Psychology, Public Policy, and Law
2001, Vol. 7, No. 2, 267–310

1076-8971/01/$5.00 DOI: 10.1037//1076-8971.7.2.267

THE EFFECTS OF CLINICAL AND SCIENTIFIC EXPERT TESTIMONY ON JUROR DECISION MAKING IN CAPITAL SENTENCING

Daniel A. Krauss
Claremont McKenna College

Bruce D. Sales
University of Arizona

The Supreme Court and many state courts have assumed that jurors are capable of differentiating less accurate clinical opinion expert testimony from expert testimony based on more sound scientific footing and of appropriately weighing these two types of testimony in their decisions. Persuasion and jury decision-making research, however, both suggest that this assumption is dubious. The authors investigated whether mock jurors are more influenced by clinical opinion expert testimony or actuarial expert testimony. Results suggested that jurors are more influenced by clinical opinion expert testimony than by actuarial expert testimony and that this preference for clinical opinion expert testimony remains even after the presentation of adversary procedures. Limited empirical evidence was found for the notion that various types of adversary procedures will have a differential impact on the influence of expert testimony on juror decisions.

The concept of dangerousness and its prediction play a significant role in many aspects of the legal system. Assessments of dangerousness affect not only civil and criminal litigation outcomes—that is, involuntary civil commitment of individuals (*O'Connor v. Donaldson,* 1975), psychotherapists' liability for their patients' actions (*Tarasoff v. Regents of the University of California,* 1976), post-jail detention of sexual predators (*Kansas v. Hendricks,* 1997), sentencing of criminal defendants—but also influence other important arenas of legal decision making (e.g., pretrial detention of criminal defendants (*United States v. Salerno,* 1987), parole of criminal defendants, and deciding whether a criminal defendant lives or dies (*Barefoot v. Estelle,* 1983).[1]

Psychologists and psychiatrists in increasing numbers have offered themselves as experts in estimating future dangerousness or future violence (Golding, 1992; Melton, Petrila, Poythress, & Slobogin, 1997; Monahan & Steadman, 1994; Showalter, 1990; Shuman & Sales, 1998), and the courts—including the United States Supreme Court (see *Estelle v. Smith,* 1981)—have uniformly recognized certain mental health practitioners (psychologists and psychiatrists) as experts in this area. Yet it is not clear that mental health practitioners have scientific warrant

Editor's Note. Norman J. Finkel served as the action editor for this article.—JGD

Daniel A. Krauss, Department of Psychology, Claremont McKenna College; Bruce D. Sales, Department of Psychology, University of Arizona.

Correspondence concerning this article should be addressed to Daniel A. Krauss, Department of Psychology, Claremont McKenna College, 500 East 9th Street, 321 Seaman Hall, Claremont, California 91711. Electronic mail may be sent to daniel_krauss@mckenna.edu.

[1]Future dangerousness of a criminal defendant is an important consideration in eight states' capital sentencing provisions, and in Oregon and Texas a finding of dangerousness is an essential requirement for the imposition of the death penalty (Worrell, 1987).

or special expertise in making these predictions. Empirical research (e.g., Monahan, 1981, 1984; Monahan & Steadman, 1994; Otto, 1992) and naturalistic experiments (see Marquart, Ekland-Olson, & Sorensen, 1989) have demonstrated that mental health clinicians are not particularly accurate in making dangerousness assessments.[2] Furthermore, many legal scholars have posited that jurors place undue weight on a mental health practitioner's unreliable dangerousness predictions because a scientific witness has special credibility. The unreliability and potential bias of mental health clinicians' dangerousness predictions have caused many critics within and outside the fields of psychology and psychiatry to charge that mental health professionals are incapable of offering truly "expert" testimony in this area and that those practitioners who do testify are engaged in unscientific and unethical practice (Faust & Ziskin, 1988; Lavin & Sales, 1998).

In a landmark case exploring the validity of mental health practitioners' predictions of dangerousness, *Barefoot v. Estelle* (1983), the United States Supreme Court refused to overturn the death sentence of a criminal defendant in a Texas trial in which dangerousness testimony was offered. Under Texas's state law, if the defendant is to be executed, the jury has to unanimously find, beyond a reasonable doubt, that "there is a probability that the defendant would commit criminal acts of violence that would constitute a continuing threat to society" (Tex. Crim. Proc. Code, Art. 37.071b).[3] The defendant's 14th Amendment due process challenge was based on the potentially constitutionally inappropriate admission of two clinicians' predictions of his future dangerousness. The clinicians testified that there was "100% chance" that the defendant would continue to be dangerous and that they were absolutely certain their prediction was correct (*Barefoot v. Estelle,* 1983).

An amicus curiae brief submitted by the American Psychiatric Association served as the evidentiary justification for the defendant's constitutional claim. The brief detailed the then-existing empirical research on the accuracy of mental health practitioners' dangerous predictions, indicating that mental health clinicians erred in their predictions two out of three times (Amicus Curiae Brief for the American Psychiatric Association, in *Barefoot v. Estelle,* 1983, citing Monahan,

---

[2]A recent study has demonstrated that clinicians may be fairly accurate in predicting dangerousness when they are confident of their dangerousness assessment and when their prediction concerns an inpatient psychiatric population over a relatively short period of time (McNeil, Sandberg, & Binder, 1998). This research does not, however, support the idea that clinicians are accurate in predicting the long-term dangerousness of a criminal defendant.

[3]At the time of the *Barefoot* decision, for the defendant to receive the death penalty in Texas, two other questions must also have been unanimously answered in the affirmative by a jury. These questions pertained to (a) whether the conduct of the defendant that caused the death of the deceased was caused deliberately and with the reasonable expectation that the death of the deceased or another would result; and (b) if raised by evidence, whether the conduct of the defendant in killing the deceased was unreasonable in response to provocation, if any, by the deceased (Tex. Crim. Pro. Code Art. 3701b, 1985). These questions were not in issue in the *Barefoot* case. Research has demonstrated that the dangerousness question is the pivotal issue in the vast majority of death penalty cases (Marquart et al., 1989; Constanzo & Constanzo, 1994). The above Texas death penalty instructions were later found unconstitutional by the Supreme Court in *Penry v. Lynaugh* (1989) and have since been modified by the Texas legislature to allow for the presentation of mitigation evidence by the defendant. However, the current Texas death penalty instructions still require a showing of dangerousness before a defendant can be executed.

1981, who found in his review of the five then-existing studies on clinical predictions of dangerousness that clinicians making these decisions had error rates of 65–85%). The United States Supreme Court, unconvinced by this evidence, upheld the death sentence of the defendant. Justice White, writing for the majority, stated that "the suggestion that no psychiatrist's testimony may be presented with respect to the defendant's dangerousness is somewhat like asking us to disinvent the wheel" (*Barefoot v. Estelle*, p. 3396) and that *"we are not persuaded* that such testimony is almost entirely unreliable and *that the fact finder and the adversary system will not be competent to uncover, recognize, and take due account of its shortcomings"* (*Barefoot v. Estelle*, p. 3398); (emphasis added). The *Barefoot* decision not only did not end the controversy but apparently has fueled it surrounding mental health practitioners' clinical predictions of dangerousness.

The purpose of this article is to explore how the fact finder and the adversary system deal with and take due account of the shortcomings of expert testimony on dangerousness. To accomplish this goal, we empirically investigated: (a) whether mock jurors are influenced by expert testimony on dangerousness in capital sentencing, (b) whether mock jurors differentially rely on clinical opinion expert testimony in comparison to actuarial based[4] expert testimony in their decision making, (c) whether traditional adversary procedures (cross-examination and a competing expert) can remove bias caused by unreliable expert testimony offered by a clinical expert, and (d) by what decisional processes the different types of expert testimony affect juror decision making. Before addressing the study, we first consider the relevant aspects of the legal literature on the admissibility of expert testimony and scientific evidence, the social and cognitive psychological literature on expertise and its effects on jury decision making, the human judgment and decision-making literature on cognitive biases and clinical predictions of dangerousness, and the recent developments in the risk assessment literature of actuarial dangerousness prediction instruments. These discussions are critical to put the empirical research and its findings in the appropriate intellectual context.

Expert Testimony and the Admissibility of Scientific Evidence

Because of the inaccuracy of clinically based predictions of dangerousness and their potential for misuse by the jury, legal and scientific controversy has surrounded the evidentiary admissibility of these predictions both before and subsequent to the *Barefoot* decision (Faust & Ziskin, 1988; Melton et al., 1997; Shuman & Sales, 1998). Prior to *Barefoot,* several legal commentators suggested that clinicians' inaccuracy in predicting dangerousness should prohibit the court from admitting this testimony because it was not scientifically based and it was prejudicial to the defendant (Ennis & Litwack, 1974; Giannelli, 1980; Morse, 1978). More recently, Shuman and Sales (1998) argued that mental health clinicians' dangerousness predictions fail to meet the newly developed scientific evidentiary standards announced in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993). To understand these criticisms, it is first necessary to examine the

[4]Actuarial prediction refers to any estimation method that uses combinations of empirically verified risk factors taken from research on large groups of dangerous individuals.

Federal Rules of Evidence (Fed. R. Evid.), which govern the admissibility of evidence, expert testimony, and scientific evidence in the federal courts. Thirty-eight states, including Texas, have relied on these rules to craft their state evidentiary requirements (Imwinkelreid, 1994).

Under the Federal Rules of Evidence, evidence is admissible if it is relevant and not prejudicial (Fed. R. Evid. 401 & 403). Relevant evidence is any evidence that makes a fact in issue more or less probable, whereas nonprejudicial evidence is evidence that is more probative than it is prejudicial. For example, in most circumstances, a criminal defendant's past crimes have been judged to be prejudicial and are inadmissible in the determination of whether the defendant has committed the act for which he or she is being tried.[5] Courts have assumed that the admission of such evidence will cause the trier of fact to place undue weight on the defendant's past criminal behavior rather than the defendant's actual behavior in deciding guilt in the present case.

Additional evidentiary constraints are placed on the admission of expert testimony because courts have reasoned that jurors lack the requisite knowledge to make an intelligent decision on this evidence and, as result, may place excessive weight on it regardless of its veracity (*Daubert v. Merrell Dow,* 1993; Strong, 1995). Individuals presenting expert testimony and scientific evidence have been granted broader leeway in presenting their testimony. Unlike other witnesses, experts are allowed to offer opinions concerning behavior they have not directly observed and to offer opinions not based on admissible evidence (see Fed. R. Evid. 702 & 703). Expert testimony, however, is only admissible if it is evidence that will assist the jury to understand other evidence or determine an important fact (Fed. R. Evid. 702).

Similarly, expert testimony on scientific evidence has been subject to heightened scrutiny by federal and state courts (Strong, 1995). Prior to 1993, a wide variety of standards were used to adjudicate the admissibility of expert testimony presenting scientific evidence (Strong, 1995). The most commonly used of these tests was the *Frye* test, based on a 1923 Washington, D.C. circuit court case concerning the admission of an interpretation of an early polygraph (Melton et al., 1997). The *Frye* court ruled that expert testimony on scientific evidence is admissible if it is "sufficiently established to have gained general acceptance in the particular field to which it belongs" (*Frye v. United States,* 292 F. 1013, D.C. 1923). By this standard, it is not clear whether clinical predictions of dangerousness would be admissible. The *Frye* test, however, has not served as an effective bar to the admission of clinical predictions of dangerousness by mental health professionals (Melton et al., 1997), and some have argued that this standard has not actually been applied to clinical predictions of dangerousness (Shuman, 1994).[6]

In *Daubert v. Merrell Dow* (1993), the United States Supreme Court interpreted Rule 702 of the Federal Rules of Evidence as imposing additional require-

[5]This evidence, however, may be admissible for evidentiary purposes other than showing a defendant's guilt; see Fed R. Evid. 404(b).

[6]Shuman contended that courts have admitted clinical predictions of dangerousness based on relevancy and prejudice standards and have not subjected this testimony to the more rigorous *Frye* standard. Further elucidation of this argument is provided later in this article.

ments on the admissibility of expert testimony concerning scientific evidence. The *Daubert* court emphasized that evidentiary reliability or scientific validity was the major concern in evaluating scientific evidence and that the judge held the responsibility for evaluating the evidentiary reliability of the evidence before it was presented to the jury. The Court suggested that judges consider four factors in determining the admissibility of expert testimony regarding scientific evidence: (a) whether the theory or hypothesis is falsifiable or testable and has been tested, (b) whether the evidence has been subject to peer review, (c) whether there is a known or potential error rate for the evidence, and (d) whether the technique or method is generally accepted within the field (*Daubert v. Merrell Dow,* 1993).

The exact impact of the *Daubert* decision on the admissibility of expert testimony evidence in state courts is still unknown (Goodman-Delahunty, 1997). Although 38 states have relied on the Federal Rules of Evidence in drafting their law, none are bound by the *Daubert* decision because it involves an interpretation of an evidentiary rule and does not involve a constitutional issue (Goodman-Delahunty, 1997; Melton et al., 1997; Shuman & Sales, 1998). As of 1997, 5 states had expressly rejected *Daubert,* 11 had adopted it, and the rest of the states were still deciding (Goodman-Delahunty, 1997; Melton et al., 1997). Further clouding the situation in state courts is the inexact relationship between social scientific evidence, expert opinion testimony offered by clinicians,[7] and the scientific evidentiary rules of admissibility.

Shuman and Sales (1998) argued that some courts have made a distinction between expert clinical opinion testimony and expert testimony based on scientific evidence. They contended that clinical opinion testimony (i.e., testimony like that which was presented in *Barefoot:* expert testimony based on the clinician's years of experience rather than scientific evidence or research) has not been subject to the more stringent scientific evidence standards. There is persuasive evidence in Florida and California for their claims (see *People v. McDonald,* 1984; *State v. Flanagan,* 1993). In both *People v. McDonald* (Cal. 1984) and *State v. Flanagan* (Fla. 1993), state courts have explicitly held that expert clinical opinion testimony does not have to meet scientific evidence admissibility standards to be admitted in trial (e.g., *Frye* or *Daubert*). These courts have assumed that jurors will weigh expert clinical opinion evidence differently than expert testimony based on scientific evidence in reaching decisions.

Other state courts seemed to have adopted an intermediate position in applying *Daubert* to expert testimony (see *Moore v. Ashland Chemicals, Inc.,* 1998, for a review of different jurisdictions' policies). These courts, including Texas's appeal courts, have applied a modified *Daubert* standard to both clinical opinion testimony and "softer" scientific testimony. These jurisdictions have applied some, but not all, of the *Daubert* criteria to determine admissibility of the proffered evidence or have created new criteria to assess the reliability of the proffered information (*Moore v. Ashland,* 1998). Under the Texas scientific

[7]Expert clinical opinion testimony refers to expert testimony that is offered by a member of the scientific community but is not necessarily based on any empirical research. This expert testimony generally offered by clinicians reflects their clinical judgment regarding an issue with regard to an individual person and may or may not be supported by any current scientific study.

evidence admissibility standard, the appropriate *Daubert* questions for clinical opinion testimony are: (a) whether the field of expertise is a legitimate one, (b) whether the subject matter of the expert's testimony is within the scope of that field, and (c) whether the expert's testimony properly relies on and/or uses the principles involved in the field (*Nenno v. State,* 1998).

Unlike some state courts, the federal courts have not made an evidentiary admissibility distinction between scientific and others forms of expert testimony. The Supreme Court in *Kumho v. Carmichael* (1999) held that the reliability standard announced in *Daubert* applies to all expert testimony regardless of whether it is scientific expert testimony or clinical opinion expert testimony. The Court clarified that the trial judge must find the expert testimony reliable and relevant to admit the expert testimony at trial and that some, all, or none of the factors mentioned in the *Daubert* case may be relevant in making this adjudication. The Court further indicated that the trial court judge has broad discretion in making these evidentiary admissibility decisions and that these lower court decisions are largely unreviewable by appellate courts because the standard for review is abuse of discretion by the trial judge (*General Electric v. Joiner,* 1998).

The broad discretion allowed to federal trial court judges to admit expert testimony still makes the admissibility of clinical opinion expert testimony unclear, even after *Kumho.* A trial court judge could weakly apply the *Daubert* admissibility standard by not applying any of the reliability factors mentioned in the *Daubert* case to the proffered testimony and admit clinical opinion expert testimony simply because other experts in the field commonly give similar testimony. In contrast, a trial court judge could more stringently apply the *Daubert* admissibility standard by applying one or more of the factors mentioned in *Daubert* to the proffered testimony and conclude that clinical opinion expert testimony is too unreliable to be admitted.

Moreover, the admissibility of clinical opinion testimony remains a "hot" topic for the courts because there are circumstances in which the admission of this testimony could lead to unjust outcomes. Inaccurate expert clinical opinion testimony based solely on a clinician's years of experience could be admitted into evidence at trial[8] under the more lenient relevancy or a modified *Daubert* standard, even though the reliability (accuracy) of such testimony is dubious.[9]

In *Nenno v. State* (1998), for example, expert clinical opinion testimony on the future dangerousness of a child molester was admitted at a capital sentencing hearing under a modified *Daubert* admissibility standard. The court reasoned that

[8]This statement assumes that scientific evidentiary rules apply to sentencing. Some jurisdictions, including Texas, apply their evidentiary rules to the sentencing process (Kelly, 1992); however, not all states do. Jurisdictions differ widely in this regard. Generally, sentencing hearings allow a greater breadth of information to be presented than is allowed during the trial. As a consequence, it is possible that some jurisdictions might not impose strict scientific evidence standards (e.g., *Daubert* or *Frye*) on scientific evidence presented during sentencing. Irrespective of current practices, a strong case could be made for the imposition of scientific evidence standards in sentencing when capital punishment is an issue.

[9]A full discussion of the inaccuracies of clinical opinion expert testimony, especially in the area of dangerousness prediction, is explored in the third section of this article, Cognitive Biases in Clinical Predictions of Dangerousness. That section also highlights research suggesting that actuarial predictions of dangerousness are likely to be more accurate than clinical ones.

the clinician's prediction of future dangerousness was reliable, even though there was no evidence presented that the clinician or any clinician had the ability to predict future dangerousness of this defendant or other child molesters.

This admissibility practice by courts could lead to biased outcomes unless, as the *McDonald* and *Flanagan* courts assumed, jurors are capable of (a) distinguishing expert clinical opinion testimony from expert testimony based on scientific evidence when they have the opportunity to hear such evidence, and (b) weighing expert clinical opinion testimony differently than other expert scientific testimony. Under these circumstances, unjust verdicts could only be avoided if adversary procedures (cross-examination and a competing expert) could correct this potential juror bias. Whether jurors can perform (a) and (b) and, if not, whether the adversary process can correct for this bias are unanswered empirical questions and the focus of this research study. Empirical studies on expert testimony and juror decision making suggest, however, that assumptions (a) and (b) are dubious.

Research on Expert Testimony's Influence on Juror Decision Making

Psychological empirical research on expert testimony and juror decision making has been primarily influenced by two theoretical perspectives: heuristic models of persuasion[10] and Hastie, Penrod, and Pennington's (1983) story model of juror decision making. Each has contributed important ideas to the understanding of jurors' perception of and use of expert testimony in reaching verdicts.[11] The heuristic model of persuasion implies that juror decision making may be unduly influenced by the credentials of an expert rather than by the quality of his or her testimony (Brekke & Borgida, 1988; Cooper, Bennett, & Sukel, 1996; Schuller & Vidmar, 1992). The theory posits that persuasiveness of a message is usually a product of the quality of the argument (Golberg, 1968; Petty & Caccioppo, 1986). However, when the message is difficult to comprehend or is complex, individuals rely on cognitive shortcuts or heuristics to evaluate the value of the communication (Chaiken, 1987; Petty & Caccioppo, 1986). This peripheral processing or heuristic shortcut focuses the individual on the perceived expertise and trustworthiness of the communicator rather than the quality of the message. In the courtroom, it is hypothesized that a heuristic method of evaluation is used to interpret expert testimony or scientific evidence because of the testimony's complexity and jurors' preexisting willingness to rely on the opinion of experts (Cooper et al., 1996; Schuller & Vidmar, 1992).

Hastie et al.'s (1983) story model suggests a more global view of expert testimony. Expert testimony influences juror decision making through its relationship with the stories the juror constructs about other evidence and testimony

[10]There are, in actuality, two major distinct heuristic models: Chaiken's heuristic model of persuasion and Petty and Caccioppo's elaboration likelihood model of persuasion. For simplicity, they are treated here as one model because they share similar predictions concerning expert testimony's influence and the mechanism by which this influence occurs. Both models suggest that complicated information is evaluated peripherally through heuristic shortcuts, whereas less complex information is interpreted centrally by focusing on the content of the message.

[11]The models are not mutually exclusive, and expert testimony's influence on juror decision making is probably a product of both models.

presented during the trial. The fit between the expert testimony, the juror's preexisting views concerning the issues, and the juror's final story determines the weight placed on the proffered expert testimony in decision making. As a consequence, for expert testimony to be effective it must be specifically linked to other evidence in the trial (Sundby, 1997).

Both the story model and the heuristic model of juror decision making have garnered empirical support. Empirical research on expert testimony and juror and mock juror decision making has demonstrated that psychological expert testimony strongly affects final outcomes when it is presented[12] on: (a) the fallibility of eyewitness identifications (Cutler, Penrod, & Dexter, 1989; Fox & Walters, 1986; Hosch, Beck, & McIntrye, 1980; Loftus, 1986; Wells, 1986; Wells, Lindsay, & Tousignant, 1980); (b) clinical syndromes (i.e., battered wife syndrome, rape trauma syndrome, child sexual abuse syndrome, and repressed memory syndrome); (Brekke & Borgida, 1988; Brekke, Enko, Clavet, & Seelau, 1991; Ewing, 1987; Gaboras, Spanos, & Joab, 1993; Kovera, Gresham, Borgida, Gray, & Regan, 1997; Schuller & Vidmar, 1992; Walker, 1990; however, see Finkel, Meister, & Lightfoot, 1991, and Follinstad, Polek, Hause, Deaton, Bugler, & Conway, 1989, for evidence of an indirect effect of expert testimony on juror decisions); (c) insanity (Rogers, Bagby, & Chow, 1992; Rogers, Bagby, Crouch, & Cutler, 1990; Greenberg & Wursten, 1988); and (d) future dangerousness of a defendant (Morier, 1987). Although expert testimony can affect juror thinking, the mechanism by which it accomplishes this is poorly understood. Psychological empirical research on expert testimony has, nonetheless, successfully highlighted a number of substantive factors that are important in influencing juror decision making.[13] These factors have bearing on the methodology of this study.

The complexity of the testimony presented by the expert and the expert's credentials have been found to affect mock juror decisions (Cooper et al., 1996; Greenberg & Wursten, 1988). As predicted by the heuristic model, the complexity of the expert testimony given is directly related to the juror's reliance on an expert's credentials in reaching a decision, and jurors' decisions are not related to the content of the expert's testimony (Cooper et al., 1996). This effect may occur even when the expert testimony is not particularly complex (i.e., where supposedly simple noncomplex testimony is difficult to understand). Greenberg and Wursten (1988), in a simulated insanity trial, found that jurors were influenced more by medical expert testimony than by psychological expert testimony even when the testimony presented by each expert was identical in form and complexity. Consequently, the level of complexity of testimony and the credentials of the expert should be controlled in experiments not directly examining these issues and were controlled in the present experiment.

The type of information presented by the expert may also influence the persuasiveness of the testimony. Case study information (i.e., information that originates from one case or one event) has been demonstrated in a number of

[12]However, see Sundby (1997) for a more skeptical view of the influence of expert testimony on juror decisions in capital sentencing.

[13]The social psychology literature has also documented a number of physical factors (i.e., attractiveness, age, demeanor, etc.) that affect the persuasiveness of an expert, but they are not discussed here. Attempts were made to control these factors in this experiment.

persuasion experiments to be more influential in decision making than more statistically based information (i.e., information that is presented as a statistical summary of a large number of events; O'Keefe, 1990). The cause of this bias is unclear. Individuals may have a special affinity for case information, a distrust of statistics, or bias against complex information, but this effect occurs even though case-specific information is often more inaccurate than statistical information. This experimental finding is particularly important in the present study because this experiment compares more case-specific testimony (i.e., clinical opinion testimony) to more actuarial or statistical testimony (i.e., actuarial testimony). In this experiment, the statistical nature and complexity of the actuarial testimony were minimized so that this bias would be lessened, but this effect may still influence juror decisions.

Psychological research investigating juror skepticism and juror sensitivity contributes further information concerning the relationship between jurors' use of heuristic processes and the influence of expert testimony on juror decisions. Generally, these studies examine whether jurors attempt to understand the content of expert testimony, and the testimony's link to facts of the case (juror sensitivity), or simply favor the side that has the expert or the more credentialed expert regardless of the content of the expert's testimony (juror skepticism).[14] Studies examining the effect of expert testimony on the fallibility of eyewitness identification in mock juror decision making have demonstrated that mock jurors do not simply base final decisions on the credentials and side of the expert (i.e., juror skepticism). These studies have discerned that, on a variety of variables, mock jurors exhibit sensitivity to expert testimony by basing their verdicts on the fit between the expert testimony and the facts of the case (Cutler et al., 1989). For example, mock jurors have shown that an expert presenting testimony suggesting that lighting may affect the accuracy of an eyewitness identification is only relied on in juror decisions when lighting problems are present in the case before them. This effect is contrary to predictions based on a heuristic model of processing and favors the story model of juror decision making because jurors are focusing on the fit between the expert testimony and the evidence presented at trial. The heuristic model would predict that this scientific testimony would focus jurors on the credentials of the expert rather than the content of the testimony.[15]

Juror sensitivity, however, may not exist for all aspects of expert testimony. The spurious relationship between witness confidence and accuracy has not always produced a sensitivity effect, even when expert testimony is presented explicitly on this issue (Cutler et al., 1989; Cutler, Penrod, & Stuve, 1988). Survey

[14]Juror skepticism appears a misnomer for this phenomenon but is nevertheless accurate because the original empirical studies were performed within the context of eyewitness identification. In this paradigm, *juror skepticism* referred to juror disbelief of any eyewitness following presentation of expert testimony on the inaccuracy of eyewitness identification. In contrast, *juror sensitivity* referred to jurors evaluating the accuracy of the eyewitness based on the factors the experts stated would either hinder or help an accurate identification. In other words, skeptics would disbelieve any witness following expert testimony on the issue, whereas sensitive jurors would only discount the eyewitness testimony of someone who was plagued by circumstances that would probably reduce the eyewitness's accuracy.

[15]This assumes that the lighting expert testimony is sufficiently complex to warrant peripheral processing of the information by jurors. This may not be the case.

and empirical research have demonstrated that jurors strongly believe that witness confidence in identification is related to accuracy of identification (Fox & Walters, 1986; Loftus, 1980, 1986; Wells, 1986), even though they are actually only weakly correlated (Bothwell, Deffenbacher, & Brigham, 1987). In several studies, expert testimony on this issue was discounted or ignored because it did not fit with the mock jurors' strongly held preconceived notions (Cutler et al., 1988, 1989). A similar finding was made by Greenberg and Wursten (1988), who discerned that expert testimony on insanity did little to change mock jurors' attitudes and final decisions when the mock jurors held strong beliefs concerning the insanity defense prior to the trial. These strong prior-beliefs mock jurors were also found to be less likely to engage in heuristic biases (i.e., expert credentials had less effect on their decisions).

Jurors' prior beliefs have also influenced the interpretation of expert testimony in actual death penalty cases in California (Sundby, 1997). Jurors expressly stated that they ignored the testimony of experts when it differed significantly from their preconceived notions of a phenomenon. The story model of juror decision making offers an explanation for these results and highlights the importance of measuring jurors' pretrial beliefs.

Past empirical research has demonstrated that mock jurors do hold strong beliefs concerning the ability of clinicians to accurately predict future dangerousness (Morier, 1987). Individuals were found to believe that clinicians were capable of predicting future dangerousness at a far more accurate rate than empirical studies suggested. This belief may cause jurors to be unfairly biased by the expert clinical opinion testimony condition in the present experiment and to discount expert testimony and cross-examination that disagrees with their beliefs.

Psychological research has also found that expert testimony specifically linked to the facts of a case is more influential in mock juror decisions than more general expert testimony (Fox & Walters, 1986; Kovera et al., 1997; Schuller, 1992; Saks & Kidd, 1980; however, see Cutler et al., 1989 for an incongruous result). This finding has been replicated in interviews with actual jurors in death penalty cases (Sundby, 1997). Sundby found that jurors rated experts who made a connection between their expert testimony and the specific facts of the case as more credible and influential. These results, too, are consistent with the story model of juror decision making and its emphasis on relating expert testimony to other evidentiary facts.

As already noted, an expert's confidence in his opinion and the influence this confidence has on jury decisions has also been explored by psychological research.[16] It is surprising that one empirical study directly addressing this issue found that moderate levels of confidence had a greater effect on mock juror decision making than higher or lower levels of confidence. In a simulated insanity case, mock jurors were more influenced by an expert's proffered 80% confidence in his opinion than by 100% or 60% confidence (Rogers et al., 1992). In contrast, survey data have suggested that for judges, jurors, and lawyers, confidence in expert testimony opinions is crucial. Judges and lawyers rated experts' failure to

[16]This question is directly relevant to the *Barefoot* situation, in which the experts stated that they were 100% confident.

state definitive conclusions as the most troubling aspect of their relationship with experts, whereas jurors rated experts who conveyed low confidence as a signal for strong concern (Champagne, Shuman, & Whitaker, 1992; Shuman, Whitaker, & Champagne, 1994). The incongruity of the experimental and research results might be explained by the fact that firm expert conclusions are likely to be most important when jurors are using heuristic processing of testimony, but overly confident expert conclusions may be less influential when jurors are evaluating the content and context of the testimony. Regardless of the exact relationship between expert confidence and juror decisions, expert confidence in their opinions was equal in all conditions in this experiment.

Deliberation conditions may also affect the impact of expert testimony on juror decision making. Although the effects of deliberation conditions on jury verdicts remain in dispute in the jury research literature (Diamond, 1997), one study has demonstrated differential effects for expert testimony pre- and postdeliberation (Brekke et al., 1991). One-sided (clearly advocating one party's position) expert testimony was viewed more favorably than balanced testimony by jurors' predeliberation, but this effect disappeared postdeliberation.[17] This finding may indicate that deliberation influences juries to avoid heuristic evaluations of the expert by allowing them to more actively explore the content of an expert's testimony. Time pressures and testimony complexity have been noted in the literature as contributing to heuristic processing (Petty & Caccioppo, 1986). Having more time to explore the content of expert testimony through multiple viewpoints makes it more likely that the complexity of expert testimony will be reduced. Jury studies that fail to use a deliberation condition may mistakenly overvalue the importance of heuristic processing in juror decisions.

Taken as a whole, psychological research on the influence of expert testimony on juror decision making suggests the importance of controlling for or evaluating the effects of (a) the physical appearance of the expert, (b) the expert's credentials, (c) the complexity of the expert's testimony, (d) the jurors' prior beliefs concerning the facts at issue, (e) the specific link between the expert testimony and the facts of the case, (f) the confidence of the expert in his or her opinion, and (g) the deliberation of jurors. With the exception of (g), all these factors were accounted for in the current investigation. In addition, several findings from psychological research on expert testimony and jury decision making suggest important direct effects in this experiment. First, jurors' preconceived notions that clinicians are capable of accurately predicting dangerousness may cause mock jurors to overly weigh clinical opinion testimony while discounting information from cross-examination and competing experts that disagrees with their position. Second, jurors' bias toward case-specific information over more statistical information may cause mock jurors to be more influenced by clinical opinion testimony than by actuarial testimony.

---

[17]The Brekke et al. study also found differences between adversarial testimony and court-appointed testimony pre- and postdeliberation. This result is not explored because nonadversarial experts were not used in this experiment (see Brekke et al., 1991, and Crowley, O'Callaghan, and Ball, 1994, for a discussion of the impact of nonadversarial experts).

Cognitive Biases in Clinical Predictions of Dangerousness

A large body of literature suggests that, as clinicians, mental health practitioners exhibit an unimpressive ability to make accurate future predictions of dangerousness (Faust & Ziskin, 1988; Grisso & Appelbaum, 1992; Hart, Webster, & Menzies, 1993; Lidz, Mulvey, & Gardner, 1993; Melton et al., 1997; Monahan, 1981; Monahan & Steadman, 1994; Otto, 1992; Showalter, 1990; Shuman & Sales, 1998). Recent developments in risk assessment research and methodology indicate, however, that the clinical judgment of practitioners may not be as inaccurate as originally believed (Mossman, 1994).[18] Monahan's (1981) review of the literature, which was presented in the American Psychiatric Association amicus brief in *Barefoot,* estimated clinical prediction error rates between 65% and 85%. These estimates of accuracy, however, appear to have been artificially low, artifacts of the low base rates of violence (2–5%) found in the early clinical dangerousness prediction studies (Borum, 1996; Monahan & Steadman, 1994). Later studies of dangerousness assessment, with base rates for violence between 25% and 50% (Borum, 1996; Klassen & O'Connor, 1988; Monahan & Steadman, 1994; Otto, 1992), found error rates closer to 44% (Otto, 1992). The difference between the base rates of violence of the early and later studies is thought likely attributable to improved outcome measures of violence and more appropriately selected samples (Borum, 1996; Melton et al., 1997).[19]

Early studies of clinical predictions of violence also suffered from a number of other methodological problems, including using administrative rather than clinical decisions to determine the release of patients and substantial portions of patients never being released into the community (Melton et al., 1997). The most recent estimates of the ability of mental health practitioners to predict dangerousness using more sophisticated statistical analysis techniques suggest that mental health practitioners are moderately better than chance in predicting future dangerousness (Gardner, Lidz, Mulvey, & Shaw, 1996; Mossman, 1994). Mossman's (1994) review of 15 clinical opinion prediction data sets, analyzed by receiver operating characteristics (ROC) curves, found significantly higher-than-chance predictive accuracy (area under the curve [AUC] = .6718).[20] Nonetheless, the accuracy of mental health practitioner's predictive assessments based on their expertise and experience are still significantly below those achieved using actuarial methods of estimation (AUC = .7130; Gardner et al., 1996; Mossman, 1994). Irrespective of their inaccuracy, mental health practitioners have continued to assert their expertise in court (e.g., Faust & Ziskin, 1988; Melton et al., 1997).

---

[18]In fact, one recent study suggested that clinicians may be quite accurate in making short-term predictions of dangerousness under certain conditions (see McNeil et al., 1998).

[19]Studies began to include self-reports of violence, hospitalization, and reports of collateral sources as evidence of violence instead of just rearrest for violent crimes (Monahan & Steadman, 1994).

[20]ROC analysis is based on graphs of true hit rates (sensitivity) as a product of the false alarm rates (1-specificity) and has been shown not to be susceptible to changes in base rates (Rice & Harris, 1995). The area under the curve (AUC) is simply the likelihood that a randomly selected violent patient would have been predicted to be more violent than a randomly selected nonviolent person (this is also known as a common language effect size; Gardner et al., 1996; Mossman, 1994; Rice, 1997; Rice & Harris, 1995).

A myriad of reasons have been advanced for mental health practitioner deficits in predicting dangerousness. The two most common explanations are: (a) mental health practitioners lack specific training and knowledge in risk assessment, and (b) mental health practitioners rely on cognitive biases or inappropriate heuristics when they attempt these predictions. These two explanations are not distinct, as it is likely that lack of training leads to more excessive use of cognitive biases. To date, there exists no consensus on appropriate standards of training or assessment for mental health practitioners interested in dangerousness assessment (Borum, 1996). Staying informed about the most promising developments in dangerousness assessment is also difficult because the relevant literature is extremely complicated and disjointed (Mulvey & Lidz, 1995). Furthermore, current understanding of dangerousness is far from complete. Even if more effective training were available, it is not clear that it would improve clinical predictions of dangerousness. In a number of areas other than violence prediction, more intensive clinical training and experience has yet to produce greater predictive accuracy (Arkes & Hammond, 1986; Goldberg, 1968; Grove & Meehl, 1996; Smith & Frank, 1997).

This conclusion is not surprising. In addition to relying on cognitive biases and heuristics that affect the judgments of ordinary people under conditions of uncertainty (Kahneman, Slovic, & Tversky, 1982),[21] mental health practitioners have been found to poorly combine information, use irrelevant information, and inappropriately vary the information they use in formulating predictions for an individual (Einhorn, 1974; Gaeth & Shanteau, 1984; Grove & Meehl, 1996). Worse, their propensity for gathering excessive and irrelevant information also likely leads mental health practitioners to have greater confidence in their conclusions than is warranted (Arkes & Hammond, 1986; Shuman & Sales, 1998).

Mental health practitioners' dangerousness estimates are undoubtedly also plagued by a number of systematic distortions inherent to their task (Shuman & Sales, 1998). The fundamental attribution error,[22] which causes individuals to incorrectly perceive that another's behavior is based on stable dispositions (i.e., traits) rather than situational contexts, is likely to bias clinical predictions of dangerousness. As a result of this error, mental health practitioners tend to view an individual's past or present violent acts as stable character traits and not as environmentally caused events, leading them to incorrectly predict future acts of violence on the basis of violent disposition. This misattribution may even be enhanced in clinicians who are specifically trained to investigate the stable

[21]This is especially troubling because laypersons have failed to make competent decisions in this area. Marquart et al. (1989) compared 107 criminal defendants judged to be dangerous by capital sentencing juries, who later had their sentences commuted, with 92 criminal defendants who were not adjudicated dangerous by capital sentencing juries. They found no differences between the two groups in prior offenses (number and violence level), severity and type of present offense, or future acts of violence both within and outside the prison environment (Marquart et al., 1989). Unfortunately, this study did not directly examine the effects of expert testimony, but it did indicate that this testimony was one of the three major factors (expert testimony, present offense, and past criminal record) used by the jurors in answering the dangerousness question and that it was frequently presented in these sentencing hearings.

[22]Shuman and Sales (1998) have described this bias as an inability to attribute causality appropriately.

psychic features of individuals rather than the effects of the environment on their behavior (Smith & Frank, 1997).[23]

Clinical estimates of dangerousness may also be fouled by a representativeness bias or heuristic (Melton et al., 1997; Shuman & Sales, 1998; Tversky & Kahneman, 1974). This bias causes the clinicians to compare the individuals they are assessing for dangerousness with their stereotypical conceptualization of a dangerous individual and construct a prediction on the basis of similarity. Unfortunately, the stereotypes that mental health practitioners have of dangerous individuals are likely to be inaccurate and contain many attributes that are not linked to future violence (Shuman & Sales, 1998). As a consequence, their predictions of violence are likely to be based on poor correlates of future violent behavior and, therefore, are likely to be erroneous. Yet, even if mental health practitioners' representation of a dangerous person were accurate, their predictions would still falter because "humans simply cannot assign optimal weights to variables and are not consistent in applying their own weights" (Grove & Meehl, 1996, p. 315).

Mental health practitioners also exhibit tendencies to ignore base rates of violence and to select information on the basis of their initial assessment while ignoring disconfirming evidence in making assessments of dangerousness (Monahan, 1981; Monahan & Steadman, 1994). Base rates of violence, which refer to calculations of how common violent behavior is within a specific population, must be taken into account in estimating violence for any individual. Numerous studies have demonstrated that if base rates are low—as they are in the occurrence of violent behavior—and are not used in estimating the occurrence of event, subsequent predictions will greatly overpredict the likelihood of a future violent event for an individual (Melton et al., 1997). Yet, mental health practitioners have been shown to be unaware of the relevance of this information, and, when made aware, have failed to incorporate it into their predictions. This failure leads clinicians to greatly overestimate the likelihood that a specific individual will commit future violent acts (Monahan & Steadman, 1994).

Another systematic distortion of information processing that can produce overestimates of dangerousness is the propensity of mental health practitioners to ignore evidence that disconfirms their initial opinion while they continue to select information that supports it (Melton et al., 1997). This overestimation bias may occur as a result of the saliency and recency of the criminal defendant's present violent criminal offense (Melton et al., 1997). The death penalty would not be at issue if the individual had not been convicted of committing a serious violent criminal act.

In sum, the literature suggests three central facts: (a) Mental health practitioners inaccurately make future violence predictions, (b) mental health practitioners lack training in making violence predictions, and (c) mental health practitioners' dangerousness predictions are biased by their reliance on a number of cognitive

---

[23]The existing actuarial instruments also suffer from this problem. They emphasize the importance of stable characteristics of the individual over the environmental constraints that may have led to violence. In the actuarial instruments, however, the stable psychic characteristics used in violence prediction have been empirically demonstrated to predict violence, whereas the characteristics used by clinicians in their violence predictions often have not been empirically tested.

heuristics, which causes them to overestimate rates of future violence. Although the cognitive errors clinicians make in assessing dangerousness are not directly manipulated in this study, the weight jurors grant to these inaccurate opinions is examined. Additionally, the causes of clinical inaccuracy in dangerousness prediction are incorporated in the cross-examination and competing expert conditions of this experiment.

Recent Developments in Risk Assessment and Violence Prediction

The development of actuarial instruments specifically designed to forecast risk of future dangerousness has significantly improved the accuracy of predictions of future violence (Borum, 1996; Gardner et al., 1996; Mossman, 1994; Rice & Harris, 1995). Actuarial instruments are created from empirically established risk factors (i.e., past history of violence, psychopathic tendencies, prior arrest, etc., for dangerousness prediction), which have been demonstrated to predict the appropriate outcome (i.e., violence) across groups within the selected population (Monahan & Steadman, 1994). These risk factors are then statistically weighted in a way that maximizes the power of the instrument to predict a specified outcome.

A small number of actuarial instruments developed during the 1970s, 1980s, and 1990s exist to predict future violence (Borum, 1996). Their construction was initiated after research demonstrated the unreliability of pure clinical opinion predictions (Melton et al., 1997; Monahan & Steadman, 1994). The development of these actuarial instruments is not surprising given that actuarial prediction has been found to outperform clinical opinion in a large number of areas. Grove and Meehl (1996) found, in a review of 136 studies in which clinical and actuarial predictive accuracy was compared, that actuarial assessments were significantly superior to clinical predictions in 64 studies whereas clinicians were more accurate only in 8 instances. As a group, actuarial dangerousness prediction instruments have been demonstrated to significantly outperform pure clinical predictions, with some instruments substantially outperforming others (Gardner et al., 1996; Mossman, 1994; Rice, 1997; Rice & Harris, 1995).

Borum (1996) reviewed the existing dangerousness prediction actuarial instruments with a standardized scoring system and reported promising results for the Violence Risk Assessment Guide (VRAG; classification accuracy = ~75%; Rice & Harris, 1995), promise for the HCR-20 (Webster, Eaves, Douglas, & Wintrup, 1995), and poor results for the Dangerous Behavior Rating Scale (DBRS; Menzies & Webster, 1995). The MacArthur Risk Assessment group is validating an actuarial instrument comprising over 30 assessment measures at five sites across the United States (Monahan & Steadman, 1994). This theory-driven instrument contains variables representing demographic, personality, personal history, contextual, and clinical domains.

Actuarial dangerousness screening instruments and decisional trees have also been advocated by psychologists to formulate more successful violence predictions. Gardner et al. (1996) demonstrated encouraging results for several screening instruments that significantly outperformed the accuracy of pure clinical opinion predictions of dangerousness (AUC = .618 for clinical, AUC = .736 for full screen). It is surprising that, in Gardner et al.'s (1996) study, a screening instrument based on historical behavior alone (prior incident of violence and seriousness of violence) performed no differently than an instrument that contained those variables plus a variety of other measures (age, hostility measured on the Brief Symptom Inventory, drug use, and a thought disorder measure). Gardner et al. (1996) has also advocated the use of a dangerousness assessment decision tree, which has demonstrated some success in predicting violence. It is based on hostility ratings, age of first offense, the number of prior violent offenses, and drug use. Melton et al. (1997) have also proposed a decision tree for the assessment of dangerousness in their handbook for mental health professionals who perform evaluations for the courts. For this instrument, risk is a product of scores on a psychopathy instrument (the Psychopathy Checklist—Revised [PCL–RI]), age of delinquency, adult criminal record, and clinical diagnosis. No reliability or validity data are yet available for this instrument. It is surprising that none of the above-mentioned instruments (actuarial or decisional) are frequently used by clinicians in their everyday practice to predict future violence (Gardner et al., 1996).

Actuarial dangerousness prediction instruments are not without their critics. Psychology commentators have criticized these instruments for ignoring case-specific information (Melton et al., 1997). For example, it is possible that some attribute of the individual not assessed by the instrument greatly reduces his or her chance of committing a future violent act (e.g., a brain tumor or being paraplegic), or some factor not assessed by the instrument protects the individual from committing a future violent act (e.g., a stable marital relationship). In such instances, an assessment based on the instrument would undoubtedly lead to an unjust outcome for that individual. Forensic experts have also suggested that actuarial evidence might be difficult to present effectively in court so that jurors can appropriately weigh it in their decision making (Melton et al., 1997). As mentioned previously, jurors clearly have difficulty evaluating and interpreting complex statistical information and may process it in a biased fashion (i.e., on the basis of the expert's credentials and final opinion rather than on the content). Alternatively, legal commentators have posited that jurors might rely too heavily on actuarial testimony in their decision making. For example, Tribe (1971) expressed concern that "when risk estimates are expressed in numerical form, the trapping of precision and objectivity may suppress appropriate doubts about validity and certainty of prediction" (p. 1353). This criticism is unsubstantiated by empirical evidence. If anything, the weight of evidence suggests that jurors place less weight on statistical measures than on other testimony, especially clinical evidence (Melton et al., 1997; Smith, Penrod, Otto, & Park, 1996).

In sum, a review of the risk assessment literature indicates that the VRAG is the most reliable and valid dangerousness assessment instrument currently available. As a consequence, it was the actuarial method of prediction used in this study.

Research Hypotheses

The various research literatures summarized above do not directly address the question of the effect of psychological expert dangerousness testimony on juror decision making, nor do these literatures answer the question of the tenability of the assumptions underlying the United States Supreme Court's *Barefoot* decision (i.e., that adversary procedures will remove the bias caused by inaccurate expert dangerousness opinions in juror decisions). They do, however, underscore the importance of these issues. In the present investigation, we directly examined the effects of expert dangerousness testimony on juror decision making in capital sentencing. To accomplish this goal, a simulated capital sentencing case was presented to mock jurors using both written and videotaped materials, and the influence of expert testimony on final mock juror dangerousness decisions was assessed using a number of dependent measures.

In this experiment, a 2 × 4 factorial crossed design was used (see Table 1 for the eight different groups used). The type of expert testimony presented on dangerousness was varied (clinical vs. actuarial) and adversarial procedures were manipulated (ineffective[24] cross-examination, effective cross-examination, competing expert of the same type as the first expert, and competing expert of a different type than the first expert). Each participant made assessments concerning (a) pre-experiment beliefs concerning the death penalty; (b) beliefs about and confidence in the criminal defendant's dangerousness following presentation of the written materials[25] ; (c) belief about and confidence in the criminal defendant's dangerousness following presentation of expert testimony; (d) evaluation of the expert's credibility, level of scientific knowledge, and persuasiveness; (e) weight placed on the present offense, past offense, and expert testimony by the mock juror in reaching his or her decision; and (f) reevaluation of questions (1–5) after the experimental adversarial manipulations (i.e., cross-examination or competing expert).

The following effects were predicted:

1. Expert testimony, regardless of type, will influence mock jurors' dangerousness ratings.
2. The influence of clinical opinion expert testimony on mock jurors' dangerousness rating will be equal to or will exceed the effects of actuarial expert testimony both immediately after presentation of this testimony and after the presentation of the adversary procedures (cross-examination or competing expert).
3. The adversarial manipulations will negatively influence mock jurors' final dangerousness ratings and their ratings of the first expert's credibility, persuasiveness, level of science, and influence.

---

[24]Lawyers have been routinely criticized for their inability to effectively cross-examine psychology experts (Melton et al., 1997).

[25]Analyses were performed on a composite variable created from the participants' belief about whether the death penalty was warranted for the defendant and how confidently participants felt in that belief. The valence of the variable was determined by whether they believed the death penalty was appropriate in this situation (plus sign for warranted, and minus sign for unwarranted), and their rating on a Likert scale determined their confidence in their belief. As a result of the use of this composite variable in all of the analyses, the study was limited to exploring participants' change in confidence in their beliefs due to experimental manipulations rather than whether they changed their overall opinion about whether the death penalty was warranted. It is not clear from this study how participants' change in confidence in their beliefs would affect their decisions on whether the death penalty was warranted. Future research is needed to explore this issue.

Table 1
Different Groups Represented in 2 × 4 Experimental Design

Type of expert testimony	Adversary manipulation: Ineffective cross-examination	Effective cross-examination	Competing expert same as first expert	Competing expert different from first expert
Clinical opinion	Clinical, ineffective cross-examination	Clinical, effective cross-examination	Clinical–Clinical	Clinical–Actuarial
Actuarial	Actuarial, ineffective cross-examination	Actuarial, effective cross-examination	Actuarial–Actuarial	Actuarial–Clinical

4. The various adversarial manipulations will have a differential negative effect on the final dangerousness rating made by mock jurors. The ineffective cross-examination should have the weakest negative effect on mock jurors' final evaluations of dangerousness, whereas the competing expert conveying a different type of testimony than the first expert should have the strongest negative effect on mock jurors' final evaluations of dangerousness.

Method

Participants

Participants in this study were 208 University of Arizona psychology undergraduates: 123 women (59%) and 85 men (41%). All participants were seeking required research credits for their undergraduate psychology classes. Participants described themselves as Caucasian (67%), Hispanic (21%), Asian (7%), or African American (5%). Participants ranged in age from 18 to 45 years (M = 20.9, SD = 1.98). Approximately 25 students were tested in each of the eight possible conditions. Mock juries ranged in size from 4 to 15 people. Prior to the administration of any materials, participants rated their initial beliefs concerning the death penalty and recorded their demographic characteristics.[26]

Procedure

Participants took part in a simulated capital sentencing hearing based on procedures and instructions used in Texas. Participants were first led into a courtroom and were questioned about their views concerning the death penalty.[27] Participants then filled out a questionnaire detailing their demographic characteristics. After completing the questionnaire, they were seated in the jury box of the courtroom. They were subsequently handed

[26]Of the participants, 27% described themselves as not politically affiliated, 39% described themselves as Democrats, 22% described themselves as Republicans, 11% described themselves as independents, and 1% described themselves as libertarians. A total of 23% of the participants had been called for jury duty, whereas 76% had not participated in jury duty. In the study, no statistically significant differences in mock jurors' dangerousness ratings were found on the basis of sex, age, ethnic background, political affiliation, age, or participation in jury duty. Consequently, these variables were ignored in later analysis of the data.

[27]Any student morally opposed to the death penalty, who could not offer a fair and impartial decision concerning this issue, was excused (see *Witherspoon v. Illinois,* 1968). Future research may wish to explore if "Witherspoon excludables" are more or less susceptible to the biases found in this study.

an envelope that contained simulated jury instructions for capital sentencing based on dangerousness as they appear in Texas and were given a written summary of a modified capital sentencing case. The participants were allowed time to read the materials contained within the envelope and to ask any questions that they might have. They were told not to discuss the case with any other jury members and were told that it was important that they make all of their decisions as if they were participating in a real death penalty case. After reading the written material, participants rated their beliefs concerning the criminal defendant's dangerousness and their confidence in that belief[28] (Time 1 or baseline).

The initial expert testimony (either actuarial or clinical) was then presented via videotape, and the participants reevaluated the criminal defendant's dangerousness and their confidence in that belief, and rated the expert's level of scientific knowledge, persuasiveness, credibility, and the influence the expert's testimony had on their decision. They also weighted the influence of three factors on their decisions (past criminal history, present offense, and expert testimony; Time 2). These evaluations were placed in their envelopes and then the adversary experimental manipulations were presented via videotape (ineffective cross-examination, effective cross-examination, competing expert of the same type as expert 1, or a competing expert of a different types than expert 1). Subsequently, the participants reassessed the same measures they had previously completed (Time 3) and were debriefed (see Table 2 for a visual presentation of the outline of the experiment).[29]

Experimental Conditions

The action depicted in the videotape occurred in a courtroom, with all the roles played by actual lawyers and clinical psychologists. To simulate the conditions of a real trial, we had mock jurors view the videotape in a law school courtroom and sit in an actual jury box while reading and reviewing the evidence and making their decisions.

Experimental Stimuli

The simulated capital sentencing materials were based on a Texas death penalty case, *Moore v. State,* 542 S.W.2d 664 (Tex. Cr. App. 1976). In this case, the criminal defendant was found dangerous and received the death penalty. The description of the crime was taken directly from the reported case with some slight modifications (i.e., names were changed, etc.). The severity of the crime was altered so that it would not appear too heinous (i.e., descriptions of a sexual assault were removed).

A confession contained in the reported case was also included in the written materials, as was a synopsis of the defendant's past criminal history. The four past crimes reported in the stimulus materials were based on Marquart et al.'s (1989) study of capital sentencing in Texas. They reported past criminal histories for 102 criminal defendants

---

[28]Diamond (1997) detailed the importance of rating both the decision of jurors (e.g., guilt or innocence, dangerousness or not dangerousness) and the confidence jurors have in their decisions.

[29]The repeated presentation of the study's measures may have sensitized the mock jurors to the measures and influenced their decision-making process. One study has found increased reliance on evidence as a result of its repeated assessment (Faigman & Baglioni, 1988), whereas another study has demonstrated the opposite effect (Pennington & Hastie, 1992). A high correlation between mock jurors' dangerousness ratings at Time 1, Time 2, and Time 3 (Time 1–Time 2, $r = .856, p < .01$; Time 1–Time 3, $r = .837, p < .01$; Time 2–Time 3, $r = .912, p < .01$) suggests that the repeated measures design may have decreased variability in participants' ratings over the three time periods. Follow-up research should vary the presentation of dangerousness ratings to determine if the repeated measures design significantly contributed to or detracted from the experimental results.

Table 2
Chronological Experimental Sequence

Sequence	Procedure
Demography	Participants rate strength of prior beliefs concerning death penalty
Written materials presented	
Time 1 or baseline	Participants rate strength of belief concerning dangerousness of the defendant
First expert testimony presented	
Time 2	Participants rate (a) strength of belief concerning dangerousness of the defendant, (b) influence of expert in decision making, and (c) expert's testimony
Second expert or cross-examination	Ineffective or effective cross-examination OR clinical or actuarial competing expert
Time 3	Participants rate (a) strength of belief concerning dangerousness of the defendant, (b) influence of both experts in decision making (if applicable), and (c) both experts' testimony (if applicable)

adjudicated as dangerous by capital sentencing juries. The modal number of crimes for these defendants was 3–5, with no past serious violent offenses (Marquart et al., 1989).[30]

Expert Testimony

Clinical opinion expert. The clinical opinion expert testimony on dangerousness was based on actual testimony presented in capital sentencing trials in Texas. The credentials of the expert were created by Daniel A. Krauss to approximate those held by actual experts in capital sentencing in Texas. The expert testimony consisted of statements of (a) his experience and education, (b) his interview of the defendant, (c) his opinion that the defendant is a severe sociopath, and (d) his belief, based on his years of experience, that the defendant definitively represents a continuing danger to society.

Expert actuarial testimony. The initial actuarial testimony[31] was presented by an psychologist with the identical experience and training as the clinical opinion expert. This

[30]In this experiment, the defendant's past crimes were (a) shoplifting at age 15; (b) disorderly conduct at age 17; (c) joyriding at age 22; and (d) possession of marijuana at age 22. He was on probation for the last two crimes.

[31]The Violence Risk Assessment Guide (VRAG) which was relied on in this experimental condition, consists of twelve variables: the Hare Psychopathy Checklist (PCL-R), elementary school maladjustment, age of present offense, diagnosis of personality disorder (*DSM–III*), separation from parents under the age of 16, failure in prior conditional release, criminal history for property offenses, marital status, diagnosis of schizophrenia (*DSM–III*), victim injury in present offense, history of alcohol abuse, and female victim in present offense. The VRAG was normed on 618 men who committed at least one serious offense and who were followed for 10 years (31% committed a violent act over this time). It accurately predicted dangerousness in 74% of the cases and demonstrated AUC = .76, r = .44, sensitivity = 60%, and specificity = 78% at a 10-point cut-off (Rice & Harris, 1995). It was revalidated on 868 men possessing similar characteristics, and results were substantially the same: clinical accuracy = 70%, AUC = .74, sensitivity = 62%, and specificity = 72% at a 10-point cut-off (Rice & Harris, 1995). The instrument has also demonstrated some success with prediction of dangerousness in sex offenders (Rice, 1997).

expert testimony consisted of statements of (a) his experience and education, (b) his interview of the defendant,[32] (c) his description of the VRAG assessment instrument, (d) his use of VRAG with the criminal defendant, and (d) his belief, based on the VRAG, that the defendant represents a continuing danger to society.

Adversary Manipulations

Ineffective cross-examination. In the ineffective cross-examination condition, the attorney attacked the credibility of the expert but not the content of his testimony. Ineffective cross-examination did not differ across conditions and was presented by the same lawyer. The ineffective cross-examination highlighted that: (a) the expert was not a medical doctor, (b) the expert worked for the state and was being compensated for his testimony, and that (c) the expert often testified in these trials.

Effective cross-examination of the clinical opinion expert. In the effective cross-examination condition, the attorney attacked the content of the expert's testimony. The effective cross-examination of the clinical expert highlighted that: (a) the dangerousness predictions made by clinicians are often wrong, (b) there is a consensus of experts in dangerousness prediction that clinicians have no special ability to predict dangerousness, (c) the testifying expert has not received any special training in dangerousness prediction, (d) clinical opinions tend to overestimate future violence and are based on faulty reasoning, (e) the testifying expert was unaware of base rates of violence of this criminal population, (f) actuarial techniques of dangerousness assessment have shown much higher validity than clinical instruments, and (g) the confidence the testifying expert has in his opinion has not been linked to accuracy in long-term dangerousness predictions.[33]

Effective cross-examination of the actuarial expert. In the effective cross-examination condition, the attorney attacked the content of the expert's testimony. The effective cross-examination condition of the actuarial expert questioned (a) the use of the VRAG on a population on which it is not validated,[34] (b) the error rate of the instrument, and (c) the failure of the VRAG instrument to take into account case-specific information.

Competing Experts

In all conditions, the competing expert was portrayed by the same actor, whose credentials, training, and experience were similar to that of the original expert.[35]

---

[32]This part of the videotape testimony was the same footage as that used in the clinical opinion expert testimony condition.

[33]This is true for empirical studies concerning long-term prediction of a criminal population's dangerousness and violence. A recent short-term study of a psychiatric population has demonstrated that confidence can be linked to accuracy of clinical predictions of future dangerousness (McNeil et al., 1998).

[34]The VRAG's validation on a serious violent offender population makes it the most appropriate instrument for criminal defendants facing capital sentencing. The VRAG, however, has not been validated on a population in the United States, and this may limit its generalizability to an American criminal population.

[35]Statistical analysis suggested that this was, in fact, the case. A t test comparing the first expert to the second expert demonstrated no significant differences overall between the two experts on a composite factor score (a factor score composed of participants' ratings of the level of science, persuasiveness, influence, and credibility of each expert and his testimony). Some differences between the experts were found on individual variables. The first expert was rated as initially more influential by mock jurors than the second expert (df = 107, p = .036, two-tailed), whereas the second expert was initially rated as more scientific than the first expert (df = 107, p = .010, two-tailed) regardless of the type of testimony delivered by the expert. After the presentation of the

Competing expert clinical opinion testimony against initial clinical opinion expert. In this condition, the competing expert disagreed with the conclusion of the first expert. His expert testimony consisted of statements of (a) his education and credentials, (b) his interview of the defendant, (c) his belief that the criminal defendant was not a severe sociopath, and (d) his belief, based on his years of experience, that the criminal defendant was not a continuing danger to society.

Competing expert actuarial testimony against initial actuarial expert. In this condition, the competing expert disagreed with the conclusion of the first expert. This expert's testimony reproduced (a) and (b) from the first competing expert condition and added statements concerning (c) his use of the VRAG to assess this criminal defendant, and (d) his opinion, based on the VRAG, that the defendant does not represent a continuing danger to society.

Competing expert clinical opinion testimony against initial actuarial expert. In this condition, the competing expert disagreed with the first expert's conclusion and attacked the content of the first expert's testimony. The responses of this expert were identical to those of the first competing expert for (a) and (b). This expert's testimony differed in that he repeated the statements in the effective cross-examination of the actuarial expert (a–c) and added statements to the effect that (d) he did not believe in actuarial predictions of dangerousness; (e) he believed no instrument can predict dangerousness better than a trained professional; and (f) he believed, on the basis of his years of clinical experience, that the defendant did not represent a continuing danger to society.

Competing expert actuarial testimony against initial clinical opinion expert. In this condition, the competing expert disagreed with the first expert's conclusion and attacked the content of the first expert's testimony. This expert's testimony repeated statements (a) and (b) from the first competing expert condition, included testimony concerning all of the statements presented in the effective cross-examination of the clinical opinion expert (a–g) condition, and added statements consisting of (h) his description of the VRAG, (i) his use of the VRAG for this criminal defendant, and (j) his opinion, based on the VRAG, that the defendant did not represent a continuing danger to society.

Dependent Measures

A series of dependent variables were collected at different times throughout the experiment: (a) prior to presentation of written materials, (b) after the presentation of written materials (Time 1 or baseline), (c) after the presentation of the expert testimony (Time 2), and (d) after the presentation of the adversary manipulation (Time 3). These measures were designed to assess (a) the mock jurors' prior beliefs concerning the death penalty, (b) mock jurors' beliefs regarding the dangerousness of the defendant, (c) the mock jurors' perception of the expert testimony, and (d) the basis of mock jurors' decisions about dangerousness.

Mock jurors' prior beliefs. Prior to the presentation of the stimuli written materials, the attitudes of all mock jurors toward the death penalty were measured by (a) a yes/no response to whether they favored the death penalty and (b) the strength of that belief on a scale of 1–8. These two responses were combined to form a composite prior beliefs score from −8 (*strongly opposed*) to 8 (*strongly in favor of the death penalty*). Participants scores ranged from −6 to 8 (M = 3.58, SD = 3.32).[36]

---

adversary manipulations, the second expert was also evaluated by the mock jurors as more scientific than the first expert (df = 107, p = .001), but no other differences between the experts were discerned. The two experts were rated by the mock jurors as contributing the same amount to their final decisions on dangerousness (df = 105, p = .517).

[36]This measure was correlated with later ratings of dangerousness (death penalty rating and

Mock jurors' dangerousness beliefs. Mock jurors' beliefs regarding the defendant's dangerousness were measured at three separate times: (a) after reading the stimulus materials (Time 1 or baseline), (b) after the presentation of the first expert testimony (Time 2), and (c) after the cross-examination or second competing expert (Time 3). At each of these junctures, mock jurors' beliefs concerning dangerousness were assessed by (a) a yes/no decision on dangerousness and (b) the mock juror's confidence in that decision (1–8). The mock jurors' responses to these questions were transformed into a scale ranging from −8 (*very confident the criminal defendant is not dangerous*) to 8 (*very confident that the criminal defendant is dangerous*).

Mock jurors' perceptions of the expert. Mock jurors also rated the credibility, influence, persuasiveness, and level of science of the first expert's testimony on two separate occasions. Each juror's assessment of the expert testimony was completed after the testimony of the first expert (Time 2) and after either the cross-examination of the expert or after the testimony of the second expert (Time 3). Additionally, in those conditions in which a competing expert was used, participants also rated the second expert testimony on credibility, influence, persuasiveness, and level of science (Time 3). Each juror's perception of the expert was assessed by their answers to the following questions: (a) How scientific do you feel the expert was? (1–10),[37] (b) How persuasive do you feel the expert was? (1–10), (c) How credible do you feel the expert was? (1–10), and (d) How influential do you feel the expert was? (1–10).

Basis of mock jurors' decisions on dangerousness. Mock jurors also rated what they believed constituted their decision on dangerousness at two separate times. They rated the percentage that expert testimony, the past offense, and the present offense (0–100%) contributed to their estimates of the defendant's dangerousness after the first expert (Time 2) and after either the cross-examination or competing expert condition (Time 3). In the experimental conditions including a competing expert, mock jurors also rated the contribution (0–100%) of the testimony of the competing expert to their view of the defendant's dangerousness (Time 3).

Results

Mock Jurors' Ratings of Dangerousness

To determine if significant changes were occurring in mock jurors' dangerousness assessments throughout the course of the experiment, a series of within-subjects, repeated measures analyses of variance (ANOVAs) were performed on the mock jurors' dangerousness ratings (see Figure 1 for a graphic presentation of mock jurors' mean dangerousness ratings at Time 1 (post-written materials), Time 2 (post-expert testimony), and Time 3 (post-adversarial manipulations)). Statisti-

Time 1 dangerousness, $r = .226, p < .05$; death penalty rating and Time 2 rating of dangerousness, $r = .185, p < .05$; death penalty rating and Time 3 rating of dangerousness, $r = .192, p < .05$). This measure did not explain, however, a significant amount of variance in Time 2 and Time 3 ratings of dangerousness when Time 1 ratings of dangerousness were taken into account in statistical analyses. Thus, mock jurors' death penalty ratings were only linked to their initial ratings of the dangerousness of the defendant and were not related to their dangerousness ratings following expert testimony and adversary procedures. Because Time 1 ratings of dangerousness were used as a covariate in all analyses because of initial group differences in dangerousness ratings following the presentation of the written materials, death penalty ratings were not used in later statistical analyses.

[37]For all scales, a rating of a 1 signified that the mock juror did not believe that the expert testimony was persuasive (or credible, scientific, etc.) and a 10 represented that the mock juror believed that the expert testimony was completely persuasive (or credible, scientific, etc.).

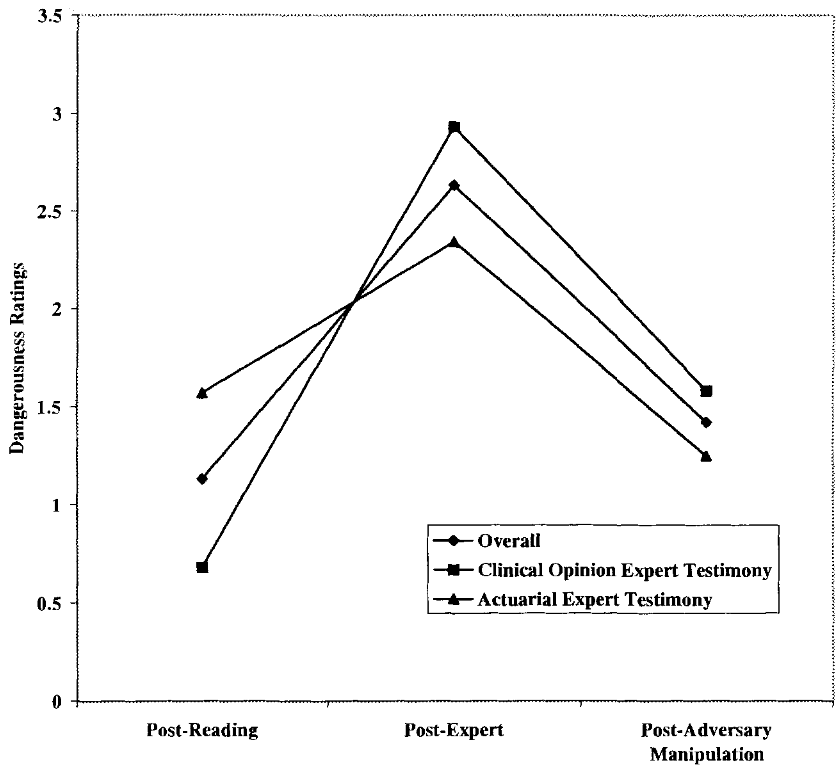

Figure 1. Mock jurors' dangerousness ratings. Ratings shown are jurors' actual ratings. When statistical analyses were performed, post-reading dangerousness ratings served as the covariate to control for initial group differences.

cal significant differences in rating were found for all within-subjects, repeated ANOVAs (see Table 3 for a summary of these data) with the exception of Time 1–Time 3. These results support Hypothesis 1 that expert testimony, regardless of type, influenced the mock jurors' dangerousness ratings. They also support the first part of Hypothesis 3, that the cross-examination and competing expert conditions negatively influenced mock jurors' dangerousness ratings.

Effect of Type of Expert Testimony on Dangerousness Ratings

After the presentation of the first expert testimony (Time 2), mock jurors' ratings of dangerousness were analyzed in an one-way analysis of covariance (ANCOVA),[38] with type of expert testimony (actuarial or clinical) as the factorial variable, to determine if the type of expert testimony mock jurors received produced significant differences in their dangerousness ratings. Clinical opinion

[38]An ANCOVA was performed because there were initial differences in baseline measures of the defendant's dangerousness between groups following presentation of the stimulus materials. This baseline or Time 1 dangerousness rating was used as the covariate in all subsequent analyses unless otherwise noted.

Table 3
Overall Within-Subjects Analyses of Variance of Mock Jurors' Dangerousness Ratings Throughout the Capital Sentencing Hearing

Condition	Dangerousness rating (*M*)	*F*	*dfs*	*p*
Time 1	1.13	80.19	2, 206	<.000[a]
Time 2	2.63			
Time 3	1.42			
Time 1	1.13	51.48	1, 207	<.000
Time 2	2.63			
Time 2	2.63	50.11	1, 207	<.000
Time 3	1.42			
Time 1	1.13	1.52	1, 207	.219
Time 3	1.42			

Note. Time 1 = post-reading; Time 2 = post-expert; Time 3 = post-adversary manipulations.
[a]Quadratic solution.

expert testimony was determined to be more effective in increasing mock jurors' initial dangerousness ratings than actuarial expert testimony (Time 2, *M*-clinical = 2.92; *M*-actuarial = 2.34), $F(1, 205) = 12.017$, $p = .001$, power = .932. A within-subjects ANOVA of both of the clinical opinion expert testimony conditions (Time 1 to Time 2, $M = .68$ to 2.92), $F(1, 103) = 40.851$, $p < .000$, power = 1.0, and the actuarial expert testimony condition (Time 1–Time 2, $M =$ 1.58 to 2.34), $F(1, 103) = 13.705$, $p < .000$, power = .956, demonstrated that each type of testimony had significant effects on mock jurors' post-expert rating of dangerousness. Figure 2 represents these findings graphically in terms of change scores from Time 1 to Time 2. The former finding supports Hypothesis 2, that expert clinical opinion testimony is more effective than actuarial expert testimony in changing mock jurors' dangerousness ratings, whereas the latter finding lends further support to Hypothesis I and the notion that expert testimony, regardless of type, will influence mock jurors' dangerousness ratings.

Effects of Type of Expert Testimony and Adversary Procedures on Dangerousness Ratings

After presentation of both expert testimony and the adversary manipulations (Time 3), a 2 × 4 (Type of Testimony × Adversary Manipulations) ANCOVA was performed to determine the singular and joint effects of the experimental manipulations. A statistically significant main effect was found for type of testimony (Time 3, *M*-clinical = 1.58, *M*-actuarial = 1.25), $F(1, 199) = 6.354$, $p = .012$, power = .708, and a main effect for the adversary manipulations (cross-examination or competing experts) approached significance: Time 3, $F(3, 199) = 2.261$, $p = .083$, power = .563. No significant effect for the interaction of type of testimony and adversary manipulation was demonstrated: Time 3, *F*(3,

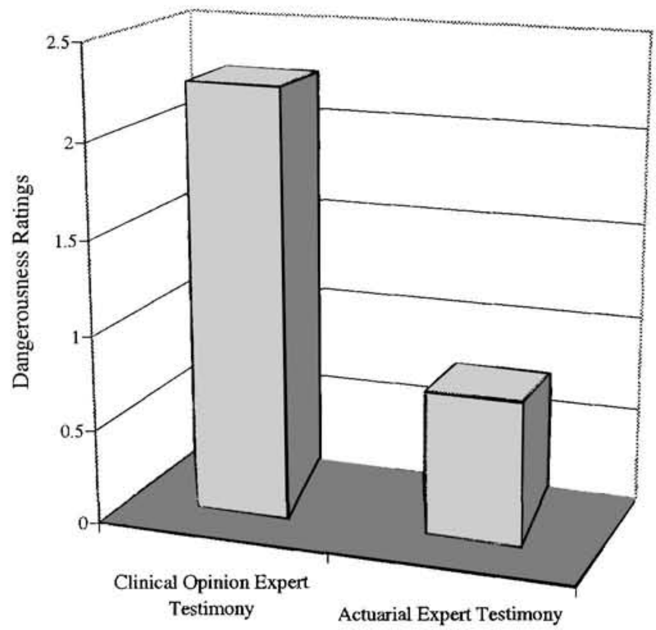

Figure 2. Change in dangerousness ratings following presentation of expert testimony.

199) = 1.078, p = .360. In addition, a within-subjects ANOVA of mock jurors' dangerousness assessments, based on the type of testimony presented, revealed that clinical opinion expert testimony influenced mock jurors' dangerousness ratings throughout the entire sentencing hearing to a greater extent (Time 1–Time 3) than did actuarial expert testimony: clinical, Ms = .68 to 1.58, $F(1, 103)$ = 7.074, p = .009, power = .750; actuarial, Ms = 1.58 to 1.25, $F(1, 103)$ = 1.102, p = .296. See Figure 1 for graphic presentation of the effects of clinical and actuarial testimony on mock jurors' rating of dangerousness throughout the sentencing hearing.) These findings suggest that the influence of expert clinical opinion testimony on mock jurors' ratings of dangerousness is greater than that of expert actuarial testimony even after adversary procedures occur, lending support for Hypothesis 2.

A one-way ANCOVA with group (clinical with ineffective cross-examination, actuarial with ineffective cross-examination, clinical with effective cross-examination, actuarial with effective cross-examination, actuarial with actuarial second expert, clinical with clinical second expert, actuarial with clinical second expert, and clinical with actuarial second expert) as the factorial variable was also performed on mock jurors' ratings of dangerousness at Time 3 to determine if any group difference existed in mock jurors' final judgments of dangerousness (see Figures 3 and 4 for visual presentation of the group means at Times 1, 2, and 3). Between-group differences were demonstrated on dangerousness ratings at Time 3, $F(7, 199)$ = 2.344, p = .025, power = .845.

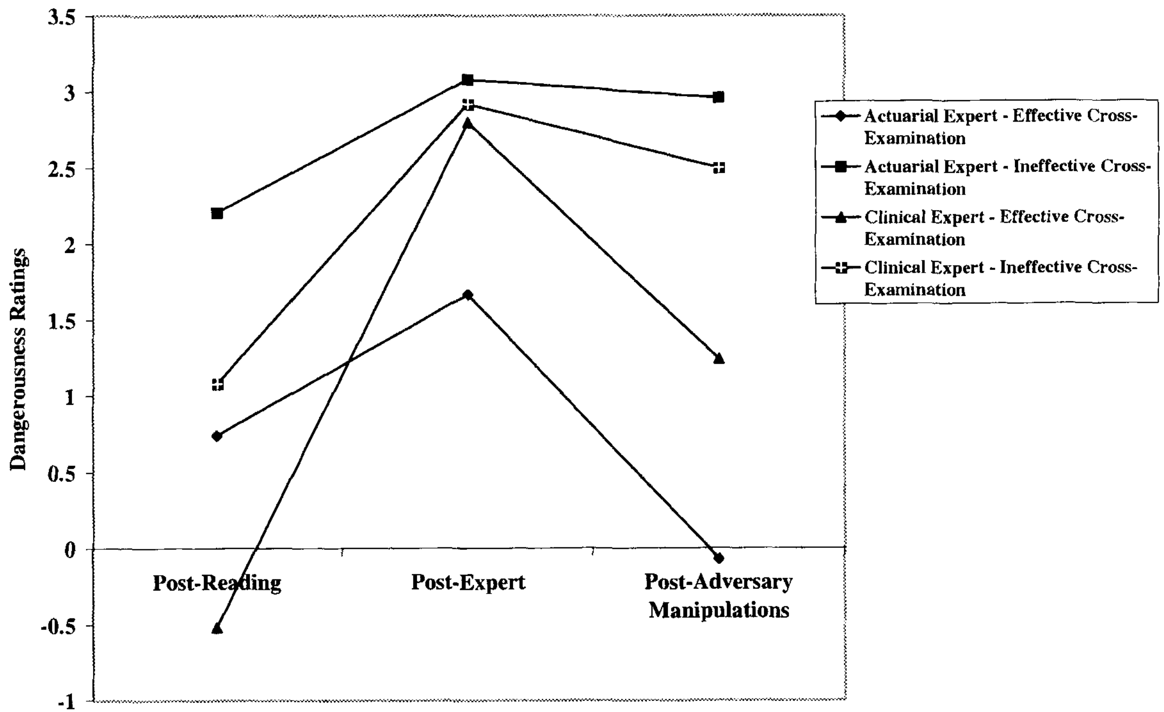

Figure 3. Mock jurors' dangerousness ratings with different types of expert testimony and different levels of cross-examination. Ratings shown are jurors' actual ratings. When statistical analyses were performed, post-reading dangerousness ratings served as the covariate to control for initial group differences.

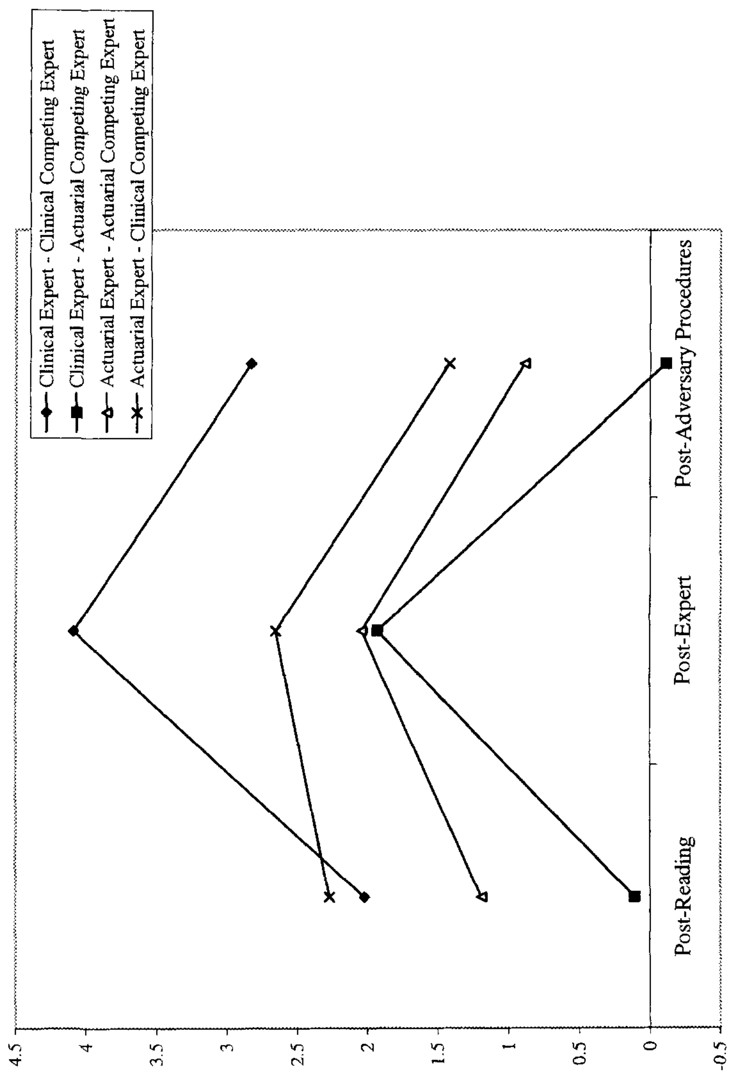

Figure 4. Mock jurors' dangerousness ratings with different types of expert testimony and various types of competing experts. Ratings shown are jurors' actual ratings. When statistical analyses were performed, post-reading dangerousness ratings served as the covariate to control for initial group differences.

The Effect of Adversary Procedures on Mock Jurors' Dangerousness Ratings

A one-way ANCOVA with adversarial manipulations (ineffective and effective cross-examination, competing expert of the same type or different type as the first expert) as the factorial variable and with Time 2 ratings of dangerousness as a covariate was performed on dangerousness ratings at Time 3 to determine if the adversary manipulations (after removing the influence of type of expert testimony) had a significant effect on mock jurors' ratings of dangerousness. A statistically significant effect was found for the adversary manipulations, $F(3, 203) = 3.489$, $p = .017$, power $= .845$. When a one-way ANCOVA with adversary manipulations as the factorial variables and with Time 2 dangerousness ratings as a covariate was performed to determine if there was a significant effect for the adversarial manipulations when the ineffective cross-examination condition was excluded, no significant effect for the adversary manipulation was found, $F(2, 156) = .475$, $p = .623$. This finding suggests that the ineffective cross-examination conditions were the major cause of differences between the adversary procedures in final mock jurors' dangerousness ratings. When this condition was excluded, the other three manipulations (effective cross-examination, a competing expert offering the same type of testimony as the original expert, and competing expert offering a different type of testimony than the original expert) had equivalent influence on mock jurors' final dangerousness ratings. This empirical finding undermines Hypothesis 4 by demonstrating that mock jurors did not differentially weight various types of adversary procedures in reaching their final dangerousness ratings. Moreover, it indirectly discredits the idea that jurors pay attention to the content of competing expert testimony when it is presented by demonstrating that more content-focused expert testimony does not have greater impact in changing juror decisions. Instead, it supports the position that jurors are simply influenced by the presence of a competing expert regardless of the content of that testimony.

Effect of Type of Testimony on Mock Jurors' Ratings of the Expert's Credibility, Influence, Persuasiveness, and Level of Science

A one-way multivariate ANCOVA (MANCOVA) with type of testimony (actuarial or clinical) as the factorial variable was performed on all four dependent measures (credibility, influence, persuasiveness, and level of science) at Time 2.[39] Multivariate statistical analysis indicated that clinical opinion expert testimony was rated more favorably on these constructs than actuarial expert testimony, $F(4, 202) = 2.548$, $p = .041$, power $= .714$. Clinical opinion expert testimony was found to be statistically superior to actuarial expert testimony on mock jurors' rating of influence, $F(1, 205) = 6.328$, $p = .013$, power $= .707$, and approached statistical superiority for persuasiveness, $F(1, 205) = 3.177$, $p = .076$, power $=$

[39]There was high correlation between ratings of each of these measures (see Table 4), and factor analysis produced only one significant factor. Due to this problem, multivariate statistical analyses were performed on these measures. Future research is needed to determine if mock jurors are in actuality rating one construct or many.

Table 4
Correlations Between Mock Jurors' Ratings of the First Expert's Credibility, Influence, Persuasiveness, and Level of Science

Characteristic	Credibility	Influence	Persuasiveness	Level of science
Credibility	—	.753	.753	.729
Influence	.753	—	.831	.650
Persuasiveness	.820	.831	—	.719
Level of science	.729	.650	.719	—

Note. All correlations were significant at $p < .000$.

.426. Mock jurors did not differentiate to a statistically significant degree between clinical opinion expert testimony and actuarial expert testimony with respect to its credibility or level of science, $F(1, 205) = .504, p > .479$, and $F(1, 205) = .027$, $p = .869$, respectively. These finding suggest that, at least on some measures, jurors respond similarly to expert clinical opinion testimony and actuarial expert testimony, whereas on other measures they tend to respond more favorably to clinical opinion expert testimony. This effect is the opposite direction of what many courts have supposed jurors would do when confronted with clinical opinion expert testimony but supports Hypothesis 2 of this study, that jurors are more influenced by clinical opinion expert testimony than by actuarial testimony.

Effect of Adversary Procedures on Mock Jurors' Ratings of the Expert's Credibility, Influence, Persuasiveness, and Level of Science

A series of within-subjects repeated measures ANOVAs were performed on all four dependent measures (credibility, persuasiveness, influence, and level of science) and their change from Time 2 to Time 3. Statically significant effects were found for all measures (credibility, *M*s = 6.4 to 5.5, $F(1, 205) = 76.007$, $p < .000$, power = 1.0; influence, *M*s = 5.9 to 5.2, $F(1, 205) = 37.692$, $p < .000$, power = 1.0; persuasiveness, *M*s = 5.6 to 5.1, $F(1, 205) = 26.429$, $p < .000$, power = 1.0; level of science, *M*s = 5.2 to 4.7, $F(1, 205) = 23.828$, $p < .000$; power = 1.0. Adversary manipulations effectively altered mock jurors' ratings on all of these measures. This is further support for the strength of the cross-examination and competing experts in negatively changing mock jurors' views of the first expert, and it is further support for Hypothesis 3, that adversary procedures lessen the influence of expert testimony.

Effects of Type of Testimony and Adversary Procedures on Mock Jurors' Ratings of the Expert's Credibility, Influence, Persuasiveness, and Level of Science

A 2 × 4 (Type of Testimony × Adversary Procedures) MANCOVA was performed on all four dependent measures at Time 3. Multivariate analysis produced statistically significant main effects for type of testimony, $F(4, 196) = 3.554$, $p = .008$, power = .863, and adversary manipulation, Wilks's lambda $F(12, 519) = 2.172$, $p = .012$, power = .912. The interaction of these two sets of variables approached significance, Wilks's lambda, $F(12, 519) = 1.658$, $p = .073$, power = .795. The individual effects of these measures are presented in

Table 5. Effective cross-examination appears to have lowered mock jurors' final ratings of the first expert's credibility and persuasiveness regardless of whether actuarial or clinical opinion expert testimony was first presented (overall credibility $M = 5.5$, effective cross-examination $M = 4.6$, ineffective cross-examination $M = 5.1$; overall persuasiveness $M = 5.1$, effective cross-examination $M = 4.5$, and ineffective cross-examination $M = 4.7$). In addition, mock jurors' final ratings of the level of science of the testimony of the first expert seemed to be most negatively affected when clinical opinion expert testimony was followed by effective cross-examination and most positively affected when clinical opinion testimony was followed by a second clinical expert (overall $M = 4.7$, effective cross-examination $M = 3.5$, and competing clinical expert $M = 6.0$).

Effect of Type of Expert Testimony on the Basis for Mock Jurors' Ratings of Dangerousness (Influence of the Present Offense, Past Offense, and Expert Testimony)

A one-way MANCOVA with type of testimony (actuarial or clinical) as the between-subjects variable was performed on the weight mock juror's *believed* they assigned to all three dependent measures (present offense, past offense, and expert testimony) in their dangerousness ratings at Time 2. No statistically significant differences were found for any of the measures on the basis of type of expert testimony first presented, $F(3, 187) = .967, p = .410$. The most striking of

Table 5
Individual Analyses of Covariance of the Type of Testimony, Adversary Manipulations, and Their Interaction on Mock Jurors' Final Ratings of the First Expert's Credibility, Influence, Persuasiveness, and Level of Science

Between-subjects variable: Final rating of first expert (Time 3)	*F*
Type of expert testimony	
Influence	2.76*
Persuasiveness	2.52
Level of science	0.52
Credibility	0.02
Adversary manipulations	
Influence	1.75
Persuasiveness	2.67**
Level of science	3.82**
Credibility	3.98***
Interaction	
Influence	1.44
Persuasiveness	0.98
Level of science	3.38**
Credibility	0.42

Note. For the type of expert testimony variable, *dfs* = 1, 199; for the adversary manipulation variable and the interaction between the two variables, *dfs* = 3, 199.
*$p < .10$. **$p < .05$. ***$p < .001$.

these nonsignificant findings is the failure of mock jurors to claim they weighted clinical opinion expert testimony more strongly than actuarial expert testimony (actuarial $M = 24\%$, $SD = 15\%$, and clinical $M = 27\%$, $SD = 19\%$), even though earlier statistical analyses demonstrate that they were more influenced in actuality by clinical opinion expert testimony than actuarial expert testimony. This suggests that mock jurors may not be cognizant of the fact that they are more influenced by clinical opinion expert testimony than actuarial expert testimony.

Effect of Adversary Procedures on the Basis for Mock Jurors' Ratings of Dangerousness (Influence of the Present Offense, Past Offense, and Expert Testimony)

Within-subjects, repeated measures ANOVAs were performed on all three dependent measures (present offense, past offense, and expert testimony) from Time 2 to Time 3. The degree to which mock jurors reported that they relied on the first expert's testimony and the past offense in reaching their dangerousness determination was significantly negatively affected by the adversary manipulations: expert testimony, *M*s = 25.4% to 20.4%, $F(1, 189) = 30.233$, $p < .000$, power = 1.00; past offense, *M*s = 20.8% to 18.9%, $F(1, 189) = 10.428$, $p < .001$, power = .895, whereas the weight mock jurors granted to the present offense positively increased after the adversary manipulations: present offense, *M*s = 21.0%–22.9%, $F(1, 189) = 12.808$, $p < .000$, power = .945.[40]

Effect of Type of Testimony and Adversary Procedures on the Basis for Mock Jurors' Ratings of Dangerousness (Influence of the Present Offense, Past Offense, and Expert Testimony)

A 2 × 4 (Type of Testimony × Adversary Manipulation) MANCOVA was performed on the mock jurors' ratings of the influence the present offense, the past offense, and the expert testimony exerted on their final decision making about dangerousness (Time 3). A multivariate main effect was found for adversary manipulation, Wilks's lambda $F(9, 453) = 12.539$, $p < .000$, but no statistically significant main effect was demonstrated for type of testimony, $F(3, 186) = 1.931$, $p = .126$) or the interaction term, Wilks's lambda $F(9, 452) = .994$, $p = .444$. See Table 6 for individual ANCOVA effects of type of testimony, adversary procedures, and their interaction on the weight mock jurors reported they placed on the present offense, the past offense, and expert testimony in mock jurors' final ratings of dangerousness.

[40]Part of this change in these dependent scores may be due to an additional rating of the second expert's influence on final dangerousness rating in the four competing expert conditions. In these four groups, mock jurors weighted four factors in their decisions (present offense, past offense, expert 1, and expert 2). No significant effects were noted in any of these measures when the four groups who rated four factors (i.e., only the other four groups that had cross-examination were used in this analysis) were removed from analyses.

Table 6
Individual Analyses of Covariance of Type of Testimony, Adversary Manipulations, and Their Interaction on the Weight Given to the Present Offense, Past Offense, and Expert Testimony in Mock Jurors' Final Decisions on Dangerousness

Between-subjects variable and basis of decision	*F*
Type of expert testimony	
Present offense	0.08
Past offense	2.06
Expert testimony	3.52*
Adversary manipulations	
Present offense	2.34*
Past offense	0.520
Expert testimony	7.15***
Interaction	
Present offense	0.86
Past offense	0.38
Expert testimony	1.73

Note. For the type of expert testimony variable, *dfs* = 1, 188; for the adversary manipulation variable and the interaction between the two variables, *dfs* = 3, 188.
*$p < .10$. ***$p < .001$.

Discussion

Hypothesis 1: Expert Testimony Will Influence Mock Jurors' Dangerousness Ratings

In a finding similar to previous empirical research investigating the influence of psychological expert testimony on mock juror decision making (Brekke & Borgida, 1988; Cutler et al., 1989; Ewing, 1987; Fox & Walters, 1986; Schuller & Vidmar, 1992; Walker, 1990), this experiment demonstrated that psychological expert testimony on the dangerousness of a criminal defendant significantly influenced mock jurors' decisions on this issue. Following the presentation of the videotaped psychological expert testimony concluding that the defendant was dangerous, mean mock jurors' ratings of a defendant's dangerousness increased substantially. This effect occurred regardless of whether actuarial testimony or clinical opinion testimony was presented. This increase in mock jurors' dangerousness ratings clearly results from the presentation of the psychological expert testimony and offers strong support for the proposition that expert testimony, regardless of type, influences mock jurors' decisions.

Although the effects of adversary procedures (cross-examination or competing experts) appear, at first glance, to have eliminated the influence of the initial expert testimony on final mock jurors' dangerousness ratings in this experiment (mock jurors' initial dangerousness ratings of the criminal defendant were not statistically different from their final ratings of the defendant's dangerousness following the adversary manipulations), three lines of evidence point to the continued influence of psychological expert testimony on dangerousness after the adversary manipulations. First, mock jurors' self-report indicates that the importance of the initial expert testimony in their decision making was still considerable when they made their final dangerousness assessments. Second, significant between-subjects effects for the influence of initial clinical opinion expert testimony on mock jurors' dangerousness ratings were found after the presentation of the adversary manipulations. Third, initial clinical opinion expert testimony was determined to have significant within-subject effects when initial mock juror dangerousness assessments (Time 1) were compared with final dangerousness evaluations (Time 3). Taken together, these results highlight the influence of expert testimony on mock jurors' decisions throughout the entire sentencing hearing.

Hypothesis 2: Clinical Opinion Expert Testimony Will Have Greater Influence on Mock Juror Decision Making Than Actuarial Expert Testimony

In this experiment, clinical opinion expert testimony was demonstrated to be more influential on mock jurors' dangerousness decisions than actuarial expert testimony. Mock jurors were found to be more influenced by clinical opinion expert testimony in their dangerousness ratings both directly after its presentation and after adversary manipulations designed to reduce that influence. Additionally, a bias in favor of the clinical opinion expert testimony was supported by mock jurors' ratings of the two types of testimony on a number of characteristics (credibility, influence, level of science, and persuasiveness). The mock jurors rated the clinical opinion expert testimony as more influential and persuasive and of equivalent level of science and credibility to the actuarial expert testimony. This supports the view that the mock jurors found the clinical opinion expert testimony more generally influential than the actuarial testimony and that juror differences in perception of the two types of expert testimony may have led them to grant greater weight to the clinical opinion testimony when they made their dangerousness evaluations.[41]

The findings of this study of a positive bias toward clinical opinion expert testimony are not inconsistent with the dominant psychological theories of persuasion. Both the heuristic processing and the story models of persuasion offer possible explanations for mock jurors' initial (i.e., before the adversary procedures) preference for clinical opinion expert testimony. The heuristic processing model of persuasion suggests that the clinical opinion testimony would be more influential because the actuarial expert testimony's use of statistics and statistical measures caused the actuarial testimony to be perceived as more complex and complicated than the clinical opinion testimony. This complexity may have, in turn, caused the actuarial information to be processed peripherally by mock jurors rather than centrally. The existing literature on peripheral processing suggests that

[41]It is interesting that mock jurors may not have been consciously aware of their bias in favor of clinical opinion expert testimony. Mock jurors' ratings of the weight they granted to the two different types of expert testimony demonstrated no statistically significant differences between the two types of testimony.

when individuals peripherally process information, they evaluate and weight it by less effortful means (i.e., heuristics) than would be the case if they analyzed its content. Research has demonstrated, in such instances, that individuals' preexisting beliefs about the world have more influence (O'Keefe, 1990).

In this experiment, several preexisting biases may have served as heuristics that mitigated the impact of the actuarial expert testimony. These include (a) a preexisting belief that actuarial instruments are incapable of explaining the complexities of future dangerousness behavior (several participants in the experiment voiced this belief during the debriefing phase of the experiment); (b) a preexisting belief that clinicians are much more accurate than they have been empirically demonstrated to be in predicting future dangerousness (Morier, 1987); or (c) a preexisting bias in favor of clinical information over more statistical information (O'Keefe, 1990). If these heuristics were used to weight the actuarial expert testimony in decision making, mock jurors may have ignored the content of the actuarial expert testimony and instead unfairly relied on preexisting biases against such testimony. As the results of our study suggest, this may have led to more influence from the less accurate clinical opinion expert testimony and less weighting of the more scientific actuarial expert testimony in mock jurors' dangerousness decision making.

The story model offers a similar rationale as to why mock jurors might be initially more influenced by clinical opinion expert testimony. As previously indicated, the story model proposes that jurors construct stories about cases based on the fit between the expert testimony, the juror's preexisting views, and the juror's final story and that this determines the weight that should be granted to proffered expert testimony in the juror's decision making. With regard to this model, the same preexisting beliefs (i.e., beliefs that instruments are incapable of predicting dangerousness, that clinicians are fairly accurate in predicting dangerousness, and a bias toward clinical information over statistical) made it more likely that clinical opinion expert testimony would more adequately fit with the mock jurors' preexisting beliefs and their stories concerning the case.

The story model and other psychological research in this area also suggests an alternative rationale for juror bias in favor of clinical opinion expert testimony over the actuarial expert testimony. The very nature of clinical opinion expert testimony may link the testimony being offered directly to the defendant in the case, whereas actuarial expert testimony based on a large sample of individuals may be more difficult for jurors to perceive as directly relevant to a specific defendant and the facts in the particular case before them. As a consequence, the clinical opinion expert testimony may be easier than the actuarial testimony for the jurors to fit into their story of the case and, therefore, may end up being more influential in their decision making. As mentioned previously, other researchers have found that expert testimony not linked to the facts of the case is less influential than expert testimony that has been more appropriately related to the case before the juror in experimental paradigms (Brekke & Borgida, 1988).

Although both of these models offer potential explanations for this study's empirical findings of mock jurors' dangerousness ratings following the initial presentation of expert testimony, neither model proposes a compelling reason as to why the clinical opinion expert testimony should have maintained its favorable bias in mock jurors' dangerousness assessments after the adversary procedures.

This experiment was not designed to produce results that would specifically test the assumptions of these theoretical models.[42] Clearly, future research is necessary to determine if the proposed mechanisms to account for the differential influence of actuarial and clinical testimony are the ones actually occurring in this experimental paradigm.

Hypothesis 3: Adversary Procedures (Cross-Examination or a Competing Expert) Will Negatively Influence Mock Jurors' Dangerousness Ratings and Evaluation of the First Expert

The presentation of cross-examination or a competing expert had a clear effect on mock jurors' final ratings of dangerousness. Overall, adversarial procedures lowered mean dangerousness ratings significantly. In fact, following the presentation of the adversary manipulations, mock jurors' dangerousness ratings did not substantially differ from their original pre-expert dangerousness ratings (Time 1).[43] In effect, adversary procedures clearly reduced the influence of the expert testimony on mock jurors' final dangerousness ratings to levels that existed before expert testimony was introduced.

Adversary procedures also had a substantial impact on mock jurors' final ratings of the first expert and his testimony (Time 3). After the presentation of cross-examination or a competing expert, mock jurors evaluated the first expert as less credible, less influential, less persuasive, and less scientific than they had evaluated him before the adversary manipulations (Time 2). Taken together, these findings suggest that adversary procedures had significant influence on both the weight mock jurors granted to psychological expert testimony and their final perception of expert psychological testimony. Nonetheless, although adversary procedures had an impact on both types of expert testimony, their influence was significantly less on the clinical opinion expert testimony. Adversary procedures failed to return mock jurors who received clinical opinion expert testimony to their initial dangerousness rating levels.

Hypothesis 4: The Type of Adversary Procedures Will Have Differential Negative Effects on Mock Jurors' Final Dangerousness Ratings and Evaluations of the First Expert

Limited empirical support was garnered for the proposition that mock jurors will differentially discount expert testimony on the basis of the kind of adversary procedure presented. In this experimental design, four levels of adversary procedures were proffered (ineffective cross-examination, effective cross-examination, a competing expert who disagreed with the first expert, and a competing expert who disagreed and attacked the content of the first expert's testimony), with the

---

[42]In fact, experimental attempts were made to limit the effects of these biases. For example, the actuarial expert testimony was scripted so that very little statistical or complex information would be presented to the mock jurors.

[43]They did, however, differ for the mock jurors who were initially presented with clinical opinion expert testimony, regardless of the type of adversary procedures presented. This suggests that clinical opinion expert testimony may be less susceptible to adversary procedures than other forms of expert testimony.

weakest negative effect predicted for ineffective cross-examination and the strongest negative effect predicted for the competing expert who both disagreed and discounted the first expert's testimony. On mock jurors' final ratings of dangerousness (Time 3), an effect for the type of the adversary procedure was demonstrated, but this effect disappeared when the influence of ineffective cross-examination was statistically controlled. This finding suggests that mock jurors differentiated between ineffective cross-examination and other adversary procedures and were less negatively influenced by the ineffective cross-examination condition. The mock jurors were equally negatively influenced, however, by the other three conditions in their final dangerousness ratings.

Mock jurors' final evaluations (Time 3) of the credibility, influence, persuasiveness, and level of science of the first expert also do not support the notion that mock jurors were differentially influenced by the type of adversary procedures. Although the ineffective cross-examination conditions tended to cause the smallest drop in mock jurors' final ratings of the characteristics of the first expert (the predicted relationship), the effective cross-examination conditions' negative impact on mock jurors' evaluations of the first expert with regard to credibility, influence, persuasiveness, and level of science was stronger than either of the two competing expert conditions (the opposite effect was predicted). Thus, the hypothesis that mock jurors would be least influenced by ineffective cross-examination and most influenced by a competing expert who disagreed and discounted the first expert's testimony was only partially supported by this study.

The existing psychological literatures on jury decision making and the heuristic processing model of persuasion both offer insight as to why the predicted effect did not occur. Empirical studies of juror skepticism and sensitivity have demonstrated that jurors sometimes pay attention to the content of expert testimony and sometimes weight their decisions solely on the basis of the presence of an expert regardless of content of the expert testimony (Cutler et al., 1988, 1989).[44] The heuristic processing model of persuasion suggests that this effect is most likely to occur when information is peripherally processed. In this study, the complexity of expert dangerousness testimony or the difficulty of evaluating and weighing the strength of two competing experts' inconsistent arguments may have caused the mock jurors to peripherally process this information. As a result of this peripheral processing, mock jurors ignored the content or quality of the competing expert's disagreement with the first expert but rather simply weighted the testimony of both experts equally (in the two-expert condition). This might potentially account for the return to baseline ratings levels (i.e., Time 1 levels) of mock jurors' final dangerousness ratings (Time 3) following the presentation of either competing expert condition. Under this theory, the testimony of the two competing experts was simply viewed as equivalent by the mock jurors based on the expert's credentials, and the mock jurors weighted the two competing expert's testimony equally (i.e., one expert canceled the other expert) in their final decisions.

---

[44]One might view juror skepticism as simply involving situations in which peripheral processing occurs and juror sensitivity as involving situations in which more central, content-based processing occurs.

Peripheral processing on the part of the mock jurors might also explain why effective cross-examination had greater negative effects on their evaluation of the first expert's credibility, persuasiveness, influence, and level of science than a competing expert condition that contained all of the same elements of the effective cross-examination and was delivered by an expert in the field. As mentioned above, the difficulty mock jurors had in balancing two experts' inconsistent opinions may have caused them to peripherally process the testimony and therefore ignore the content of the testimony in the competing expert conditions. In the cross-examination conditions, however, the mock jurors may have encountered less difficulty in evaluating the complexity of the expert testimony presented (they had to evaluate and weight only one expert) and thus were able to centrally process the content of the cross-examination. This central processing of the cross-examination would allow the mock jurors to focus on the content and quality of cross-examination and discount the first expert on the basis of the quality of the cross-examination. Again, it should be noted that this experiment did not intend to directly examine the effects of peripheral and central processing of expert testimony and, in fact, tried to minimize the impact of this problem. As a consequence, this post hoc explanation for empirical results should be taken as conjecture. Future research in this area should attempt to examine at what point and under what conditions certain types of testimony become more centrally or peripherally processed by jurors.

Limitations of Research

Before discussing the broader policy implications of the results of this investigation, it is important to note some of the limitations of this study. First, college students rather than actual jury members participated in this experiment. Actual jurors may have reacted differently to the experimental conditions. One ought to view this investigation as Phase I research (Diamond, 1997; however, see Bornstein, 1999). It should be replicated and further refined with a more representative population before any policy decisions are made on the basis of its results. In a similar view, it should be remembered that although considerable effort was expended to mimic a capital sentencing hearing, attempted control of the experimental variables and possible confounds may have caused experimental conditions to fall short of verisimilitude on a number of dimensions. These include: (a) the selection of jury members; (b) the absence of the guilt phase of the trial before the sentencing hearing; (c) the brevity of the sentencing hearing itself (it lasted approximately an hour and did not include testimony other than the expert, cross-examination, or a competing expert); (d) the simplicity of the capital sentencing hearing instruction (the instructions were rewritten and focused only on the dangerousness component of the Texas penalty instruction to avoid the confusion that is often found in understanding sentencing instructions); (e) the lack of a deliberation condition; (f) the use of one particular capital sentencing case; (g) the use of scripted testimony; (h) the absence of both cross-examination and competing expert testimony, which occur together in most trials; and most important, (i) the lack of consequence (i.e., no one dies). In other words, perhaps the greatest weakness of the study is that it simulates the process by which life or death capital sentencing decisions are made, using an unrepresentative population in an experimental situation that clearly did not convey the gravity and finality of reality.

Future Directions and Conclusions

Even after taking into account the weakness in this research design, this study's empirical findings do have several important legal and policy implications that require further empirical exploration. Most important, the research findings strongly point to the possibility that jurors weigh clinical opinion expert testimony more heavily than actuarial expert testimony and that adversarial procedures may be insufficient to remove this bias. If this is indeed the case, the Supreme Court may have taken an incorrect view concerning the constitutionality of dangerousness predictions in *Barefoot* when they stated that, "we are not persuaded . . . that the fact finder and the adversary system will not be competent to uncover, recognize, and take due account of its shortcomings" (*Barefoot v. Estelle,* 1983). At least in the one capital sentencing case explored in this experiment, mock jurors failed to fully uncover the weaknesses of the less accurate clinical opinion expert testimony that was presented. Although mock jurors in the clinical opinion expert testimony were influenced by adversary procedures, these jurors were less influenced by these procedures than mock jurors who received actuarial expert testimony. Unlike the mock jurors who received actuarial expert testimony, mock jurors receiving clinical opinion expert testimony did not return to their initial ratings of dangerousness levels following adversary procedures.

The persuasive advantage that clinical opinion expert testimony holds over actuarial expert testimony in influencing juror decision making has implications that extend far beyond capital sentencing in Texas. Courts in Florida and California have assumed that jurors routinely differentiate between clinical and other forms of expert testimony and have assumed that jurors consider clinical opinion expert testimony to be less influential than other testimony (*State v. Flanagan,* 1993; *People v. McDonald,* 1984). This research suggests that the opposite may, in fact, be true. Jurors not only do not appear to consciously discriminate clinical opinion expert testimony from other forms of expert testimony but they may also weigh less accurate clinical expert testimony more heavily than more accurate actuarial expert testimony in their final decisions. In these jurisdictions, such a bias could lead to juror decisions, in a variety of different contexts, that are unfairly based on jurors' predilection for less accurate clinical opinion testimony.

These research findings may also have bearing on the appropriate legal admissibility standard for clinical opinion expert testimony. If clinical opinion expert testimony is more heavily weighted by jurors than other testimony, the *Daubert* standard may inadvertently make it easier for the clinical opinion expert testimony to be admitted. In this case, clinical opinion expert testimony would not face the same admissibility requirements as more scientific and accurate expert testimony, and might not be perceived as any less scientific than other expert testimony, but may well be more heavily weighted by jurors in their decision making.

Future empirical research will be especially important in determining how adversary procedures could eliminate or minimize the unfair advantage that accrues to clinical opinion expert testimony. Alternatively, factors other than adversary procedures (e.g., a judicial instruction to jurors suggesting the inaccuracy of clinical predictions or application of *Daubert* criteria to the admissibility of all expert testimony) should be investigated to limit the impact of inaccurate clinical opinion testimony. Finally, future research should also explore ways in which actuarial based testimony might be presented to laypersons so that it is more appropriately weighed by jurors.

References

Arkes, H., & Hammond, K. (1986). *Judgement and decision making: An interdisciplinary reader.* New York: Cambridge University Press.

Barefoot v. Estelle, 463 U.S. 880 (1983).

Bornstein, B. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23,* 75–92.

Borum, R. (1996). Improving the clinical practice of violence risk assessment. *American Psychologist, 51,* 945–956.

Bothwell, R., Deffenbacher, K., & Brigham, J. (1987). Correlations of eyewitness accuracy and confidence: Optimality hypothesis revised. *Journal of Applied Psychology, 72,* 691–695.

Brekke, N., & Borgida, E. (1988). Expert psychological testimony in rape trials: A social–cognitive analysis. *Journal of Personality and Social Psychology, 55,* 372–386.

Brekke, N., Enko, P., Clavet, G., & Seelau, E. (1991). Of juries and court appointed experts: The impact of nonadversarial versus adversarial expert testimony. *Law and Human Behavior, 15,* 451–475.

Brief for the American Psychiatric Association as amicus curiae in support of petitioner, in Barefoot v. Estelle, 463 U.S. 880 (1983).

Carmichael v. Samyang 131 F.3d 1433 (11th Cir. 1997).

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario Symposium* (Vol. 5, pp. 3–39). Hillsdale, NJ: Erlbaum.

Champagne, A., Shuman, D., & Whitaker, E. (1992). Expert witnesses in the courts: An empirical examination. *Judicature, 75,* 5–10.

Constanzo, S., & Constanzo, M. (1994). Life or death decisions: An analysis of capital jury decisions under the special issues sentencing framework. *Law and Human Behavior, 18,* 151–170.

Cooper, J., Bennett, E., & Sukel, H. (1996). Complex scientific testimony: How do jurors make decisions? *Law and Human Behavior, 20,* 379–395.

Crowley, M., O'Callaghan, M., & Ball, P. (1994). The juridical impact of psychological expert testimony in a simulated child sexual abuse case. *Law and Human Behavior, 18,* 89–103.

Cutler, B., Penrod, S., & Dexter, H. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13,* 311–332.

Cutler, B., Penrod, S., & Stuve, T. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12,* 41–55.

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

Diamond, S. (1997). Illumination and shadow from jury simulations. *Law and Human Behavior, 21,* 561–571.

Einhorn, H. (1974). Expert judgement: Some necessary conditions and an example. *Journal of Applied Psychology, 59,* 562–571.

Ennis, G., & Litwack, R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review, 62,* 693–718.

Estelle v. Smith, 451 U.S. 454 (1981).
Ewing, P. (1987). *Battered women who kill: Psychological self-defense as a legal justification.* Lexington, MA: Heath.
Faigman, D., & Baglioni, A. (1988). Bayes theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior, 12,* 1–17.
Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241,* 31–35.
Federal Rules of Evidence 401–403, 404(b), and 702–706.
Finkel, N., Meister, K., & Lightfoot, D. (1991). The self-defense and community sentiment. *Law and Human Behavior, 15,* 585–602.
Follingstad, D., Polek, D., Hause, E., Deaton, L., Bugler, M., & Conway, Z. (1989). Factors predicting verdicts in cases where battered women kill their husbands. *Law and Human Behavior, 13,* 253–269.
Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10,* 215–228.
Frye v. United States, 292 F. 1013 (D.C. Cir. 1923).
Gaboras, N., Spanos, N., & Joab, A. (1993). The effects of complaint age and expert psychological testimony in a simulated child sexual abuse trial. *Law and Human Behavior, 17,* 103–119.
Gaeth, G., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experienced decision makers. *Organizational Behavior and Human Performance, 33,* 263–282.
Gardner, W., Lidz, C., Mulvey, E., & Shaw, E. (1996). Clinical versus actuarial predictions of violence in patients with mental illness. *Journal of Consulting and Clinical Psychology, 64,* 602–609.
General Electric v. Joiner, 522 U.S. 136 (1998).
Giannelli, P. (1980). The admissibility of novel scientific evidence: Frye v. United States, a half-century later. *Columbia Law Review, 80,* 1197–1250.
Goldberg, L. (1968). Simple models or simple process? Some research on clinical judgement. *American Psychologist, 23,* 483–496.
Golding, S. L. (1992). Increasing the reliability, validity, and relevance of psychological expert evidence. *Law and Human Behavior, 16,* 253–256.
Goodman-Delahunty, R. (1997). Forensic psychological expertise in the wake of *Daubert. Law and Human Behavior, 21,* 121–140.
Greenberg, J., & Wursten, A. (1988). The psychologist and the psychiatrist as expert witnesses: Perceived credibility and influence. *Professional Psychology: Research and Practice, 19,* 373–378.
Grisso, T., & Appelbaum, P. (1992). Structuring the debate about ethical predictions of future violence. *Law and Human Behavior, 17,* 482–485.
Grove, W., & Meehl, P. (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.
Hart, S., Webster, C., & Menzies, R. (1993). A note on portraying the accuracy of violence predictions. *Law and Human Behavior, 17,* 695–700.
Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.
Hosch, H., Beck, E., & McIntyre, P. (1980). Influence of expert testimony regarding eyewitness accuracy on jury decisions. *Law and Human Behavior, 4,* 287–296.
Imwinkelreid, E. (1994). The next step after *Daubert:* Developing a similarity epistemological approach to ensuring reliability of nonscientific expert testimony. *Cardozo Law Review, 15,* 2271–2294.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kansas v. Hendricks, 521 S. Ct. 2072 (1997).

Kelly, R. (1992). Applicability of the rules of evidence to the capital sentencing proceeding: Theoretical and practical support for open admissibility of mitigating information. *University of Missouri—Kansas City Law Review, 60,* 411–465.

Klassen, D., & O'Connor, W. (1988). A prospective study of predictors of violence in adult male mental patients. *International Journal of Law and Psychiatry, 12,* 143–158.

Kovera, M., Gresham, A., Borgida, E., Gray, E., & Regan, P. (1997). Does expert psychological testimony inform or influence juror decision making? A social cognitive analysis. *Journal of Applied Psychology, 82,* 178–191.

Kumho v. Carmichael, 1999 WL 152275.

Lavin, M., & Sales, B. (1998). Moral justifications for limits of expert testimony. In S. J. Ceci & H. Hembrooke (Eds.), *Expert witnesses in child abuse cases: What can and should be said in court* (pp. 59–81). Washington, DC: American Psychological Association.

Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of the American Medical Association, 269,* 1007–1011.

Loftus, E. (1980). Impact of expert psychological testimony on the unreliability of eyewitness identifications. *Journal of Applied Psychology, 65,* 9–15.

Loftus, E. (1986). Ten years in the life of an expert witness. *Law and Human Behavior, 10,* 241–263.

Marquart, J., Ekland-Olson, S., & Sorensen, J. (1989). Gazing into the crystal ball: Can jurors actually predict dangerousness in capital cases? *Law & Society Review, 23,* 449–468.

Melton, G., Petrila, J., Poythress, N., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers.* New York: Guilford Press.

McNeil, D. E., Sandberg, D. A., & Binder, R. L. (1998). The relationship between confidence and accuracy in clinical predictions of psychiatric patients' potential for violence. *Law and Human Behavior, 25,* 655–671.

Menzies, R., & Webster, D. (1995). Construction and validation of risk assessments in a six-year follow-up of forensic patients: A tridimensional analysis. *Journal of Consulting and Clinical Psychology, 63,* 766–778.

Monahan, J. (1981). *The clinical prediction of violent behavior.* Beverly Hills, CA: Sage.

Monahan, J. (1984). The prediction of violence: Toward a second generation of research. *American Journal of Psychiatry, 141,* 10–15.

Monahan, J., & Steadman, H. (1994). *Violence and mental disorder: Developments in risk assessment.* Chicago: University of Chicago Press.

Moore v. Ashland Chemicals, Inc., 151 F.3d 269 (5th Cir. 1998).

Moore v. State, 542 S.W.2d 664 (Tex. Crim. App. 1976).

Morier, D. (1987). *The role of expert psychiatric predictions of dangerousness in a capital murder trial: A multimethod investigation.* Unpublished dissertation, University of Minnesota.

Morse, S. (1978). Analysis of mental health law. *California Law Review, 51,* 590–627.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783–792.

Mulvey, E., & Lidz, C. (1995). Conditional prediction: A model for research on dangerousness to others in a new era. *International Journal of Law and Psychiatry, 18,* 129–143.

Nenno v. State, 970 S.W.2d 549 (Tex. Crim. App. 1998).

O'Connor v. Donaldson, 422 U.S. 563 (1975).
O'Keefe, D. (1990). *Persuasion theory and research.* London: Sage.
Otto, R. (1992). Predictions of dangerous behavior: A review and analysis of 'second generation' research. *Forensic Reports, 5,* 103–133.
Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.
Penry v. Lynaugh, 109 S. Ct. 2934 (1989).
People v. McDonald, 37 Cal.3d 351 (Cal. 1984).
Petty, R., & Caccioppo, J. (1986). *Communication and persuasion: Central and peripheral routes in attitude change.* New York: Springer-Verlag.
Rice, M. (1997). Violent offender research implications for the criminal justice system. *American Psychologist, 52,* 414–423.
Rice, M., & Harris, G. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63,* 737–748.
Rogers, R., Bagby, R., & Chow, M. (1992). Psychiatrists and the parameter of expert testimony. *International Journal of Law and Psychiatry, 15,* 387–396.
Rogers, R., Bagby, R., Crouch, M., & Cutler, B. (1990). Effects of ultimate opinions on juror perceptions of insanity. *International Journal of Law and Psychiatry, 13,* 225–232.
Saks, M., & Kidd, R. (1980). Human information processing adjudication: Trial by heuristics. *Law and Society Review, 15,* 123–160.
Schuller, R. (1992). The impact of battered women syndrome evidence on jury decision processes. *Law and Human Behavior, 16,* 597–620.
Schuller, R., & Vidmar, N. (1992). Battered wife syndrome evidence in the courtroom: A review of the literature. *Law and Human Behavior, 16,* 273–291.
Showalter, C. (1990). Psychiatric participation in capital sentencing procedures: Ethical considerations. *International Journal of Law and Psychiatry, 13,* 261–280.
Shuman, D. (1994). *Psychiatric and psychological evidence* (2nd ed.). Colorado Springs, CO: Shephard/McGraw-Hill.
Shuman, D. W., & Sales, B. D. (1998). The admissibility of expert testimony based upon clinical judgment and scientific research. *Psychology, Public Policy and Law, 4,* 1226–1252.
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.
Smith, B., Penrod, S., Otto, A., & Park, R. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior, 20,* 49–82.
Smith, D., & Frank, D. (1997). Eliminating overconfidence in psychodiagnosis: Strategies for training and practice. *Clinical Psychology, 4,* 335–345.
State v. Flanagan, 625 So.2d 827 (Fla. 1993).
Strong, J. (1995). *Evidence law: Cases and materials* (5th ed.). St. Paul, MN: West Publishing.
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.
Tarasoff v. Regents of the University of California, 551 P.2d 334 (Cal. 1976).
Texas Criminal Procedure Code, Article 370.071b–f (1985).
Tribe, L. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review, 84,* 1329–1393.
Tversky, A., & Kahneman, D. (1974). Judgements under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.
United States v. Salerno, 481 U.S. 739 (1987).

Walker, L. (1990). *Terrified love: Why battered women kill and how society responds.* New York: Harper & Row.

Webster, C., Eaves, D., Douglas, K., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk.* Burnaby, British Columbia, Canada: Simon Frasier University and Forensic Psychiatric Services Commission of British Columbia.

Wells, G. (1986). Expert psychological testimony: Empirical and conceptual analyses of effects. *Law and Human Behavior, 10,* 83–95.

Wells, G., Lindsay, R., & Tousignant, J. (1980). Effects of expert psychological advice on human performance in judging the validity of eyewitness testimony. *Law and Human Behavior, 4,* 275–285.

Witherspoon v. Illinois, 391 U.S. 510 (1968).

Worrell, C. (1987). Psychiatric predictions of dangerousness in capital sentencing. *Behavioral Sciences and the Law, 5,* 433–446.

Received January 15, 2000
Revision received May 6, 2000
Accepted May 26, 2000 ■