, no significant differences were found between the mock jurors’ RVEI score in this study and those from McCabe et al. (2010) on either the rational scale \( (p = .39) \) or the experiential scale \( (p = .65) \).

**Final Verdict, Demographics, Confidence Ratings, and RVEI Scores**

**Final verdict.** A binary logistic regression was performed on final dichotomous verdict using testimony type \( (actuarial \ or \ clinical) \) and gender as predictors while controlling for initial verdict-confidence. Testimony type trended toward significance \( (Wald = 3.25, df = 1, p = .07) \). Gender was not a significant predictor of final verdict, controlling for initial verdict-confidence, in this analysis \( (Wald = .15, p > .70, ns) \). Unremarkably, initial verdict-confidence was a significant predictor of final verdict \( (Wald = 22.49, df = 1, p < .01, \text{odds ratio} = 1.39) \).

**Final verdict, confidence, and demographics.** A hierarchical multiple regression analysis was performed to assess the impact of demographic factors and testimony on final verdict confidence when verdict confidence at Time 1 was controlled for. Verdict confidence at Time 1 was entered in step 1. All demographic variables, including the dichotomous marital status, political affiliation, and race variables, were entered on step 2, and testimony condition was entered on step 3. Results demonstrated that even when controlling for initial verdict

\(^{11}\) There are a number of different ways to adjust for the failure of equivalency between groups prior to manipulations. A covariate statistically controls for any pre-existing initial group differences due to the co-varied variable, and is the preferred method of analysis where initial group differences are likely due to chance or measurement issues. Another method would be to use difference scores subtracting the first assessment from the second. Generally, difference scores include more error in analyses than a covariate \( (Cook \ & \ Campbell, 1979) \), and have been found to evidence less reliability. Still, all analyses were also performed with difference scores and results did not differ materially.
confidence and all demographic variables, testimony type was still a significant predictor of final verdict confidence ($\beta = -1.77$, $t = -2.14$, $p = .03$, $R^2$ Change = 2.1%).

An Analysis of Covariance (ANCOVA) was performed on the continuous final verdict-confidence variable as the dependent variable and testimony type (actuarial or clinical) and gender as factors, while covarying initial verdict-confidence and RVEI scores. In this analysis, testimony type was a significant predictor of final verdict-confidence ($F(1, 152) = 4.86$, $p = .03$, $\eta^2 = .03$) with jurors that received clinical expert testimony ($M = 5.78$, $SE = .55$) significantly more confident in their verdict ratings than those who received actuarial expert testimony ($M = 4.00$, $SE = .57$). Gender was again not a significant predictor of final verdict-confidence ($p > .48$). And, again unremarkably, initial verdict-confidence was a significant predictor of final verdict-confidence ($F(1, 152) = 73.94$, $p < .01$, $\eta^2 = .34$).

**RVEI effects and final verdicts.** A series of analyses was performed on participants’ RVEI scores to compare this study’s findings to that of McCabe, Krauss, and Lieberman (2010). A hierarchical multiple regression was performed to test if participants’ rational and experiential scale scores were predictive of final verdict-confidence, controlling for initial verdict confidence. Initial verdict-confidence was entered at step 1 and rational and experiential scale scores were both entered at step 2. Neither rational scores nor experiential scores were predictive of final verdict-confidence, controlling for initial verdict-confidence (all $ps > .29$).

An ANCOVA was performed on final verdict-confidence using the cognitive processing style variable (rational or experiential, $n = 64$) and testimony type (actuarial or clinical) and their interaction, while controlling for initial verdict-confidence. The rational/experiential variable and testimony type produced a significant interaction, $F(1, 59) = 7.21$, $p < .01$, $\eta^2 = .11$; however, in the opposite direction of that anticipated. *Rationals* who heard clinical testimony ($M = 7.67$) had higher verdict-confidence than those who heard actuarial testimony ($M = 3.43$). Also, *experientials* who heard actuarial testimony ($M = 5.94$), had higher verdict-confidence than those who heard clinical testimony ($M = 4.25$).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Gender</th>
<th>Covaried means</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td>Males</td>
<td>5.69</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>5.88</td>
<td>.75</td>
</tr>
<tr>
<td>All Clinical jurors</td>
<td></td>
<td>5.78&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.55</td>
</tr>
<tr>
<td>Actuarial</td>
<td>Male</td>
<td>3.58</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>4.42</td>
<td>.76</td>
</tr>
<tr>
<td>All Actuarial jurors</td>
<td></td>
<td>4.00&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.57</td>
</tr>
<tr>
<td>Actuarial + Clinical</td>
<td>Male</td>
<td>4.63</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>5.15</td>
<td>.53</td>
</tr>
</tbody>
</table>

<sup>a</sup> Comparison is significantly different at $p < .05$. (covariate in model initial confidence × verdict ratings = 1.65, rational scale = 71.83, experiential scale 41.05).
Perceptions of the Expert

**Evaluations of the expert testimony.** Jurors made evaluations of the experts based on four distinct factors (influence, confident, scientificness, and credibility). Because assessments of how scientific the testimony was and the other assessments of influence, confidence, and credibility seemingly tap distinct constructs, these last three scores were analyzed separately (all r’s > .72, p < .05). Participant’s scores for the expert’s testimony on influence, confidence, and credibility were found to be highly reliable (α1 = .91). As a result, these three scores were averaged to form a composite variable. The mean score for expert rating composite variable was 6.41 (SD = 1.71; minimum = 1; maximum = 9). When asked to rate how scientific the testimony was, the participants’ mean score was 5.71 (SD = 1.94; minimum = 1; maximum = 9). An independent samples t test using testimony types as the grouping variable found no significant differences on both the expert rating composite variable scores (clinical testimony, M = 6.50, actuarial testimony, M = 6.36; t(154) = .53, p > .60) and the scientificness scores (clinical testimony, M = 5.71, actuarial testimony, M = 5.74; t(154) = .08, p > .94).

A 2 (expert testimony) × 2 (gender) Multivariate Analysis of Variance (MANOVA) was conducted on jurors’ perception of the composite expert variable (confidence, influence, and credibility) and jurors’ perception of the level of science conveyed by the expert testimony. A main effect was found for gender with regard to the composite expert variable, F(1, 152) = 4.67, p < .05, η^2 = .03 such that female jurors perceived the experts more favorably than male jurors (female M = 6.67, SE = .19, male M = 6.12, SE = .20). This relationship did not exist for female jurors’ perception of the level science the expert conveyed, nor was there a significant main effect for expert testimony type or an interaction between gender and expert testimony (all ps > .10).

Jurors’ Assessment of Influential Factors

The weight jurors reported they placed on the respondent’s past, most recent offense, and the expert testimony in their decision-making was also examined. Jurors were to assign percentages to these variables that added to 100%. Overall, of the jurors who correctly completed this item (N = 138), the average of the 100% apportioned by the jurors to the most recent offense in their decision making was 21%, whereas 47% was apportioned to the respondent’s past offenses, and 32% was apportioned to the expert testimony that was proffered.

A 2 (expert testimony) × 2 (gender) MANOVA was performed on jurors’ assessments of the weight they placed on the respondent’s past criminal acts, most recent crime, and the expert testimony in their decision-making. No significant main effects or interactions were discovered based upon juror gender, expert testimony they received, or an interaction between these two factors (all ps > .10).12

---

12 These results of this analysis were similar regardless of whether the assessments of all jurors were used or only the jurors (N=138) whose evaluations correctly added to 100%.
Discussion
The primary goal of this research was to examine the effects of expert testimony of differing scientific quality on the decision-making of jurors in an SVP hearing using a highly representative sample of jurors. It was hypothesized that jurors would problematically place greater weight on less reliable expert testimony when compared to more accurate expert testimony in their decision-making. Although past research has sometimes (McCabe et al., 2010), but not always (Guy & Edens, 2003, 2006), found that poorer quality expert testimony was more influential in mock jurors’ SVP decisions than more scientifically valid testimony, no study to date has completed what Diamond (1997) would characterize as Stage Two research (i.e., research on representative juror samples with a realistic simulation). This study demonstrates that a representative juror sample favors less scientifically valid unstructured clinical expert testimony over more accurate actuarial assessment in a realistic mock SVP hearing. This favoritism occurred even though a substantial majority of jurors favored commitment (82%) regardless of condition, indicating that even in cases where commitment is likely, expert testimony still has an important effect. In essence, a fundamental disconnect in juror decision-making appears to exist in this context—jurors are more confident in their verdicts when presented with less accurate expert testimony. These findings are of particular importance because they have been obtained using a highly realistic methodology that builds on previous research findings. Given the reluctance of courts to consider social science findings, particularly when research is conducted in an artificial manner, it is critical that research demonstrate that findings regarding jurors over reliance on clinical testimony generalize to a representative sample drawn from actual venires (Diamond, 1997; Wiener et al., 2011).

Further, consistent with sizable group of research studies examining both college student and representative mock juror decisions involving comparisons of actuarial and clinical expert testimony on future dangerousness (Krauss & Sales, 2001; Krauss & Lee, 2003; Lieberman et al., 2007; McCabe et al., 2010), jurors’ self-report of the weight they placed on the expert testimony they received—approximately one third of the reason for their decision in both conditions. This finding lends support to the notion that jurors may be unaware of their bias in favor of clinical expert testimony, or at least, unable to accurately report on their decision-making process. A large body of psychological research argues that such ignorance of one’s own decision-making is far from unusual (Nisbett & Wilson, 1977). Lastly, this result suggests a concern for the collection of juror interview data in this context because jurors may not be able to accurately report on the actual factors that influenced their decision-making. Yet, consistent with past jury interview research in SVP hearings in Texas (Meeks et al., 2009), our mock jurors placed more weight on the respondent’s past and most recent crime. This may indicate

---

13 The limited variance in jurors’ final verdicts regardless of condition may also explain why the less sensitive binary measure of verdict only trended towards significance and the more sensitive confidence verdict measure was significant.
that expert testimony, while important in juror decisions in SVP cases, may not be as important as other pieces of evidence, and that jurors may be making judgments based on their own naïve theories of what factors (e.g., heinousness of past crimes) are associated with greater future risk. To some extent, this can be viewed as jurors making their own “clinical judgments” about what factors will predict future risk, and the weight to give those factors, based on their own experiences.

Cognitive Processing Style Differences in Evaluation of Expert Testimony

A secondary goal of this research was to explore individual differences in cognitive processing style as a possible mechanism for jurors’ bias in favor in unstructured clinical expert testimony. Previous research by McCabe and colleagues (2010), based on Cognitive-Experiential Self-Theory (CEST; Epstein, 1994; Epstein & Pacini, 1999), suggested that college student mock jurors who were assessed as more rational in their thinking style on the RVEI (Pacini & Epstein, 1999) would be more likely to favor actuarial expert testimony while those who more heavily favored experiential processing would be more influenced by clinical expert testimony. Results did not support this relationship for juror venire participants. In fact, jurors rated as high in an experiential and low in rational processing were significantly more influenced by actuarial expert testimony while jurors low in experiential processing favored clinical expert testimony. This may reflect a spurious finding. There were more rationals in the clinical testimony condition, and the reported effect may simply be a result of the clinical expert testimony’s appeal overwhelming the individual cognitive processing variable among these participants.

Alternatively, it may be that ratings of an individual’s cognitive processing style, especially with more representative juror samples, are outweighed by a number of other more important contextual variables when applied to specific decisions in a SVP hearing. It is not unusual for personality factors to exert a relatively small influence on verdict decisions, particularly when general factors are used (e.g., dogmatism, locus of control) as opposed to case specific personality or attitudinal factors (e.g., attitudes toward the death penalty; Lieberman & Sales, 2007). In addition, and again contrary to prediction, no differences were found between the processing styles of the more ecologically valid sample in this study and the undergraduate student sample in McCabe et al. (2010). Further, jurors overall perceptions of evidence strength may override any individual differences (Visher, 1987). Finally, it is possible that if the evidence was more equivocal, the effect of individual differences would be more powerful (Moran, Cutler, & DeLisa, 1994).

Very limited support was found for the proposition that those who were high in rational and low in experiential processing (and vice versa) differed on verdict-confidence ratings and were differentially affected by divergent forms of expert testimony. Using the same analysis as McCabe et al., we found a significant relationship between processing style and verdict-confidence only before participants heard either actuarial or clinical testimony (i.e., rationals were less punitive). After hearing the testimony, the relationship was no longer present. Again, this finding suggests a significant difference between our representative juror
sample and previous college student samples with regard to the effects of processing style.

This does not mean that CEST-based manipulations (e.g., the CEST jury instructions from Lieberman et al., 2007) may not serve as an effective intervention to counteract bias associated with different forms of expert testimony in different sample populations. It is important that jurors give adequate attention to more reliable actuarial-based testimony. Encouraging jurors to process information in a more rational and analytic manner may be a more effective means to counter clinical testimony’s persuasive advantage, as opposed to attempting to remove experientially oriented jurors through *voir dire* challenges. Additional research on the RVEI’s assessment of cognitive processing style, its stability across age, and its relationship to specific decision-making tasks needs to be conducted with ecologically valid samples in additional Stage Two research.

**The Effects of Juror Gender and Demographics on Decisions**

Another goal of this study was to more fully examine the effects of jurors’ gender and demographic characteristics on their decision-making in an SVP hearing. Using a sample of jurors selected from an actual jury pool, it was assumed that greater insight into the factors that affect their decisions might be gained. Somewhat consistent with our hypothesis (i.e., the gender effect approached but did not reach significance in our sample) and a growing body of research generally (Schutte & Hosch, 1997) and specifically in SVP trials (McCabe et al., 2010), female jurors evidenced a greater tendency to recommend civil commitment for a sexual offender. They also rated the expert testimony they received more highly than their male juror counterparts. This favoritism for the expert could either reflect that female jurors were more influenced by the expert testimony in the case or it could indicate that they justified their preexisting commitment beliefs by later self-reporting a more positive impression of the expert. Future research on this issue is necessary to determine which of these two alternatives or some combination of both is a more accurate explanation of female jurors’ behavior. No other characteristics of female jurors easily explained either of these relationships. For example, having or caring for children was not predictive of initial verdicts for female jurors, or for male jurors.

Juror race also demonstrated an important initial effect on juror commitment decisions. After opening statements, minority jurors were significantly more likely to believe the respondent should be committed, but their increase in verdict-confidence after hearing expert testimony, cross-examination, and receiving judicial instructions was similar to that of nonminority jurors. Beyond juror gender and race, no other individual demographic factor (i.e., age, education level, socioeconomic status, etc.) was predictive of initial or final verdicts. This null result is not entirely surprising given the overall weak relationships between demographics and juror decisions in a wide variety of legal contexts (Lieberman & Sales, 2007).

**Limitations of Research**

As with any simulated trial and mock juror decision-making study, this research suffers from a number of limitations. Although every attempt was made to create a realistic SVP trial stimulus (e.g., basing it on a real transcript,
employing attorneys for all the roles, including real jury instructions, etc.) and to foster realistic juror decision-making (e.g., using a real jury pool sample, completing the research at a court facility, etc.), a number of factors simply cannot be accounted for in this paradigm. These include but are not limited to: (a) the lack of *voir dire* prior to juror selection, (b) the abbreviated nature of the trial (most of these types of SVP trials last at least 2–3 days), (c) compensated, volunteer participants, (d) a lack of a deliberations phase, (e) multiple juror ratings of verdict and confidence in verdict, (f) an unusual jury population in terms of education (50% possessed college degrees) and socioeconomic status (40% earned under $25,000 and 27% earned over $100,000), and (g) perhaps, most importantly, the jurors knew they were not deciding a case that had real repercussions for the individuals involved. Even with all these problems, all juror participants appeared to take their responsibilities very seriously. There were no reported issues with jurors talking to other participants during the presentation of the trial video, and all participants completed their paperwork diligently with the vast majority of participants accurately identifying the type of testimony they were presented with during the hearing.

Additionally, because the case stimuli was based on an Arizona SVP hearing and Arizona law, the results of this study may not generalize to other jurisdictions and juror populations. Further, the generalizability of these results is clearly affected by the particular case facts presented. The case was selected because the crimes committed were not perceived to be as heinous as other Arizona SVP transcripts. As a result, in this case the effects of expert testimony may be more pronounced than in a more severe case where the expert testimony might have little effect on the jurors’ already strong views toward civil commitment. Further, the use of a confidence and verdict composite variable instead of a simple binary variable may also overstate the effects of the expert testimony. Yet, it is important to remember that no deliberations occurred in this case, and significant changes in juror confidence may still substantially impact final verdict after deliberations occur (e.g., Krauss & Lee, 2003). Even with all these limitations, however, the results of this study offer a more accurate assessment of the specific effects of different forms of expert testimony in a SVP hearing with most other important

---

14 Arizona law is similar to ten other jurisdictions and uses beyond a reasonable doubt as its burden of proof in SVP hearings. The remaining ten jurisdictions employ clear and convincing evidence in their adjudications. Although it is possible such a legal distinction could have important effects in a particular case, existing psychological research has demonstrated subtle differences in law very rarely have significant effects on decisions (see e.g., Finkel, 1991; Lieberman & Sales, 1997).

15 Some might argue that the study does not offer a true test of expert testimony of different scientific quality because the cross-examination between conditions was not identical, and the results could simply be the result of the cross-examinations employed. While acknowledging this limitation, the cross-examinations were taken from SVP trial transcripts and it seemed illogical to use identical cross-examinations for different expert content (but see Krauss & Sales, 2001, where identical cross-examinations produced significant and similar differences between clinical and actuarial expert testimony). Moreover, even if the cross-examination rather the expert testimony was the cause of the juror differences, from a practical standpoint the result is the same—under ecological valid conditions, jurors are more influenced by less scientific clinical expert testimony after a standard cross-examination, than they are by more scientific actuarial expert testimony after a standard cross-examination.
factors being controlled for, and they provide much needed data on the relationship between different types of expert testimony and juror decisions.

Policy Implications and Conclusions

Given the convergent validity of findings indicating that jurors in SVP hearings are more likely to be influenced by less accurate expert testimony and the fact that these decisions have important impacts on the liberty interests of the individuals involved (i.e., commitment for an indefinite period), this problem has important legal and policy implications. Potentially, courts could bar unstructured clinical expert testimony from being presented at SVP trials. Unfortunately, past judicial practices suggest that this is unlikely to occur because judges exhibit an affinity similar to jurors for unstructured clinically based expert testimony (Janus & Prentky, 2003). This deferential practice, however, may be changing.

Special evidentiary rules govern the admissibility of expert testimony. The federal system as well as many state jurisdictions have adopted expert testimony rules that focus admissibility decisions on the scientific validity of the expert information being presented, and this standard offers a flexible list of factors judges can consider in reaching a determination, including: falsifiability, peer review, error rate, and general acceptance (Daubert v. Merrell Dow Pharmaceuticals, 1993; General Electric v. Joiner, 1998; Kumho Tire v. Carmichael, 1999). The remainder of state jurisdictions determine the admissibility of expert testimony based on some variant of the Frye standard (see Frye v. United States, 1923), which holds that expert evidence is admissible if it is generally accepted in the field from which it originates (see Frye v. United States, 1923). At their heart, both these admissibility frameworks are concerned with prohibiting expert testimony that is not truly “expert” from entering the courtroom, and preventing jurors from granting undue weight to testimony because it comes from an expert. While under either of these evidentiary admissibility standards, unstructured clinical expert testimony offered on dangerousness in a SVP hearing might be considered suspect (e.g., the error rates associated with its use are too high in a Daubert jurisdiction, or the use of unstructured clinical judgment is no longer generally accepted in the field in a Frye jurisdiction; see ATSA, 2001), until recently, courts have been extremely reluctant to adjudicate this expert testimony inadmissible or even apply their expert evidentiary admissibility standards thoughtfully to this expert testimony (Slobogin, 2006).

Historically, it appears that more accurate expert testimony based on actuarial instruments in SVP hearings are even more likely to face evidentiary admissibility hearings and are also more likely to be successfully challenged based upon these standards than unstructured clinical judgment (Otto & Petrila, 2006; Janus & Prentky, 2003). For example, in People v. Miller (2005), a California appellate court refused to even apply its version of the Frye test to an expert’s offer of unstructured clinical judgment on dangerousness in a SVP trial while the same court held that expert testimony based on an actuarial instrument would have to
pass this admissibility threshold (Faigman & Monahan, 2009). Other scholars have detailed the reasons for why these anomalous practices may have been occurring, such as: (a) judicial inability to accurately evaluate scientific validity, (b) deference to other areas of law which have long allowed questionable clinical expert testimony on dangerousness (e.g., Barefoot v. Estelle, 1983), (c) limited appellate review of evidentiary admissibility decisions, (d) a judicial culture of acceptance of experts and their opinions, and (e) judicial ignorance of relevant empirical research (see Slobogin, 2006, and Sales & Shuman, 2007, for a detailed review of these issues).

Yet, a Texas court decision and several recent American Psychological Association amicus briefs (APA, 2007; APA, 2011) offer a more optimistic view for the future. In a recent death penalty decision (Coble v. Texas, 2010), the Texas court of Criminal Appeals held that one expert’s unstructured clinical expert testimony failed its Daubert standard for admissibility. This case’s appeal to the United States Supreme Court and the APA amicus brief which accompanies it, detailing the research relevant to both the problems with unstructured clinical expert testimony and the superiority of actuarial and structured professional judgment approaches as well as jurors favoritism for the less scientific expert testimony, is likely to lead to greater scrutiny of unstructured clinical expert testimony by judges in a variety of legal contexts. This is especially likely to occur if the Court were to eventually find that in the Texas case, the admission of unstructured clinical prediction expert testimony was a constitutional violation. If this happens, it would be extremely difficult for lower court to accept unstructured clinical predictions of future dangerousness in a SVP hearing as meeting their evidentiary admissibility standard. The APA briefs will also make it more difficult for courts in Frye and Daubert jurisdictions to continue to accept unstructured clinical expert testimony as generally accepted in the SVP context, because the Coble APA brief clearly notes “It is now widely accepted that ‘[un]structured clinical judgment by itself is no longer a useful or necessary approach to appraising violence risk.’ Heilbrun, Yasuhara, & Shah, 2010, p. 5” (APA, 2011, p. 8).

Hopefully, the present research as well as other future research will help convince judges that the effects of admitting poor quality expert testimony, at least in the SVP context, are important and potentially prejudicial, and that the legal system needs to reevaluate their deference toward this form of expert testimony. Without greater scrutiny of the admissibility of unstructured clinical expert testimony on dangerousness occurring, courts face the ironic possibility that more accurate actuarial expert testimony may not reach jurors while less accurate expert testimony does. Regardless of this possibility, given our research supporting juror bias in favor of the less valid expert testimony, even if both forms

16 Faigman and Monahan (2009) argue that Arizona and California courts have decided to apply general rules of relevancy (a much lower threshold than Frye) to unstructured clinical predictions of dangerousness largely because these predictions are not scientific expert testimony and are therefore not subject to the same evidentiary admissibility rules.

17 Subsequent to the acceptance of this article, the Supreme Court denied certiorari in the Coble case.
of expert testimony are admitted, there is still a substantial likelihood of injustice occurring based on jurors’ favoritism for the less accurate testimony.

Beyond the legal system prohibiting expert testimony of poor scientific quality, mental health professionals could be more circumspect in the expertise they offer during their courtroom testimony. By articulating the extreme limitations of unstructured clinical judgment to predict dangerousness as well as the significant problems that still exist for actuarial prediction methods, mental health professionals could provide more reliable and valid expert testimony. Future research should also strive to improve the ability of clinicians to predict future dangerousness and examine the effects of other types of more scientific expert testimony on future dangerousness (e.g., structured or professional judgment instruments) on juror decisions (see, e.g., Lieberman et al., 2007). Researchers also need to continue to examine the extent, and the psychological means by which expert testimony influences jury verdicts in these important legal decisions.

References


Frye v. United States, 292 F. 1013 (D. C. Cir. 1923).


DANGEROUSLY MISUNDERSTOOD


(Appendix follows)
Dr. Brent, were you appointed by the Superior Court to conduct an evaluation for the purpose of a potential SVP commitment on James Stone?
A: Yes.
Q: Did you have a chance to meet with Mr. Stone for an interview?
A: Yes, I did.
Q: How long did you have occasion to meet with Mr. Stone?
A: Approximately two or three hours.
Q: And did you prepare a report as a result of that assessment?
A: Yes, I did. I do a kind of a standard interview.
Q: Why don’t you tell us about that, then?
A: Well, that covers such things as the referring offense, Mr. Stone’s sexual history, whether he’s had any past paraphilia, committed any sex offenses, prior treatment that he’s received, family life, work history, medical history, adjustment to prison, and additional treatment considerations, other related behaviors, and interview behavior in the time that we met at the Arizona State Hospital.
Q: Dr. Brent, as part of your interview screening you need to be familiar with the provision of the sexually violent persons’ statute, I presume?
A: That’s correct.
Q: What are the criteria you are looking for when you conduct the screening?
A: A person must have a mental disorder. A mental disorder can be a paraphilia or a personality disorder. They must have been convicted of a sexually violent crime.
Q: Did you have any information when you reviewed Mr. Stone’s file about whether or not he had been convicted of any sexually violent crimes?
A: Yes, I did.
Q: What did you find?
A: He had been convicted of at least two.
Q: And can you give us the details about the dates of those convictions and the circumstances which surrounded them?
A: Sexual conduct with a minor. This was in 1988 he was arrested, 2-4-88. He was arrested January 19th, 1983. There was a plea on April 29th, 1988. And it consisted of engaging in inappropriate sexual activity with a couple of neighbors. He would go to a fence that separated the two properties and engage in sexual activities through the fence with these two young children.
Q: Now, you have already indicated that in addition to ascertaining whether or not there are convictions for sexually violent offenses, you are asking whether an individual under the statute suffers from a mental disorder?
A: That’s correct.
Q: What are the definitions of a mental disorder according to the statute?
A: According to the statute it means a paraphilia, personality disorder, or a conduct disorder or any combination.
Q: What is paraphilia?
A: A paraphilia is a disorder that consists of having either pretty intense urges, fantasies, thoughts, and/or behavior towards, in this case, inappropriate objects or nonconsenting people. Could be adults or children. So in other words you have lots of urges and behaviors toward some inappropriate types of sexual objects, including children.
Q: Is there a subcategory of paraphilia for individuals who have intense urges, fantasies, or behaviors specifically toward children?
A: Yes, there is.
Q: And what is that called?
A: Pedophilia.
Q: How did you reach that conclusion?
A: Well, I reached the conclusion based upon a number of statements by Mr. Stone indicating that he had probably engaged in activity with a total of 11 children over 11 years, eight of them being females from the ages of three to eight and three of them being males between the ages of five and 10. Those were all prepubescent children. When you talk about prepubescent children, you are talking about children who are younger than about 13 years. And for pedophilia, they must be prepubescent children. And this kind of activity must have occurred over a six-month period, and this was over an 11-year period so those criteria were easily met in this case.
Q: In addition to your finding that Mr. Stone has two convictions for sexually violent offenses, that he suffers from a specific mental disorder known as pedophilia, were you also asked to render a recommendation about whether Mr. Stone is likely to engage in future acts of sexual violence?
A: Yes, we were.
Q: Now, is his diagnosis important to that?
A: Yes, it is.
Q: Why is it important?
A: Well, it’s important because by definition pedophilia really indicates that an individual has engaged in this kind of activity, has engaged in some kind of activity on a fairly long-term basis, and individuals like that are more likely to engage in that activity again than not, hence the diagnosis.

(Appendix continues)
Q: Now, among the records that you had for your review is his participation in the Sex Offender Treatment Program. The fact that he participated in that program for at least two years, if not beyond, how does that affect your ability or affect your ultimate diagnosis of Mr. Stone?
A: Well, the diagnosis doesn’t change, but his participation is a positive factor.
Q: Did you factor that in while reaching any kind of conclusion as to whether or not he was a risk to commit future acts of sexual violence?
A: Yes, I did.
Q: How did you factor it in?
A: Well, in the record—the record indicated that he completed the sex offense treatment program but with deficiencies and I’m not sure exactly what the deficiencies were. They were not specified, but deficiencies are troublesome. His participation is helpful, so it’s sort of like some good things and some bad things associated with that.

Clinical Expert Testimony Manipulation
Q: Having diagnosed Mr. Stone as a pedophile and being familiar with his history, you went ahead and used your clinical judgment anyway?
A: Yes, I did.
Q: Now, based upon your interview and judgment, did you form a judgment of his likelihood of re-offense?
A: Yes, I did.
Q: And what is that?
A: His background and my experiences suggest that he has a high risk of recidivating.
Q: What about your clinical interview was significant in terms of your conclusion that Mr. Stone is more likely than not to re-offend if released into the community?
A: Well, the thing about the clinical interview that I think in part was significant was that he acknowledged the behaviors in question even prior to his first conviction while he was on probation, the behavior leading to his second conviction, and even the continued problems that he really needed to continue to deal with, treatment-wise.
Q: Do you have an opinion specifically as it relates to Mr. Stone as to the relative degree of risk he poses to commit sexually violent acts if he is released into the community? That is, based on your interview and examination, your conclusion about mental disorders that you described for Mr. Stone and the predicate offense, the ’88 offense for sexual conduct with a minor, would you believe to a reasonable psychological certainty that he has a likelihood to recidivate?

Actuarial Expert Testimony Manipulation
Q: Having diagnosed Mr. Stone as a pedophile and being familiar with his history, you went ahead and used actuarial instruments anyway?
A: Yes, I did.
Q: Now, tell us about the instrument you used.

(Appendix continues)
A: The shorter instrument is called the Rapid Risk Assessment of Sexual Offenders. It’s commonly known as the RRASOR with two R’s in it.
Q: Okay. And what is its primary design?
A: The primary design was to assist individuals who are doing these kind of screening evaluations to do it in a rather expeditious manner by looking at some of the factors that are the factors that we find are highly associated with individuals who re-offend.
Q: What are those factors?
A: The first factor is what’s called prior sex offenses.
Q: What’s the second category?
A: The second category is the age of release.
Q: Okay. What is the third factor?
A: The third factor is the victim’s gender.
Q: Why does that make a difference based on your training?
A: We find that individuals who molest males and females have a higher risk of re-offending than individuals who only molest females.
Q: And the fourth factor?
A: Relationship to the victim.
Q: And why is that important?
A: We find that an individual who will molest somebody they are not related to has a higher risk factor to re-offend than someone who molest only people they are related to.
Q: Any other variables or any other factors on this particular instrument?
A: Just four.
Q: What score did you give Mr. Stone on the RRASOR?
A: I believe that was a 3.
Q: What was the other instrument that you used?
A: The other instrument I used is called the Sexual Offender Risk Appraisal Guide, commonly known as the SO-RAG.
Q: How many variables are contained within that instrument?
A: There are 14 variables.
Q: Once all these were added together, what total score did Mr. Stone receive?
A: He received a score of 8.
Q: Now, based upon the research and how this instrument is utilized, is there a corresponding recidivism estimate for Mr. Stone’s score of 8?
A: Yes, there is.
Q: And what is that?
A: A score of 8, according to the research literature, indicates that individuals with this score have a high risk of recidivating.
Q: Is that estimate an overestimate or underestimate?
A: It still would be an underestimate. We are talking arrests as opposed to actual offenses or actual activity that people don’t get arrested for.
Q: Do you have an opinion specifically as it relates to Mr. Stone as to the relative degree of risk he poses to commit sexually violent acts if he is released into the community? That is, based on your review of records, your conclusion about mental disorders that you described for Mr. Stone and the predicate offense, the '88 offense for sexual conduct with a minor, would you believe to a reasonable psychological certainty that he has a likelihood to recidivate?
A: Yes, I do.
Q: What is your opinion?
A: The law asks us to make a determination as to whether a person is likely or not likely and my opinion is that he is likely to.

Cross-examination
Q: Well, Mr. Stone is a pedophile; right?
A: That’s the diagnosis.
Q: And Mr. Stone, regardless of what type of treatment he got, there isn’t a pill that he can take that’s going to eliminate the pedophilia, is there?
A: Not quite that simply. There are medications—I’m not a physician—but that decrease sexual drive.
Q: Well, that’s a good sign, is it not, that he is aware of it, struggling with it, and trying to deal with it?
A: Again, I think that can be answered in two ways. One, it’s better to deal with a problem than not deal with it, but you cannot ignore the fact that there is a significant problem.
Q: Well, again, though, those are all in fact not negative factors to be viewed? You are not telling this jury that in regards to Mr. Stone, are you?
A: I’m very much suggesting to the jury that persistent deviant sexual interest is one of the strongest predictors of sexual recidivism.
Q: The fact that Mr. Stone sought treatment in prison and is already seeking treatment options in the community, is this not a good sign for his future?
A: Again, it is good that Mr. Stone is seeking treatment, but the fact is that he has this problem and that in itself is a negative.
Q: Okay. Well, the State’s pointed out he is a pedophile and that’s one of the criteria under the statute; right?
A: Pedophilia is one of the criteria under the statute; correct.
Q: And that he has a prior conviction, in this case two, for a sexually violent offense pursuant to the statute; correct?
A: He has two sexual convictions, correct.
Q: So the third factor you are bringing into play, it’s more likely than not that he is going to re-offend? Is that not correct?
A: I don’t know if I’m bringing that into play, but that is my opinion based on my interview with him.

Clinical Cross-Examination Manipulation
Q: Now, you made some comments about some of the things you got out of your interview, is that correct, Doctor?

(Appendix continues)
A: Yes.

Q: Your interview with him—how long was that?
A: I spoke with him for a period of two hours or three hours.

Q: So based on this reasonably short time period, you believe that you can make an accurate assessment of Mr. Stone’s likelihood to reoffend?
A: Well, yes, I believe the time was sufficient and that enough was accomplished and discussed in our session to allow me to make this type of judgment.

Q: So, even though Mr. Stone has expressed remorse for what he did, is seeking outside help, and has the full support of his family, you still believe that he cannot be released into society because he poses too great a danger?
A: Like I said, those are all positive factors. However, he has admitted to still feeling interest toward inappropriate stimuli, in his case, children. The fact that he still has these feelings and temptations leads me to believe that he still poses a danger to the children in the community. These types of urges are very hard to overcome, even if he wants to.

Q: And all this is just based upon your opinion?
A: Yes, it is based upon my opinion as a licensed psychologist.

Q: And as such, you believe that after a two or three hour session, you can make this type of determination about the future of Mr. Stone?
A: Yes, I do.

Actuarial Expert Testimony—Cross-Examination Manipulation

Q: Now, you made some comments about some of the things you got out of the presentence report, is that correct, Doctor?
A: I did review the presentence report, yes.

Q: And I think you related some of the contents of it to the jury, did you not?
A: I believe I did.

Q: And again, all you are doing is in effect relaying the information in the presentence report based on what’s there in written form because you have no personal knowledge of this at all?
A: That is correct.

Q: You didn’t find anything in the records and you did not do one of the HARE psychopathy tests on Mr. Stone, did you?
A: No, I did not.

Q: The actual testing—and we will talk about that in a minute, but other than the RRASOR and the SORAG, you didn’t do any of the other actuarial testing on Mr. Stone, did you?
A: No, I did not.

Q: So when we talk about somebody re-offending six years into the future, again, you are using basically an estimate, aren’t you?
A: Yes.
Q: All right. So it’s safe to say, sir, that somebody who is not a professional would have gotten the same result using the same documents you got—as you did on those SORAG and the RRASOR documents; isn’t that right?
A: If they were trained in it, it’s likely that they could have gotten the same scores, yes.

Q: And all of this is predicting the future. There is no hard data and/or any specific information you can give this jury based on—other than the actuarial tables that you prepared, as to whether or not you know exactly whether or not this individual will re-offend or not, is there?
A: The answer is no, I cannot tell you precisely. No.

Q: Well, is there some question within the industry as to whether actuarial instruments should be used in this manner?
A: Yes.

Q: So there are some that don’t feel that it’s an appropriate way to use them and some that do; is that right?
A: That’s correct.

Q: Yeah, you determined it was more likely than not he could re-offend pursuant to the SVP law; correct?
A: That’s correct.

Q: And the fact that you did not interview him for more than two or three hours in your opinion didn’t really or would not have changed that one way or the other?
A: Well, I based my information entirely upon what was in front of me, so I can’t say what I would have done had I had different information, so I don’t know the answer to that question. But based upon the information that I had, that was my opinion.

Q: Thank you, no further questions.