

# Venire Jurors' Perceptions of Adversarial Allegiance

Nicholas Scurich  
University of California–Irvine

Daniel A. Krauss  
Claremont McKenna College

Lauren Reiser and Robert J. Garcia  
University of California–Irvine

LillyBelle Deer  
Claremont McKenna College

Psychological assessment instruments designed to assist mental health experts in conducting forensic evaluations have proliferated in the past 2 decades. These instruments are intended to increase objectivity and reliability and therefore theoretically reduce bias. However, a burgeoning body of research suggests that when such instruments are applied in adversarial settings, reliability is poor and their results are systematically biased toward the retaining party, so-called “adversarial allegiance.” In this experiment, venire jurors read a synopsis of a sexually violent predator proceeding in which 2 experts presented the results of a structured risk assessment. The experts were either *adversarial experts* (i.e., called by the attorneys in the case) or *court-appointed experts* (i.e., called by the court), and their assessments were in conflict (i.e., the criteria for commitment either were or were not met). The order in which the conflicting expert assessments were presented was counterbalanced. Overall, jurors considered the assessments produced by court-appointed experts to be more objective, credible, and persuasive than the assessments produced by adversarial experts. With respect to the likelihood that jurors would commit the respondent, an interaction was observed such that jurors essentially discounted the assessment of adversarial experts, rendering them indistinguishable, but relied on the assessment provided by court-appointed experts in making their decisions. The results suggest that although adversarial allegiance is troubling to researchers, jurors appear to anticipate the presence of adversarial bias and adjust their evaluation of risk-assessment testimony accordingly.

*Keywords:* adversarial allegiance, structured risk assessment, decision making

Mental health experts commonly testify on a variety of legal issues, including whether criminal defendants are competent to stand trial, whether criminal defendants meet the legal standards for insanity, whether individuals should be involuntarily civilly committed, and whether individuals are likely to engage in future injurious behavior (Weiner & Otto, 2014). Such testimony has long been a source of controversy in American jurisprudence. The title of Margaret Hagen’s (1997) infamous book captures the essence of the issue, *Whores of the Court: The Fraud of Psychiatric Testimony and the Rape of American Justice*. Similar sentiment was espoused by John Henry Wigmore—a leading American treatise writer—who lamented that experts were nothing more than paid advocates (Wigmore, 1923), and similar views appear in appellate court dictum (see Mossman, 1999). Suffice it to say, the view that avarice and chicanery drive mental health expert testimony is not recent or uncommon.

## Violence-Risk-Assessment Instruments

The past two decades have seen a proliferation of psychological assessment instruments designed to assist mental health experts in conducting forensic evaluations (Melton et al., 2007). Although these instruments are diverse and are designed to be used for disparate legal purposes (e.g., competency to stand trial, criminal responsibility), they fundamentally all have the same objective, which is to improve consistency and objectivity of the evaluations (Borum & Grisso, 1995; Heilbrun, Grisso, & Goldstein, 2008). Forensic assessment instruments designed to assess the risk of violence have received the most scholarly attention, and consequently will be the focus of this inquiry. Although there is necessarily a loss of generality when focusing on a particular forensic context, the instruments used in violence-risk assessment are similar in many ways to tests and instruments used in other areas of forensic assessment.

A violence-risk assessment has essentially three components: (a) selecting and measuring empirically related risk factors, (b) combining risk factors, and (c) producing a final estimate of risk (Monahan, 2008). A clinical risk assessment structures none of the components. The clinician uses his or her experience and intuition to select risk factors that he or she deems relevant and combines them into a holistic judgment of risk. At the other end of the spectrum is “actuarial” risk assessment, which uses a statistical algorithm to identify and weight risk factors (determined from an

---

Nicholas Scurich, Department of Psychology and Social Behavior and Department of Criminology, Law & Society, School of Law, University of California–Irvine; Daniel A. Krauss, Department of Psychology, Claremont McKenna College; Lauren Reiser and Robert J. Garcia, Department of Psychology and Social Behavior, University of California–Irvine; LillyBelle Deer, Department of Psychology, Claremont McKenna College.

Correspondence concerning this article should be addressed to Nicholas Scurich, 4213 Social & Behavioral Sciences Gateway, Irvine, CA 92697-7085. E-mail: [nscurich@uci.edu](mailto:nscurich@uci.edu)

empirical study and often a particular sample of participants) and ultimately provide a probabilistic estimate of risk. There is an intermediary approach known as *structured professional judgment*. This approach selects and specifies how to measure a defined set of risk factors, but it does not specify how to combine the risk factors or how to generate a final estimate of risk, which is typically expressed in nominal terms (e.g., low, medium, or high risk). The actuarial and structured professional judgment approaches are collectively referred to as *structured risk assessments* because they structure at least one component of the risk-assessment process (Skeem & Monahan, 2011).

The Hare Psychopathy Checklist–Revised (PCL-R; Hare, 2003) is the most well-known and widely used forensic risk-assessment instrument that applies the structured professional judgment approach (Skeem, Polaschek, Patrick, & Lilienfeld, 2011). Technically, the PCL-R is a structured instrument designed to assess the personality disorder “psychopathy.”<sup>1</sup> However, given the moderate correlation between psychopathy and violence, the instrument is commonly used to assess the risk of violence and sexual recidivism (Jackson & Hess, 2007), and even appears as an individual risk factor in the most common structured risk-assessment instruments (e.g., the HCR-20, the Violence Risk Appraisal Guide [VRAG], etc. Monahan, 2005). It is also required by statute to be used for certain evaluations in some jurisdictions (e.g., Texas; DeMatteo et al., 2014). The PCL-R provides a list of 20 risk factors, which an expert rates on a 0–2 point scale, with a 0 indicating that the risk factor *is definitely not present*, a 1 indicating that the risk factor *might be present*, and a 2 indicating that the risk factor *definitely is present*. Thus, scores can span a range of 0–40. The PCL-R manual specifies a threshold of 30, above which the individual is deemed to be a psychopath. Dichotomizing the PCL-R score in forensic settings is controversial in part because the exact threshold appears to be a value judgment and because of the stigmatizing effect such a label can have on legal decision makers (see, e.g., Edens & Petrila, 2006).

In general, structured approaches to risk assessment are superior to unstructured clinical judgment in terms of predictive validity (Mossman, 1994; Grove et al., 2000). The theoretical reasons for this result are comprehensively discussed by Grove and Meehl (1996; see also Meehl, 1954). One reason for the superiority in predictive validity is the increase in reliability afforded by structured risk assessments. Indeed, research indicates that the interrater reliability of the PCL-R is quite high. For instance, Hare (2003) reported that the intraclass correlation coefficient (ICC) for the PCL-R is .87, and Gacono and Hutton (1994) found that 92% of pairs of PCL-R scores differed by less than 2 points. Thus, it seems that structured risk assessments do increase objectivity by reducing variability and subjectivity in the assessment itself.

It is often assumed that the high levels of reliability reported in the research literature would generalize when the instrument is applied in the real world. In part, this assumption seems tenable because structured instruments should leave little room for disagreement about the constituent risk factors. However, vast differences between the research context and actual forensic applications abound. In the research context, instruments such as the PCL-R (Hare, 2003) are often administered by well-trained graduate students and the results have no potential legal implications. It is also not uncommon for researchers to deviate from the official PCL-R protocol, for instance by administering the instrument

without conducting an in-person interview or by using a different threshold (e.g., a score of 20) to make a binary classification. Further complicating the issue is the fact that forensic evaluators are retained by an interested party (e.g., a prosecutor or a defense attorney), and the results of the assessment can have profound implications, including the potential for lifetime commitment and even the imposition of capital punishment. A recent and burgeoning body of research has cast doubt on the assumption that the reliability observed in the research context generalizes to real-world applications.

### Adversarial Allegiance

Murrie, Boccaccini, Johnson, and Janke (2008) examined the reliability of the PCL-R (Hare, 2003) when it was applied in sexually violent predator (SVP) commitment proceedings in Texas. This is an optimal context in which to examine the performance of a structured instrument because (a) the legal issues to be resolved are typically controversial, (b) serious consequences may result from the decision (i.e., indeterminate incarceration), (c) the PCL-R is required to be administered by statute, and (d) there is typically an assessment conducted by both the petitioner (prosecution) and the respondent (defense), thus allowing a direct comparison of scores between experts retained by different parties. Of the 43 trials that occurred between 2000 and 2007 in Texas, 23 had scores from experts retained by both the petitioner and the respondent.

Laboratory research indicates that the standard error of measurement (SEM) for the PCL-R is  $\pm 3$  points (Hare, 2003). Therefore, one would expect a difference of 3 points or less in PCL-R scores 68% of the time (i.e., 1 SEM), and one would expect a difference of 6 points or less 95% of the time (i.e., 2 SEMs). In Murrie et al.’s (2008) study, the average score for the prosecution’s expert was 25.86, whereas the average score for the defense’s expert was 18.04, a difference of nearly 8 points which is greater than 2 SEMs. This difference was observed in over 60% of the cases. A follow-up study with more comprehensive data files observed that 37% of the cases had differences in scores greater than 2 SEMs (Murrie et al., 2009). Furthermore, the average ICC for these studies was .39 and .42, both of which are substantially lower than the impressive .87 reported in research studies (Hare, 2003). Additional field research has suggested that these discrepancies in scoring the PCL-R are not limited to Texas, with researchers finding similar differences in scoring in Florida (Miller, Kimonis, Otto, Kline, & Wasserman, 2012), Canada (Lloyd, Clark, & Forth, 2010), and across a sample of cases in eight states in the United States (DeMatteo et al., 2014).

Although the data from field studies are high in ecological validity, they cannot provide a causal explanation of the observed discrepancy in the scores. It is possible that attorneys simply proffer experts who give the most extreme scores because such scores are favorable to their positions; experts with unfavorable or

<sup>1</sup> The PCL-R also differs from other structured professional judgment (SPJ) instruments in that it specifies how to combine the risk factors to reach a judgment about the presence of psychopathy. Whereas other SPJ instruments explicitly eschew summing the item scores (i.e., “actuarially scoring” the instrument), the PCL-R does not; it might therefore be more appropriately categorized as an actuarial instrument. However, this distinction is unimportant for our purposes.

less favorable scores are not called to testify. Thus, the observed differences between the prosecution and defense experts' scores could be attributable to a selection effect. On the other hand, it is possible that the evaluator's assessment was subtly, even if unconsciously, influenced in a way that is biased toward the side that retained them (Brodsky, 2013). This potential explanation is consistent with so-called *adversarial allegiance*. Random assignment is necessary to disentangle selection effects from allegiance effects: If evaluators were randomly assigned to prosecutor/defense, and nonchance disparities in scores were observed, then one might be fairly confident that disparities in scores are attributable to an allegiance effect, as opposed to a simple selection effect.

To address the causality issue, Murrie et al. (2013) conducted an experiment in which 100 experienced forensic psychologists and psychiatrists were retained by an attorney to conduct evaluations on four offenders. The experimental manipulation was whether the retaining attorney—who was a confederate—was ostensibly a prosecutor or a defense attorney. Participants thus believed that they were conducting an assessment for either a prosecutor or defense attorney. In three of the four cases, an allegiance effect was observed with prosecution-retained experts giving, on average, a score 3 points higher than defense-retained experts (the effect sizes ranged from .55 to .85, which is consistent with the effect sizes observed in the field studies (.63–.83; Murrie et al., 2009). Of course in court, scores are not averaged across experts; rather, a single or pair of experts typically testifies in a given case. Thus, Murrie et al. (2013) conducted an additional analysis in which they estimated the likelihood that a random pairing of scores would differ by more than 6 points (i.e., 2 *SEM*). Results indicated that in 13–33% of the cases, there was a 6-point discrepancy favoring the prosecution, which is a far greater discrepancy than would be predicted by random error alone (Murrie et al.'s Table 2, p. 6). Experts' attitudes toward sex offenders and clinical experience were not related to allegiance effects.

### Jurors and Expert-Witness Testimony

The impact of expert witnesses on legal decisions has been the subject of extensive empirical investigation. It has been studied in both naturalistic (e.g., Shuman, Whitaker, & Champagne, 1994; Sundby, 1997) and experimental (e.g., Brekke & Borgida, 1988) paradigms across a variety of psycholegal issues, including future dangerousness (e.g., Krauss & Sales, 2001), insanity (e.g., Rogers, Bagby, & Chow, 1992), repressed memories of childhood sexual assault (e.g., Buck & Warren, 2010), and eyewitness testimony (e.g., Cutler, Dexter, & Penrod, 1989). Across these legal contexts, testimony by expert witnesses often has a significant effect on the decisions rendered by jurors.

Nietzel, McCarthy, and Kern (1999) conducted a meta-analysis of all the experimental studies involving an expert witness and found that jurors typically show agreement with expert testimony on the whole, but their agreement fluctuates and depends on specific traits of the experts and features of their testimony. For instance, Ivković and Hans (2003) subsequently found that jurors rely on credentials, clarity of expert presentations, and possible motives for testimony in determining the value of an expert's comments. Other research on characteristics of a lone expert's testimony has demonstrated that expert gender is inconsequential (e.g., Couch & Sigler, 2010), that academic credentials have a

persuasive effect only when paired with complex testimony (e.g., Cooper, Bennett, & Sukel, 1996), and that jurors generally fail to detect and respond to methodological flaws in an expert's research (McAuliff & Kovera, 2008; McAuliff & Duckworth, 2010; McAuliff, Kovera, & Nunez, 2009).

One limitation of these studies, however, is that they do not address the fact that many trials involve expert witnesses providing contradictory testimony from competing legal sides. Fewer expert witness studies have examined these effects. For instance, Levett and Kovera (2008) manipulated expert testimony in the context of a child sexual abuse case to see if adversarial experts were an effective mechanism for illuminating methodological flaws in expert-witness testimony about the suggestibility of children. The researchers manipulated both the methodological validity of a defense expert's testimony and how an opposing prosecution expert responded to the defense expert's testimony. Instead of enlightening participants of potential methodological flaws in evidence cited in the defense witness's testimony, testimony from contradictory experts increased participant skepticism of both experts. In other words, the presence of two experts simply leads participants to discount the quality and credibility of experts from both sides. This canceling out, or "skepticism effect," has been observed in other domains, such as expert testimony on eyewitness identifications (e.g., Pezdek, Avila-Mora, & Sperry, 2010).

There is a small body of work that evaluates how jurors respond to opposing expert testimony about future dangerousness. In a simulated capital sentencing proceeding, Krauss and Sales (2001) tested whether having an opposing expert would influence juror verdicts. Results indicated that contradictory expert testimony, with greater information and analysis, did not affect decisions regarding the imposition of capital punishment, suggesting that the two expert conditions simply cancelled each other out, regardless of the quality of the testimony presented. In the context of an SVP proceeding, Boccaccini, Turner, Murrie, Henderson, and Chavalier (2013) found, similarly, that jurors expressed greater skepticism of risk assessment when a defense expert presented testimony than when no opposing expert testimony was presented. In other words, although defense experts were deemed less credible than prosecution experts, the presence of a defense expert reduced the credibility of the prosecution expert as well.

### The Present Study

Observing allegiance effects with the use of structured risk-assessment instruments is troubling. It suggests that, even with the use of tools designed to minimize variability, experts' opinions—in applied settings—are swayed by the side that retained them. Indeed, such effects were even observed in an experiment in which practicing clinicians had minimal exposure to the retaining party, based their assessment strictly on a case file, and had no current financial incentive to be biased (Murrie et al., 2013). If anything, these dynamics surely understate the true effect size of adversarial allegiance. Although the existence of adversarial allegiance might be troubling to researchers and scientists, whether adversarial allegiance has any impact on legal decisions is an empirical question. Jurors might already anticipate that such differences are attributable to adversarial allegiance, even with the use of ostensibly objective instruments, and they might adjust their perceptions of the testimony accordingly. In fact, the adversarial

system is predicated on the assumption that jurors are most able to discern truth by hearing competing and conflicting accounts of evidence. Thus, adversarial allegiance might not be problematic when considered within the broader context in which such testimony is evaluated by triers of fact.

For the present experiment, we systematically examined whether jurors anticipate and discount risk-assessment testimony that might be influenced by adversarial allegiance. We were able to accomplish this by holding the substance of the testimony itself constant, but experimentally manipulating who called the expert to testify. We hypothesized that jurors would favor risk-assessment testimony from a court-appointed expert over that of an expert called by the attorney representing one side in the dispute, even though the testimony and risk-assessment conditions would be identical. If differences were observed, it would suggest that jurors presume that adversarial experts are biased toward the side that retained them.

## Method

### Participants

Participants were drawn from a pool of venire jurors who reported for jury duty in Orange County, California during July, 2014. Before releasing those potential jurors who were not called for duty, a court employee announced that we were conducting a study that would take approximately 15 min to complete and that participants would be compensated \$15.00 for their time. Data were collected on six different days.

We collected data from 217 participants for the experiment. Embedded within the materials were two different comprehension questions probing basic facts related to the experimental manipulations. Of the 217 participants, 42 failed these questions and were removed from the analyses reported (see Oppenheimer, Meyvis, & Davidenko, 2009). The resulting sample consisted of 175 jury-eligible adults, of which 48.6% were women and 51.4% were men, ranging from 19 to 69 years of age. More than half of the participants were 36 years old or younger. In terms of racial/ethnic composition, the sample was 45.1% White, 1.1% Black, 10.9% East Asian, 6.3% South Asian, 4.0% Pacific Islander, 22.9% Latino, 8.6% multiracial, 0.6% Middle Eastern, and 0.6% Native American. The participants' educational attainments were distributed as follows: 5.6% had a high school education or less, 36.6% had some college or a technical degree, 34.3% had a bachelor's degree, and 23.4% had a graduate or professional degree. In terms of prior jury experience, 84.0% reported no prior jury experience, 12.0% reported having served on a jury once, and 4.0% reported serving more than once. The median annual income was \$50,000 among those who chose to disclose this information. Political orientation was distributed as follows: 33.1% Independent, 31.4% Democrat, 28.6% Republican, 5.7% Libertarian, 1.1% Green Party. Religious affiliation varied as follows: 38.1% Christian, 23.9% Catholic, 11.9% agnostic, 9.1% atheist, 5.1% Buddhist, 1.1% Muslim, 6% Hindu, and 9.1% other.

### Procedure and Design

Participants were presented with the jury instructions that are used in SVP proceedings in California. These explain the role of a

juror, the elements of the SVP statute, and the standard of proof (beyond a reasonable doubt). Participants then read a synopsis (~1,300 words) of an actual SVP commitment case that had been used in previous research (Scurich & Krauss, 2013). In brief, the respondent in the case, Robert Hanson, had two previous convictions for sexual assault on a minor after he engaged in sexual intercourse with two different 15-year-old girls. One of the relationships resulted in a pregnancy. At the end of his 5-year prison sentence, the State sought to have Mr. Hanson committed as an SVP. Mr. Hanson stipulated to two of three SVP elements (that he had a previous felony conviction for a specified sex offense and that he had a mental abnormality), but he disputed that he was "likely to commit another criminal act of sexual violence."

To this end, two different forensic psychologists conducted a risk assessment using a fictional instrument, the *Prediction Checklist–Recidivism (PCL-R)*. It was explained that the PCL-R is a widely used instrument with 20 items scored on a 0–2 scale, depending on whether the item is *definitely not present*, *possibly present*, or *definitely present*, respectively. In addition, it was explained that, according to the manual, a score of 25 or higher indicates that the person is likely to commit a criminal act of sexual violence. This hypothetical instrument is virtually identical to the *Psychopathy Checklist–Revised*—the most widely used instrument for assessment purposes—except that it eschews the term "psychopathy" which researchers have suggested has a considerable tainting effect on jurors decisions (DeMatteo et al., 2014; Edens, Colwell, DesForges, & Fernandez, 2005).

The two forensic psychologists, Dr. Campbell and Dr. Knox, were both described as board-certified forensic psychologists with decades of experience conducting risk assessments of sexual offenders. Dr. Campbell testified that Mr. Hanson received a score of 20 on the PCL-R and accordingly believed that he *did not* meet the criteria for commitment. Dr. Knox testified that Mr. Hanson received a score of 30 on the PCL-R and accordingly believed that he *did* meet the criteria for commitment. Both experts added that there was no reason to believe the results or their conclusions were invalid. Note that the substance of each risk assessment was identical; the only difference was in the outcome of the assessment and the expert's opinion with respect to whether or not Mr. Hanson met the criteria for commitment.

The experiment employed a mixed factorial design, in which *the affiliation of the expert* (either appointed by the court or an adversarial attorney) was a between-participants factor and *the result of the assessment* (either meets the criteria or does not meet the criteria for commitment) was a within-participants repeated-measures factor. That is, participants heard the testimony from both Dr. Campbell and Dr. Knox, and participants were told that Dr. Campbell and Dr. Knox were retained either by the court or an adversarial attorney, but not both. For the adversarial attorney condition, as would be the case in real life, Dr. Knox always testified for the state (because he found that Mr. Hanson did meet the criteria for commitment) whereas Dr. Campbell always testified for the respondent (because he found that Mr. Hanson did not meet the criteria for commitment). In both the court and adversarial conditions, the order in which the experts testified was counterbalanced, in that half of the participants first heard the testimony of Dr. Knox followed by Dr. Campbell, and the other half experienced the opposite order.

After receiving the testimony of the first expert, participants were asked how likely they were to commit the respondent and three questions probing the credibility, objectivity, and persuasiveness of the risk assessment, all of which were rated on an 11-point Likert scale. The first likelihood of commitment rating served as a manipulation check to ensure that the differences in the experts' scores were sufficiently large to produce variation in the ratings before they heard the second expert. Participants then heard the testimony of the second expert, and were subsequently told that they had heard all of the evidence in the case. They again completed the same four questions (likelihood, credibility, objectivity, persuasiveness), as well as answered several demographic questions.

**Results**

**Time 1**

After hearing the testimony of a single expert, participants rated how likely they were to commit the respondent on an 11-point Likert scale. A two-way ANOVA with affiliation and result of assessment detected a significant main effect for result of assessment,  $F(1, 173) = 15.84, p < .001, d = .61$ . Participants who heard the testimony of Dr. Knox (who presented the PCL-R score of 30) were more likely to commit the respondent ( $M = 8.63, 95\% \text{ CI } [8.16, 9.09]$ ) than participants who heard the testimony of Dr. Campbell (who presented the PCL-R score of 20;  $M = 7.35, 95\% \text{ CI } [6.93, 7.78]$ ), suggesting that participants were responsive to the risk-assessment testimony. The main effect for source was not significant,  $F(1, 173) < 1$ , nor was the interaction,  $F(1, 173) < 1$ .

The three items probing the objectivity, credibility, and persuasiveness of the risk assessment were entered into a principal component analysis using varimax rotation and yielded a single-factor solution (eigenvalue = 2.07; all others less than 1). Accordingly, the three items were combined into a composite measure ( $\alpha = .774$ ), which is referred to as "credibility." For a comparison of credibility scores and likelihood of commitment ratings by condition at Time 1, see Table 1. Note that the credibility scores and likelihood of commitment ratings are modestly correlated ( $r = .189, p = .014$ ).

A two-way ANOVA with expert affiliation and result of assessment as the independent variables and credibility as the dependent variable detected a main effect for affiliation,  $F(1, 167) = 9.78, p = .002, d = .49$ , and a significant main effect for result of assessment,  $F(1, 167) = 41.43, p < .001, d = 1.00$ . The interaction was not significant,  $F(1, 167) < 1$ . The court-appointed expert was deemed more credible ( $M = 7.66, 95\% \text{ CI } [7.28, 8.04]$ ) than the noncourt expert ( $M = 6.82, 95\% \text{ CI } [6.46, 7.19]$ ), and the

expert, Dr. Knox, who testified that the respondent met the criteria for commitment was deemed more credible ( $M = 8.10, 95\% \text{ CI } [7.7, 8.49]$ ) than the expert, Dr. Campbell, who testified that the respondent did not meet the commitment criteria ( $M = 6.38, 95\% \text{ CI } [6.03, 6.73]$ ). These latter findings are consistent with previous research, which found that people are inclined to commit previously convicted sex offenders and tend to favor evidence that facilitates that outcome (Scurich & Krauss, 2013, 2014; Varela et al., 2014).

**Time 2**

After participants read the testimony of the second expert, they again indicated the likelihood that they would commit the respondent. A mixed-model ANOVA with affiliation as a between-participants factor and result of assessment as a repeated-measures factor detected a significant main effect for result of assessment,  $F(1, 171) = 60.71, p < .001, d = 1.19$ , and a significant interaction,  $F(1, 171) = 6.77, p < .05, d = .31$ . The main effect for affiliation was not significant,  $F(1, 171) < 1$ . These findings are graphically displayed in Figure 1.

As is apparent in Figure 1, participants were more likely to commit the respondent when the second expert resolved that the respondent met the commitment criteria. However, the interaction indicated that this pattern was true only for the court-appointed experts. For the adversarial experts, the risk-assessment scores did not affect the likelihood of commitment,  $t(87) = .81, p = .42, d = -0.17$ . In other words, participants were equally likely to commit the respondent after hearing the conflicting testimony of two adversarial experts. On the other hand, participants were significantly more likely to commit the respondent when the court-appointed expert indicated that the respondent met the criteria for commitment,  $t(83) = 3.62, p < .001, d = -0.78$ .

The three items probing the objectivity, credibility, and persuasiveness of the assessment were entered into a principal component analysis using varimax rotation and yielded a single-factor solution (eigenvalue = 1.96; all others less than 1). Accordingly, they were combined into a composite score ( $\alpha = .715$ ) referred to as "credibility." For a comparison of credibility scores and likelihood of commitment ratings by condition at Time 2, see Table 2. Note that the credibility scores and likelihood of commitment ratings are modestly correlated ( $r = .164, p = .032$ ).

A mixed model ANOVA with affiliation as a between-participants factor and result of assessment as a repeated-measures factor detected a significant main effect for affiliation,  $F(1, 171) = 10.07, p < .001, d = .50$ , with court-appointed experts deemed more credible ( $M = 7.51, 95\% \text{ CI } [7.15, 7.86]$ ) than adversarial experts ( $M = 6.72, 95\% \text{ CI } [6.38, 7.06]$ ). The main effect for result

Table 1  
Table of Credibility and Likelihood of Commitment by Experimental Condition at Time 1

Condition	Adversarial experts			Court-appointed experts		
	Credibility	Likelihood of commitment	<i>n</i>	Credibility	Likelihood of commitment	<i>n</i>
Does meet commitment criteria	7.76 (1.40)	8.68 (1.55)	44	8.44 (1.41)	8.58 (1.97)	37
Does not meet commitment criteria	5.88 (1.90)	7.37 (2.40)	46	6.88 (1.97)	7.34 (2.31)	47

Note. Numbers in parentheses indicate 1 SD.

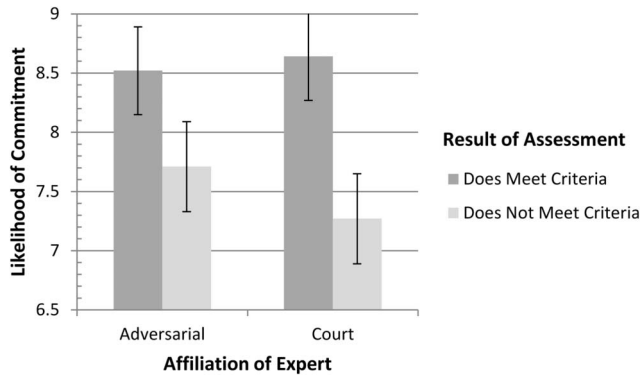


Figure 1. Likelihood of commitment as a function of expert affiliation and result of risk assessment at Time 2. Note that error bars indicate 95% confidence intervals.

of assessment and the interaction were not significant,  $F(1, 165) < 1$ ,  $F(1, 165) < 1$ , respectively. In other words, court-appointed experts—even though their risk assessments were identical to the assessments provided by adversarial experts—were perceived to be more credible than adversarial experts.

### Discussion

This experiment examined what role adversarial allegiance plays in legal decisions. It found that venire jurors are sensitive to the affiliation of experts. In particular, risk assessments conducted by court-appointed experts were overall deemed more objective, credible, and persuasive than adversarial experts. With regard to the influence on the likelihood of commitment, venire jurors seem to discount risk assessments conducted by adversarial experts. Consistent with previous research (e.g., Levett & Kovera, 2008; Krauss & Sales, 2001), when presented with the conflicting risk assessments from adversarial experts, participants seemed to “average out” the two judgments. However, when presented with conflicting assessments conducted by court-appointed experts, participants were responsive to the risk-assessment testimony. Particularly noteworthy is the finding that participants were significantly less likely to commit the respondent when a court-appointed expert opined that the respondent did not meet the criteria for commitment. This finding is remarkable given jurors’ strong propensity to commit SVP respondents (Scurich & Krauss, 2014; Boccaccini et al., 2013). It suggests that one of the few ways SVP proceedings will potentially not end in commitment is if an assessment conducted by a court-appointed expert indicates the respondent does not meet the criteria for commitment. It also implies that the

so-called “skepticism effect” might be limited to adversarial experts only.

These findings raise questions about whether the adversarial system needs to be changed to a more inquisitorial-type system, as some scholars have recently argued (Slobogin, 2014). In the inquisitorial system, the judge acts as a neutral body collecting information from both sides, calling witnesses, evaluating evidence, and rendering a verdict, all in an effort to determine what truly happened in a certain situation. The adversarial system, in contrast, pits prosecution and defense counsel against each other and puts judges in a more removed role. The inquisitorial system is primarily concerned with determining the pure truth (i.e., factual accuracy) of the situation, whereas the adversarial system is more concerned with implementing a just process to reach a fair outcome (King, 2001). The adversarial system has deep roots in American jurisprudence, so is unlikely to be limited in any significant capacity (Sanders, 2007). Moreover, it could be argued that the use of a single court-appointed expert is undesirable because it presupposes that there is a singularly correct answer, when in fact experts can legitimately have differences in opinion on matters.

In light of this conundrum, Australian federal courts have adopted a novel approach to present expert testimony, known as concurrent evidence, or more colorfully, as “hot tubbing experts” (Yarnall, 2009). Under this approach, adversarial experts engage in a dialogue—with each other, the attorneys, and the judge—under oath to probe and attempt to reconcile discrepancies in opinions. Hot tubbing is said to afford several benefits, including a focus of expert testimony on the actual issues in dispute, allowing experts to contemporaneously testify (thus avoiding recall issues), and reducing the time and expense of legal proceedings by up to 50% (Yarnall, 2009), all while allowing each party to call and present its own expert witness. Still, some question whether this approach could feasibly be implemented in a jury trial, particularly criminal trials (see Reifert, 2011).

The reported results cannot answer whether or not the adversarial approach is necessarily preferable to the inquisitorial approach. Such a comparison would assume that there are concrete objectives against which to measure performance of both approaches. Although the results suggest that jurors consider expert testimony from a more quasi-inquisitorial approach (i.e., court-appointed experts) to be more credible, it cannot be said that this is necessarily preferable, because the adversarial system is concerned with a panoply of issues besides ground truth (see Tribe, 1971). The current study is also not a direct test of the effect of hot tubbing experts, since the experts did not interact with one another. However, the findings do suggest that hot tubbing might be unnecessary, given that jurors in our study discounted the testimony

Table 2  
Table of Credibility and Likelihood of Commitment by Experimental Condition at Time 2

Condition	Adversarial experts			Court-appointed experts		
	Credibility	Likelihood of commitment	<i>n</i>	Credibility	Likelihood of commitment	<i>n</i>
Does meet commitment criteria	6.86 (1.62)	8.07 (1.92)	43	7.31 (1.44)	7.19 (2.11)	37
Does not meet commitment criteria	6.48 (2.01)	8.37 (1.57)	46	7.55 (1.81)	8.73 (1.81)	48

Note. Numbers in parentheses indicate 1 SD.

of adversarial experts. Still, this possibility should be examined with future research. Hot tubbing might still be desirable, given the potential increases in efficiency it affords.

As with all laboratory-based juror decision-making experiments, elements of verisimilitude were eliminated to produce greater internal validity. The brevity of materials and corresponding lack of complexity of the trial process (e.g., lack of deliberations, opening and closing statements, etc.) limits both the generalizability and the immediate practical import of the findings. Yet, as this is a context with a lack of well-developed psychological theories and research, and a few already existing corroborative field studies, a focus on internal validity and causal explanations was more heavily weighted in the design. Furthermore, the use of a venire sample of jurors adds an important element of external validity that significantly increases the generalizability of results beyond the laboratory.

### Conclusion

In the end, these findings indicate that concerns about the adversarial allegiance effect with regard to jurors' decisions in SVP cases may be overblown, and that jurors already account for this effect in weighing adversarial experts' risk-assessment testimony. What is less obvious is how jurors process, weigh, and understand conflicting risk-assessment testimony from court-appointed experts, and whether the use of court-appointed experts enhances the overall satisfaction and perceived legitimacy of the process. Although adversarial allegiance is potentially troubling to scientists and researchers, it should be kept in mind that it is accounted for both by the adversarial system and, apparently to some extent, by jurors.

### References

Boccaccini, M. T., Murrie, D. C., & Turner, D. B. (2014). Jurors' views on the value and objectivity of mental health experts testifying in sexually violent predator trials. *Behavioral Sciences & the Law*, *32*, 483–495. <http://dx.doi.org/10.1002/bsl.2129>

Boccaccini, M. T., Turner, D. B., Murrie, D. C., Henderson, C. E., & Chevalier, C. (2013). Do scores from risk measures matter to jurors? *Psychology, Public Policy, and Law*, *19*, 259–269. <http://dx.doi.org/10.1037/a0031354>

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice*, *26*, 465–473. <http://dx.doi.org/10.1037/0735-7028.26.5.465>

Brekke, N., & Borgida, E. (1988). Expert psychological testimony in rape trials: A social-cognitive analysis. *Journal of Personality and Social Psychology*, *55*, 372–386. <http://dx.doi.org/10.1037/0022-3514.55.3.372>

Brodsky, S. L. (2013). *Testifying in court: Guidelines and maxims for the expert witness* (2nd ed.). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/14037-000>

Buck, J. A., & Warren, A. R. (2010). Expert testimony in recovered memory trials: Effects on mock jurors' opinions, deliberations and verdicts. *Applied Cognitive Psychology*, *24*, 495–512.

Cooper, J., Bennett, E., & Sukel, H. (1996). Complex scientific testimony: How do jurors make decisions? *Law and Human Behavior*, *20*, 379–395. <http://dx.doi.org/10.1007/BF01498976>

Couch, J. V., & Sigler, J. N. (2010). Gender of an expert witness and the jury verdict. *The Psychological Record*, *52*, 281–287.

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences & the Law*, *7*, 215–225. <http://dx.doi.org/10.1002/bsl.2370070206>

DeMatteo, D., Edens, J. F., Galloway, M., Cox, J., Smith, S. T., & Formon, D. (2014). The role and reliability of the Psychopathy Checklist–Revised in U.S. sexually violent predator evaluations: A case law survey. *Law and Human Behavior*, *38*, 248–255. <http://dx.doi.org/10.1037/lhb0000059>

Edens, J. F., Colwell, L. H., Desforges, D. M., & Fernandez, K. (2005). The impact of mental health evidence on support for capital punishment: Are defendants labeled psychopathic considered more deserving of death? *Behavioral Sciences & the Law*, *23*, 603–625. <http://dx.doi.org/10.1002/bsl.660>

Edens, J. F., & Petrila, J. (2006). Legal and ethical issues in the assessment and treatment of psychopathy. In C. J. Patrick (Ed.), *Handbook of psychopathy* (573–588). New York, NY: Guilford Press.

Gacono, C. B., & Hutton, H. E. (1994). Suggestions for the clinical and forensic use of the Hare Psychopathy Checklist–Revised (PCL-R). *International Journal of Law and Psychiatry*, *17*, 303–317. [http://dx.doi.org/10.1016/0160-2527\(94\)90032-9](http://dx.doi.org/10.1016/0160-2527(94)90032-9)

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323. <http://dx.doi.org/10.1037/1076-8971.2.2.293>

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19–30. <http://dx.doi.org/10.1037/1040-3590.12.1.19>

Hagen, M. A. (1997). *Whores of the court: The fraud of psychiatric testimony and the rape of American justice*. New York, NY: Harper Collins.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised manual* (2nd ed.). North Tonawanda, NY: Multi-Health Systems.

Heilbrun, K., Grisso, T., & Goldstein, A. M. (2008). *Foundations of forensic mental health assessment*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/med/9780195323092.001.0001>

Ivković, S. K., & Hans, V. P. (2003). Jurors' evaluations of expert testimony: Judging the messenger and the message. *Law & Social Inquiry*, *28*, 441–482. <http://dx.doi.org/10.1111/j.1747-4469.2003.tb00198.x>

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse*, *19*, 425–448.

King, M. T. (2001). Security, scale, form, and function: The search for truth and the exclusion of evidence in adversarial and inquisitorial justice systems. *International Legal Perspective*, *12*, 185–236.

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychology, Public Policy, and Law*, *7*, 267–310. <http://dx.doi.org/10.1037/1076-8971.7.2.267>

Levett, L. M., & Kovera, M. B. (2008). The effectiveness of opposing expert witnesses for educating jurors about unreliable expert evidence. *Law and Human Behavior*, *32*, 363–374. <http://dx.doi.org/10.1007/s10979-007-9113-9>

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

McAuliff, B. D., & Duckworth, T. D. (2010). I spy with my little eye: Jurors' detection of internal validity threats in expert evidence. *Law and Human Behavior*, *34*, 489–500.

McAuliff, B. D., & Kovera, M. B. (2008). Juror need for cognition and sensitivity to methodological flaws in expert evidence. *Journal of Applied Social Psychology*, *38*, 385–408. <http://dx.doi.org/10.1111/j.1559-1816.2007.00310.x>

McAuliff, B. D., Kovera, M. B., & Nunez, G. (2009). Can jurors recognize missing control groups, confounds, and experimenter bias in psycholog-

- ical science? *Law and Human Behavior*, 33, 247–257. <http://dx.doi.org/10.1007/s10979-008-9133-0>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press. <http://dx.doi.org/10.1037/11281-000>
- Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (3rd ed.). New York, NY: Guilford Press.
- Miller, C. S., Kimonis, E. R., Otto, R. K., Kline, S. M., & Wasserman, A. L. (2012). Reliability of risk assessment measures used in sexually violent predator proceedings. *Psychological Assessment*, 24, 944–953. <http://dx.doi.org/10.1037/a0028411>
- Monahan, J. (2005). Clinical and actuarial predictions of violence: Scientific status. In D. Faigman, D. H. Kaye, M. Saks, & J. Sanders (Eds.), *Modern scientific evidence* (pp. 90–109). St. Paul, MN: Thomson West.
- Monahan, J. (2008). Structured risk assessment of violence. In R. Simon & K. Tardiff (Eds.), *Textbook of violence assessment and management* (pp. 17–33). Washington, DC: American Psychiatric Association.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783–792. <http://dx.doi.org/10.1037/0022-006X.62.4.783>
- Mossman, D. (1999). “Hired guns,” “whores,” and “prostitutes:” Case law references to clinicians of ill repute. *Journal of the American Academy of Psychiatry and the Law*, 27, 414–425.
- Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science*, 24, 1889–1897. <http://dx.doi.org/10.1177/0956797613481812>
- Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, 32, 352–362. <http://dx.doi.org/10.1007/s10979-007-9097-5>
- Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater dis(agreement) on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, 15, 19–53. <http://dx.doi.org/10.1037/a0014897>
- Nietzel, M. T., McCarthy, D. M., & Kern, M. J. (1999). Juries. In *Psychology and law* (pp. 23–52). New York, NY: Springer.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867–872. <http://dx.doi.org/10.1016/j.jesp.2009.03.009>
- Pezdek, K., Avila-Mora, E., & Sperry, K. (2010). Does trial presentation medium matter in jury simulation research? Evaluating the effectiveness of eyewitness expert testimony. *Applied Cognitive Psychology*, 24, 673–690. <http://dx.doi.org/10.1002/acp.1578>
- Reifert, E. (2011). Getting into the hot tub: How the United States could benefit from Australia’s concept of “hot tubbing” expert witnesses. *University of Detroit Mercy Law Review*, 89, 103–116.
- Rogers, R., Bagby, R. M., & Chow, M. M. (1992). Psychiatrists and the parameters of expert testimony. *International Journal of Law and Psychiatry*, 15, 387–396. [http://dx.doi.org/10.1016/0160-2527\(92\)90019-W](http://dx.doi.org/10.1016/0160-2527(92)90019-W)
- Sanders, J. (2007). Expert witness ethics. *Fordham Law Review*, 76, 1539–1584.
- Scurich, N., & Krauss, D. A. (2013). The effect of adjusted actuarial risk assessment on mock-jurors’ decisions in a sexual predator commitment proceeding. *Jurimetrics Journal*, 53, 395–413.
- Scurich, N., & Krauss, D. A. (2014). The presumption of dangerousness in sexually violent predator commitment proceedings. *Law, Probability and Risk*, 13, 91–104. <http://dx.doi.org/10.1093/lpr/mgt015>
- Shuman, D., Whitaker, E., & Champagne, A. (1994). An empirical examination of the use of expert witnesses in the courts: Part II. A three part study. *Jurimetrics*, 34, 193–208.
- Skeem, J. L., & Monahan, J. (2011). Current directions in violence-risk assessment. *Current Directions in Psychological Science*, 20, 38–42. <http://dx.doi.org/10.1177/0963721410397271>
- Skeem, J. L., Polaschek, D. L., Patrick, C. J., & Lilienfeld, S. O. (2011). Psychopathic personality bridging the gap between scientific evidence and public policy. *Psychological Science in the Public Interest*, 12, 95–162. <http://dx.doi.org/10.1177/1529100611426706>
- Slobogin, C. (2014). Lessons from inquisitorialism. *Southern California Law Review*, 87, 13–36.
- Sundby, S. (1997). The jury as critic: An empirical look at how capital juries perceive expert and law testimony. *Virginia Law Review*, 83, 1109–1188. <http://dx.doi.org/10.2307/1073729>
- Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84, 1329–1393. <http://dx.doi.org/10.2307/1339610>
- Varela, J. G., Boccaccini, M. T., Cuervo, V. A., Murrie, D. C., & Clark, J. W. (2014). Same score, different message: Perceptions of offender risk depend on Static-99R risk communication format. *Law and Human Behavior*, 38, 418–427. <http://dx.doi.org/10.1037/lhb0000073>
- Weiner, I. B., & Otto, R. (Eds.). (2014). *Handbook of forensic psychology* (4th ed.). Hoboken, NJ: Wiley.
- Wigmore, J. (1923). *A treatise on the Anglo-American system of evidence in trials at common law: Including the statutes and judicial decisions of all jurisdictions of the United States and Canada*. Boston, MA: Little, Brown.
- Yarnall, M. A. (2009). Dueling scientific experts: Is Australia’s hot tub method a viable solution for the American judiciary? *Oregon Law Review*, 88, 311–340.

Received January 12, 2015

Revision received February 14, 2015

Accepted February 16, 2015 ■