Law Hum Behav (2008) 32:363–374
DOI 10.1007/s10979-007-9113-9

ORIGINAL ARTICLE

The Effectiveness of Opposing Expert Witnesses for Educating Jurors about Unreliable Expert Evidence

Lora M. Levett · Margaret Bull Kovera

Published online: 17 October 2007

Abstract We tested whether an opposing expert is an effective method of educating jurors about scientific validity by manipulating the methodological quality of defense expert testimony and the type of opposing prosecution expert testimony (none, standard, addresses the other expert's methodology) within the context of a written trial transcript. The presence of opposing expert testimony caused jurors to be skeptical of all expert testimony rather than sensitizing them to flaws in the other expert's testimony. Jurors rendered more guilty verdicts when they heard opposing expert testimony than when opposing expert testimony was absent, regardless of whether the opposing testimony addressed the methodology of the original expert or the validity of the original expert's testimony. Thus, contrary to the assumptions in the Supreme Court's decision in *Daubert*, opposing expert testimony may not be an effective safeguard against junk science in the courtroom.

Keywords Juries · Experts · Scientific evidence

Courts have been concerned with the possible harm of admitting junk science into the courtroom since their decision in *Frye v. United States* (1923). In a series of recent rulings (*Daubert v. Merrell Dow Pharmaceuticals, Inc.* 1993; *General Electric Co. v. Joiner* 1997; *Kumho Tire Co. v. Carmichael* 1999), the Supreme Court ruled that judges were responsible for evaluating the quality of scientific evidence proffered and for admitting only evidence that they deemed to be relevant and reliable. As evidentiary gatekeepers, judges are to evaluate the methodology, the error rates, and the general acceptance of the research to be presented; however, research suggests that they might not be able to judge the validity of proffered expert evidence (Kovera and McAuliff 2000; Wingate and Thornton 2004). Thus, it is likely that some jurors will confront the task of evaluating the validity of flawed expert evidence.

In *Daubert*, the Supreme Court justices noted that even if unreliable expert evidence was admitted at trial, there are at least three procedural safeguards that will help jurors evaluate the validity of scientific evidence and weigh it appropriately: cross-examination, presentation of contrary evidence (including opposing expert testimony), and judicial instruction on the burden of proof. Expert testimony about a variety of topics (e.g., economic damages, eyewitness identification, repressed memories, child sexual abuse, child suggestibility) influences juror knowledge of the topic in question and/or juror decisions (Greene et al. 1999; Griffith et al. 1998; Kovera et al. 1994; Leippe 1995; McAuliff et al. 2006).[1] As jurors have difficulty differentiating between expert testimony based on flawed or valid research (Groscup and Penrod 2002, Kovera et al. 1999; McAuliff and Kovera, in press-b), it is important to determine the efficacy of these safeguards.

L. M. Levett (✉)
Department of Criminology, Law, and Society, University of Florida, Box 115950, Gainesville, FL 32611-5950, USA
e-mail: llevett@ufl.edu

M. B. Kovera
Department of Psychology, John Jay College, City University of New York, New York, NY, USA
e-mail: mkovera@jjay.cuny.edu

[1] Despite the general tendency of expert testimony to influence juror judgments, some studies find that the effects of expert testimony may occur only under certain circumstances (Devenport and Cutler 2004; Leippe et al. 2004), such as when the expert testimony appeared after the evidence (Leippe et al. 2004), whereas other studies find effects for expert testimony only when it appears before the remaining evidence (Brekke and Borgida 1988; Schuller and Cripps 1998).

The evidence supporting the efficacy of these suggested safeguards is equivocal. Cross-examination may not effectively assist jurors in making sound decisions concerning scientific evidence (Kovera et al. 1999). Traditional jury instructions did not sensitize jurors to reliability issues in one study; however revised instructions including the *Daubert* criteria sensitized jurors to unreliable scientific evidence (Groscup and Penrod 2002). Several studies have examined the influence of opposing experts on juror decisions (Cutler and Penrod 1995; Devenport and Cutler 2004; Greene et al. 1999; Griffith et al. 1998), demonstrating that opposing expert testimony may not effectively counter the testimony offered by the initial expert. To date no studies have varied the validity of the initial expert testimony to test whether opposing expert testimony is an effective safeguard that helps jurors distinguish between valid and flawed scientific evidence.

Can Jurors Differentiate Between Valid and Invalid Scientific Evidence?

Research in social psychology suggests that laypeople may not be successful in detecting errors in scientific research (Nisbett 1993). People do not recognize that the results obtained from smaller samples are less reliable than those obtained using larger samples (Tversky and Kahneman 1974). For example, individuals ignore sample size information when making probability judgments, seeing small and large samples as equally representative of the larger population (Kahneman and Tversky 1972; Tversky and Kahneman 1971). Other studies have demonstrated that individuals often have a flawed perception of chance and do not understand probabilistic information (e.g., Gilovich et al. 1985; Tversky and Kahneman 1971). These experiments demonstrate that individuals often are unable to evaluate statistics or methodology properly. Therefore, it is reasonable to assume that jurors may be unsuccessful in independently detecting flaws in research presented by an expert in court.

Two studies have examined experimentally the ability of eligible jurors to evaluate scientific evidence. Kovera et al. (1999) manipulated the construct validity, general acceptance, and the ecological validity of research presented by the plaintiff's expert in a sexual harassment case. Although jurors noticed the variations in the information provided about construct validity in the expert evidence, they did not use this information in their decision-making. In contrast, information about ecological validity and general acceptance did influence jurors' decisions. If the evidence was generally accepted and the methods that produced the evidence were ecologically valid, participants judged the expert testimony to be valid. Additional research suggests that jurors are also insensitive to internal validity threats such as confounds and non-blind experimenters (McAuliff and Kovera, Unpublished manuscript). In this study, only jurors who were high in the need for cognition (i.e., how much people enjoy engaging in effortful cognitive activities; Cacioppo and Petty 1982) differentiated between research that contained a proper control group and research that did not.

The Effectiveness of the Opposing Expert Safeguard against Unreliable Expert Evidence

Most research on opposing expert testimony to date has studied the influence of opposing expert testimony versus no opposing testimony on juror decisions (Griffith et al. 1998; Raitz et al. 1990). In these studies, researchers compared conditions with one expert, an opposing expert, and no expert. The findings from these studies suggest that opposing expert testimony does not effectively counter the testimony offered by the initial expert. Other studies found that although the presence of an opposing expert influenced participants' credibility ratings of the initial expert, overall trial judgments did not differ (Devenport and Cutler 2004; Greene et al. 1999).

Only two studies have manipulated both the presence of opposing expert testimony and the quality of the evidence to which the expert speaks. In the first of these studies (reported in Cutler and Penrod 1995), jurors heard a case in which an eyewitness viewed the crime under good or poor viewing conditions. Jurors either heard a defense expert, both a defense and an opposing expert, or no expert testimony. The defense expert testified about the reliability of eyewitnesses under differing viewing conditions and the prosecution expert discussed the limitations of the research presented by the defense expert. The defense expert witness sensitized the jurors to the factors that affect eyewitness reliability; however, adding the opposing expert caused jurors to become more skeptical of the eyewitness identification than jurors who heard no opposing expert, regardless of the condition under which the witness viewed the crime (Cutler and Penrod 1995). In another study, researchers varied the presence of foil and instruction bias in a lineup administration as well as the presence of expert testimony (none, defense only, defense and opposing expert) on characteristics of lineups that influence eyewitness reliability (Devenport and Cutler 2004). The defense expert did not influence juror judgments yet the opposing expert caused jurors to evaluate the defense expert's credibility more negatively. Although these studies varied the conditions under which the witness viewed the crime (i.e., varied the quality of the eyewitness evidence), they did not vary the methodological quality of the expert's research, so we still must ask the empirical question: Does opposing expert testimony help jurors evaluate the quality of expert evidence? Social psychological research provides hypotheses about how jurors may process the information contained in both direct and opposing expert testimony and how that testimony will influence their trial judgments.

Dual-process Models of Persuasion and the Evaluation of Expert Evidence

When evaluating the persuasiveness of a message, individuals who lack the motivation and/or ability to process information systematically rely on information like environmental cues or heuristics rather than the quality of the information presented (Chaiken et al. 1989; Petty and Cacioppo 1986). The Heuristic-Systematic (HSM) and the Elaboration Likelihood (ELM) Models of persuasion both suggest that decision makers use one of two processes to evaluate new information. In the first process, termed systematic processing or the central route to persuasion, information is processed deliberately and carefully (Petty and Wegener 1998). With systematic processing, people attend to variations in argument quality rather than more superficial characteristics of the message or the source. Individuals are more likely to use the central route or process systematically when they are highly motivated to do so and have the ability to understand the information. Since jurors are in a position in which they must make important decisions, they should be highly motivated to make a sound decision.

Even if we assume that jurors are highly motivated to make good decisions, they may not have the ability to do so. For example, expert testimony may be complex, so jurors may not have the ability to understand it. When their ability to understand the information presented is low, individuals are likely to engage in heuristic processing or use the peripheral route to persuasion (Petty and Wegener 1998). With this type of processing, individuals use heuristics or factors other than argument quality (e.g., source credibility, message length) in making a decision. Heuristics are cognitive shortcuts that allow the decision maker to arrive at a decision quickly and without much thought (Chaiken et al. 1989).

Several studies have applied social psychological models of persuasion to jury decision-making (Cooper et al. 1996; DeWitt et al. 1997; Kovera et al. 1999) and it is possible that this framework may also be useful for understanding the influence of opposing expert testimony on juror decisions about the scientific validity of scientific testimony. If an opposing expert educates jurors regarding the methodological quality of the original expert's testimony, these jurors may be better equipped to evaluate the validity of the evidence, and have no need to rely on heuristics such as ecological validity of the research or the expert's credibility when evaluating expert evidence. Research suggests that when jurors understood the *Daubert* criteria, they were sensitized to evidentiary reliability (Groscup and Penrod 2002). Thus, teaching jurors about the scientific concepts behind the presented expert testimony through an opposing expert may be an effective means of sensitizing jurors to variations in scientific quality.

However, it is possible that an opposing expert may serve a function other than a method of increasing jurors' abilities to process (Petty and Wegener 1998). An opposing expert could act as a heuristic cue that makes jurors skeptical of all scientific evidence, regardless of whether the expert attempts to educate the jury about scientific concepts. Recall that in the Cutler and Penrod (1995) study, an opposing expert caused jurors to be more skeptical of eyewitness identification evidence, regardless of the witnessing conditions. It is possible that the opposing expert acted as a heuristic in this case with the mere presence of an opposing expert acting as a cue for jurors to discount all eyewitness testimony. Jurors observed two experts who disagreed about eyewitness identification and were more likely to conclude that eyewitness identifications must be unreliable. It is possible that an opposing expert who attempts to educate the jury about scientific evidence may cause a similar skepticism effect in that jurors may be skeptical of the scientific evidence rather than sensitized to variations in the quality of the expert's science.

Overview

Is opposing expert testimony an effective safeguard that educates the jury about the methodology of the research presented in the initial testimony? To answer this question, we presented jurors with a written summary of a child sexual abuse trial, within which a defense expert testified about research she had conducted demonstrating the effects of suggestive interviewing techniques on child reports of sexual abuse. Within the trial, we manipulated the quality of the defense expert's testimony (valid versus lack of counterbalancing versus lack of control group) and the content and presence of the prosecution expert's testimony (addressed the defense expert's methodology versus did not address the defense expert's methodology versus no opposing testimony).[2] Participants rendered a verdict and rated their perceptions of the trial and expert testimony on a variety of dimensions.

[2] We also manipulated the credibility of the defense expert, but despite successful pilot testing with a student sample, the manipulation failed in the data collected from community members. The credibility manipulation did not interact with our other independent variables to affect our dependent measures so we removed the credibility manipulation from the analyses.

We chose to use a child sexual abuse trial with an expert testifying about child suggestibility because expert testimony on the effects of suggestibility has been admitted to court in the past (e.g., *Barlow v. State* 1998; *State of Ohio v. Gerson* 1996), and jurors seem to need the information presented by these experts, as research has demonstrated that jurors lack the knowledge that such expert testimony would provide to them (McAuliff and Kovera, in press-a). In addition, previous research has shown that this type of expert testimony has increased jurors' sensitivity to variations in factors that affect a child's susceptibility to suggestion (McAuliff et al. 2006; McAuliff et al. 2007). Thus, this type of expert testimony has demonstrated potential to change jurors' knowledge of factors that affect witness suggestibility.

We predicted that if opposing expert testimony sensitized jurors to the quality of the other expert's methodology, there would be a two-way interaction of opposing expert testimony type and the quality of the defense expert evidence. Specifically, we predicted that in the condition in which the opposing expert testimony addressed the methodology of the defense expert, jurors would be able to distinguish between valid and invalid methods (i.e., a simple main effect of validity when the opposing expert addresses the methodology of the defense expert on jurors' ratings of the science presented by the defense expert and verdict). In the conditions in which there is no opposing expert or the opposing expert does not address the methodology of the defense expert, we predicted that jurors would not distinguish between valid and invalid defense expert testimony, as suggested by previous research (Kovera et al. 1999).

A few studies reported that opposing experts produced a skepticism effect, causing jurors to devalue the evidence it addresses even if that evidence is of good quality (e.g., Cutler and Penrod 1995; Cutler et al. 1990b). As the presence of an opposing expert may act as a peripheral or heuristic cue, indicating a lack of consensus in the field, jurors who hear opposing expert testimony might not distinguish between valid and flawed expert testimony and therefore may render more guilty verdicts and be more likely to believe the defense expert is incompetent, regardless of the quality of scientific testimony provided by the defense expert than jurors who hear no opposing expert. Evidence for this skepticism effect would be demonstrated by a main effect of opposing expert presence (collapsed across opposing expert type) on juror judgments.

Method

Design

The experiment used a 3 (defense expert validity: valid versus missing control versus lack of counterbalancing) × 3 (opposing expert: none versus standard versus address methodology) between subjects factorial design.

Participants

Community members recruited through www.studyresponse.com participated in exchange for entry into a raffle for one of four \$25 gift certificates, two \$50 gift certificates, and two \$100 dollar gift certificates to an online store. Participants had previously indicated to studyresponse.com that they would like to receive solicitations to participate in online studies. An employee of studyresponse.com randomly selected a group of 1000 participants from this database to solicit. To ensure that participants had attended to the stimulus materials, we eliminated those participants who spent either too much or too little time on the materials. We had five undergraduate research assistants pilot the study; it took 30–45 m for them to complete the study, with 6–7 m spent reading the trial stimulus. Based on that information, we programed the web pages to time out if a participant took more than 60 m to complete the study and we eliminated the data of any participants who spent fewer than five minutes on the stimulus page. Finally, we programed the webpage to allow each respondent to participate only once by requiring that they provide their studyreponse.com identification numbers and then preventing anyone with those numbers from participating a second time. As the raffle was linked to participants' identification numbers, it is unlikely that anyone would attempt to fabricate a new number so that they could participate again because they would not increase their chances of winning the raffle by doing so.

The remaining 262 participants[3] were more likely to be women (59%) and Caucasian (89%). Other represented ethnicities were Black (2%), Hispanic (4%), and Asian (3%). The average age of the sample was 39 years (*SD* = 12 years). Participants had a wide range of educational experiences: some high school (1%), high school diploma (11%), some college (36%), college degree (29%), technical degree (9%), some graduate school (5%), and graduate degree (9%). About 25%of them had served on a jury before, with previous service evenly split between criminal (49%) and civil (51%) trials.

[3] Collecting data using the internet versus traditional survey methods generally results in a lower response rate; however, several studies have demonstrated that results from internet research yield very similar results to laboratory-based research (Birnbaum 2004; Krantz and Dahal 2000). In this study, participants were sent e-mail invitations and reminders to participate in the study. In one study investigating such methods, over one-third of non-respondents did not respond to the invitation because they did not read their e-mail during the data collection period (Welker 2001, reported in Birnbaum 2004).

Trial Summary

The trial summary was based on *Pavel v. Hollin* (2001), a child sexual abuse case with a 4-year-old alleged victim. The child, the child's mother, a medical expert, the interviewing therapist, the defendant (the father), the defendant's girlfriend, and a defense expert on the reliability of child witnesses testified in all versions of the trial summary. The summary included judicial instructions on the burden of proof and the elements of the crime.

The child and the mother alleged that the defendant sexually abused the child during visits with the defendant. The therapist testified about her interview with the child following the initial accusation. On cross-examination, she admitted that she often repeated questions when interviewing the child and that she suggested specific behaviors to the child during the interview process. The medical expert presented ambiguous evidence about the child's medical condition, testifying that there was a slight redness around his anus but that it could be due to diarrhea or to sodomy. The defendant testified and denied the allegations. The defendant's girlfriend testified about the visits between the defendant and the child, and vouched for the defendant's character.

The defense raised the issue that the accuracy of the child's report may have been negatively influenced by suggestive questioning. An expert testified about the influence of suggestive questioning on children's reports of sexual abuse, relying heavily on an adapted description of the 'Sam Stone' study (Leichtman and Ceci 1995). In this study, a stranger (i.e., Sam Stone), visited young children between the ages of three and six. Later, the teacher showed the children a ripped book and a soiled teddy bear and asked if Sam Stone had damaged either item (he had not). In one condition, children were then asked leading questions about Sam's visit and the book/bear over several weeks. In the second condition, they were not asked leading questions but still were questioned about the events. When interviewed at the end of the experiment, children made the most false reports about the events during Sam Stone's visit when they had been asked leading questions about those events. This study was chosen as the centerpiece of the defense expert's testimony because it is an internally valid study typical of the type of study presented as evidence of the suggestibility of children (e.g., Bruck and Ceci 1995).

Validity of Defense Expert Testimony

In the valid condition, the original methods used in the Sam Stone study were adapted to a valid within subject design. Specifically, the study contained a control group and was appropriately counterbalanced, meaning the events and objects about which leading and non-leading questions are asked were counterbalanced across children. In those conditions in which the expert's study lacked a control condition, the research was missing the critical control group (i.e., all participants in the research were questioned using only suggestive questioning techniques). Based on this research, the expert concluded that suggestive questions caused the children to be inaccurate.

In those conditions in which the expert's study contained a confound, the topics of the leading and misleading questions were not counterbalanced. Children were always asked leading questions about the same set of events or objects and non-leading questions about different sets of items. This particular confound was chosen because it has been a problem in published child suggestibility research (McAuliff et al. 1998).

Type of Opposing Expert Testimony

In all conditions that contained opposing expert testimony, the opposing expert presented arguments based on published criticisms of the relevance of most child witness suggestibility research to typical child sexual abuse cases (Lyon 1999). For example, there is no empirical evidence to support the assumption made by child witness researchers that suggestive interviewing is the norm in sexual abuse cases. Further, researchers in this area ignore the fact that false claims in sexual abuse cases are highly unlikely because of the nature of sexual abuse. In the trial stimulus, the opposing expert addressed the relevance of the defense expert's research to the current case, arguing it could not be proved that this child was asked leading questions. The expert also argued that even if the child was asked leading questions, it would not prove that the child was mistaken about being abused.

In those conditions in which the opposing expert addressed the methodological quality of the defense expert's research, the opposing expert commented on the internal validity of the research presented by the defense in addition to presenting the criticisms raised above. When the defense expert's methodology was invalid, the opposing expert explained the importance of counterbalancing or control groups during direct examination. When the defense expert's testimony was valid, the opposing expert acknowledged that it was so during cross-examination. Despite acknowledging the validity of the defense expert's testimony in those conditions in which the testimony was valid, the opposing expert continued to disagree with the original expert, stating that her study results were irrelevant to trial facts.

Dependent Measures

Verdict

Participants indicated whether they believed the defendant was guilty or not guilty of the charges against him.

Ratings of Trial Participants

Participants rated the experts on a series of bipolar adjective pairs. Participants rated witness credibility using these adjective pairs: convincing–unconvincing (R), not credible–credible, not believable–believable, competent–incompetent (R), and certain–uncertain (R) (defense expert $\alpha = .89$; prosecution expert $\alpha = .93$). They also rated the trustworthiness of each expert on the following bipolar adjective pairs: honest–dishonest (R), sincere–insincere (R), immoral–moral, and trustworthy–untrustworthy (R) (defense expert $\alpha = .90$; prosecution expert $\alpha = .90$). All ratings were on 7-point scales and were recoded when necessary so that a higher rating always indicated a more favorable rating of the witness. Items followed by (R) were reverse coded.

Ratings of the Expert Testimony

For all remaining ratings, participants indicated the level of their agreement with a series of statements, with one indicating strong disagreement and six indicating strong agreement. Items were recoded when necessary so that higher numbers represented a more positive evaluation. To determine if jurors were able to distinguish between valid and flawed scientific evidence, we directly measured jurors evaluations of the expert testimony presented. This measure was the most sensitive measure to assess whether jurors used the opposing expert safeguard to assist them in making scientifically sound decisions about the expert testimony proffered. Presumably, jurors would give more weight to valid evidence than flawed evidence, and these differing weights would figure into their final verdicts. However, it is possible that jurors' evaluations of the scientific evidence would not affect their final verdict. So, the most direct measure assessing their ability to distinguish between valid and flawed science is their evaluation of the scientific evidence. To measure participants' assessments of the quality of the expert evidence, we created a score reflecting participants' ratings of the defense expert's research by averaging their responses to the following statements ($\alpha = .92$): (a) In my opinion, the research techniques used by the defense expert were valid; (b) In my opinion, the research presented by the defense expert was not reliable (R); (c) The dependent measures used by the defense expert were very appropriate; (d) The procedure used by the defense expert was inappropriate (R); (e) In my opinion, the findings of the research presented by the defense expert were applicable in this trial; (f) The findings of the research presented by the defense expert were applicable to child sexual abuse cases in general; (g) In my opinion, the research presented by the defense expert was not "good science" (R); and (h) The methods that the defense expert used to test the effects of suggestive questioning on children's memory were sound.

We created a scale score for participants' ratings of the prosecution expert's testimony by averaging their responses to the following statements ($\alpha = .82$): (a) In my opinion, the testimony of the prosecution expert was applicable in this trial; (b) The testimony of the prosecution expert was applicable to child sexual abuse cases in general, in my opinion; and (c) In my opinion, the testimony of the prosecution expert was not "good science" (R).

Pilot Tests

We pilot tested the trial stimulus to ensure participants noticed our manipulations. After reading the trial stimulus, participants ($N = 60$ undergraduate students at a southeastern public university) answered a series of questions assessing their perception of our manipulations. Participants in the pilot study were successful in noting that the valid research contained two sets of questions whereas the research lacking the control group had only one set, $\chi^2(1, N = 40) = 4.10$, $p < .05$, $\phi = .32$. Despite several revisions, participants were unable to distinguish between valid research and research that lacked proper counterbalancing of questions (63% vs. 52% of participants reported that both suggestive and nonsuggestive questions had been asked about each event for valid and non-counterbalanced conditions, respectively), $\chi^2(1, N = 40) = 2.71$, $p > .05$, $\phi = .26$. We included the condition despite the inability of participants to notice the manipulation without opposing expert education because the purpose of the opposing expert was to assist the jurors in recognizing the flaw.

Procedure

Participants received an e-mail from studyresponse.com inviting them to participate in a study that entailed both reading a trial summary and rendering trial judgments. Participants were told that should they choose to participate, they would be entered in a raffle for prizes. If participants chose to participate, they clicked on a link in the e-mail, which brought them to the website with the consent form. Participants entered their anonymous studyresponse.com identification number as an electronic signature on the consent form. Then, the program randomly assigned them to condition and presented the trial stimulus. After they read the stimulus, participants answered questions by pointing and clicking with the mouse. After participants completed answering the questions, they were debriefed and thanked.

Results

Manipulation Checks

Participants in the study were successful at distinguishing between valid research and research that lacked a control group (84% vs. 38% reported that both suggestive and non-suggestive questions had been asked), $\chi^2(1, N = 178) = 32.27$, $p < .05$, $\phi = .39$. Participants in the study did not recognize the difference between valid testimony and testimony that lacked proper counterbalancing of questions (84% vs. 74% of participants reported that both suggestive and nonsuggestive questions had been asked about each event for valid and non-counterbalanced conditions, respectively), $\chi^2(1, N = 177) = 2.29$, $p = .13$, $\phi = .12$. Given that research has been published with this flaw, it is not surprising that participants did not notice the improper counterbalancing of questions.

Does Opposing Expert Testimony Increase Juror Sensitivity to Variations in Scientific Validity?

We conducted separate 3 (defense expert validity: valid versus missing control versus lack of counterbalancing) × 3 (opposing expert: none versus standard versus address methodology) MANOVAs for each set of continuous dependent measures (defense expert ratings and prosecution expert ratings) and conducted a logistic regression for the verdict measure.

We regressed verdict on opposing expert type, defense expert testimony validity, and the interaction of these two variables. A significant interaction term would be evidence of a sensitization effect (i.e., a significant interaction would indicate that the variations in the validity of the defense expert's evidence affected verdict in those conditions with an opposing expert who addressed the methods used by the defense expert). Neither the interaction nor either of the main effects were significant predictors of verdict, all $\chi^2(8) > 4.38$, $p > .43$.

A sensitization effect would predict a simple main effect of defense expert testimony validity in those conditions in which the opposing expert addressed the methodology of the original expert. Further, jurors would not distinguish between invalid and valid science in those conditions in which the opposing expert was absent or the opposing expert did not address the methods used by the defense expert. However, the omnibus interaction terms in the MANOVAs for each of our sets of ratings (including juror evaluations of the expert testimony, our most sensitive measure of whether jurors distinguished between valid and flawed science) were nonsignificant, all $F\text{s} < .81$, all $p\text{s} > .56$, all $\eta^2 < .02$. Before concluding that we had no evidence that opposing experts increased juror sensitivity to variations in scientific validity, we wanted to test the theory-driven sensitivity hypothesis using the most powerful tests we had at our disposal. Therefore, we conducted planned comparisons by examining the multivariate and univariate simple main effects of study validity within opposing expert type for each of the ratings.

First, we used a MANOVA to test the simple main effect of study validity within the three levels of opposing expert on the scales representing participants' evaluations of the defense expert's credibility, trustworthiness, and research quality. None of the multivariate simple main effects were significant, all $F\text{s}\ (6, 400) < 1.70$, $p > .12$, $\eta^2 < .03$. When looking at the univariate tests for defense expert credibility and trustworthiness, there was a simple main effect of study validity only in the conditions in which the opposing expert addressed the defense expert's methodology. Specifically, participants judged the defense expert to be more trustworthy and more credible in the valid condition than in the invalid due to control condition, although the latter effect was only marginally significant. See Table 1 for all means and univariate tests.

The second set of analyses examined the multivariate and univariate simple main effects of study validity on participants' ratings of the prosecution expert within opposing expert type (standard versus address methodology only as participants in the no expert condition had no prosecution expert to evaluate). None of the multivariate or univariate simple effects were significant, with multivariate $F\text{s}\ (6, 266) < 1.41$, $p > .21$, $\eta^2 < .03$, and univariate $F\text{s}\ (2, 128) < 1.87$, $p > .16$, $\eta^2 < .03$.

Does Opposing Expert Testimony Make Jurors Skeptical of All Expert Testimony?

To test for a skepticism effect on verdict, we conducted a logistic regression with opposing expert (present versus absent), study validity, and the interaction of these variables as predictors. Only the opposing expert manipulation was a significant predictor of verdict, $\chi^2(1, N = 258) = 4.37$, $p < .04$. The likelihood odds ratio was 1.83, indicating that opposing expert testimony increased

Table 1 Univariate effects of study validity within type of opposing expert on defense expert ratings

Measure	Means (SE)			Univariate effect of study validity			
	Valid	Invalid(count)	Invalid(control)	*F*	df	*p*	η^2
No opposing expert							
Credibility	5.79 (.26)	5.74 (.27)	5.24 (.24)	1.48	2, 194	.23	.02
Trustworthiness	6.09 (.24)	6.20 (.27)	5.56 (.23)	2.03	2, 194	.13	.002
Research ratings	4.47 (.20)	4.42 (.22)	4.21 (.19)	.53	2, 194	.59	.01
Opposing expert—does not address methodology							
Credibility	5.07 (.27)	5.19 (.26)	5.19 (.32)	.07	2, 194	.94	.001
Trustworthiness	5.70 (.24)	5.44 (.25)	5.79 (.30)	.46	2, 194	.63	.01
Research ratings	3.82 (.20)	3.95 (.20)	4.06 (.25)	.29	2, 194	.75	.003
Opposing expert—addresses methodology							
Credibility	5.32 $(.23)_a$	4.92 (.30)	4.43 $(.29)_a$	2.83	2, 194	.06	.03
Trustworthiness	5.73 $(.22)_a$	5.54 (.29)	4.79 $(.28)_a$	3.65	2, 194	.03	.04
Research ratings	3.98 (.18)	3.90 (.23)	3.56 (.23)	1.09	2, 194	.34	.01

Note: Means sharing subscripts differ at $p < .05$

the odds of participants rendering a guilty verdict by 83% (25% vs. 37% guilty for no opposing expert and opposing expert conditions, respectively).

We used MANOVAS to test the contrast of the no opposing expert condition against the conditions that contained an opposing expert. There was a significant effect of opposing expert on participants' ratings of the defense expert, $\lambda = .94$, F (3, 200) = 4.30, $p < .01$, $\eta^2 = .06$. Univariate analyses showed that participants rated the defense expert as more credible, more trustworthy, and as having presented higher quality research when the opposing expert was absent than when the opposing expert was present. See Table 2 for means and univariate tests. The interaction between opposing expert presence and study validity was not significant.

Discussion

Does opposing expert testimony sensitize jurors to variations in scientific validity? Or does opposing expert testimony just make jurors devalue expert testimony, irrespective of its validity? We found strong evidence for a skepticism effect; whether an opposing expert was present affected jurors' verdicts and ratings of the defense expert regardless of the content of the defense expert's testimony. We found little to no support for a sensitization effect, using the most powerful statistical tests available. When the main effect of study validity was significant in the address methodology condition, jurors generally rated the defense expert more favorably in those conditions with a valid study than in those conditions in which the expert's study lacked a control group. These analyses were theory driven and conducted despite a nonsignificant omnibus test to provide the strongest possible test of the sensitization effect. Despite the powerful test and the significant manipulation check indicating that most participants noticed whether a control group was present or absent, there was no evidence that opposing expert testimony helped jurors distinguish between flawed and valid scientific testimony.

This finding has several important implications. First, opposing expert testimony, at least as we operationalized it, seems to have very limited potential to educate jurors about the quality of scientific evidence. For the opposing expert safeguard to be effective, the opposing expert must address the testimony of the original expert, although even under these conditions, the opposing expert was only effective in helping jurors distinguish between flawed and valid science in their ratings of the credibility and trustworthiness of the defense expert but not for their ratings of the quality of the expert evidence itself. Second, in our study, jurors failed to distinguish between valid research and research that lacked proper counterbalancing of questions, even in those conditions in which an opposing expert addressed the validity of the defense expert's research. Others found that in attempting to educate jurors about the validity of science, jurors had trouble distinguishing between valid science and flawed science when the flaw was a more difficult methodological concept (e.g., confound, experimenter bias) than when the flaw was easier to comprehend and more often discussed in media reports about research findings (e.g., a missing control group; McAuliff and Kovera, Unpublished manuscript). Similarly, in those very few instances in which we found a sensitization effect, it was for evaluations of research that lacked a control group, not for failure to counterbalance questions (a confound). Thus, it may be reasonable to hypothesize that an opposing expert's

Table 2 Univariate effects of opposing expert presence on defense expert ratings

Measure	Means (SE)		Test of opposing expert main effect			
	No opposing expert	Opposing expert	*F*	df	*p*	η^2
Credibility	5.59 (.15)	5.02 (.11)	9.12	1, 194	.01	.05
Trustworthiness	5.95 (.14)	5.48 (.11)	6.33	1, 194	.01	.03
Research ratings	4.36 (.12)	3.87 (.09)	11.24	1, 194	.01	.06

potential to sensitize jurors to flaws in research are somewhat dependent on the difficulty of the concept. More research is needed to investigate the effect of the difficulty of statistical concept on juror decision-making.

Although evidence of a sensitization effect was weak to non-existent, there was much stronger evidence of a skepticism effect. There were more convictions and the defense expert was viewed to be more credible, more trustworthy, and as having presented higher quality research when jurors heard an opposing expert, regardless of the type of evidence proffered by either expert. These results suggest that jurors were influenced by the presence of opposing expert testimony, but were not using the content of that testimony to make better decisions about the quality of proffered expert evidence.

Our research is not the first to demonstrate that opposing experts may produce juror skepticism about the evidence that an expert proffers. In another study, a defense expert witness testified about factors that affect eyewitness reliability, and this expert evidence sensitized jurors to those factors (Cutler and Penrod 1995). In those conditions in which an opposing expert testified, jurors became skeptical of all eyewitness identification regardless of the variation in the witnessing conditions of the eyewitness in the case. That is, in those conditions with an opposing expert, jurors thought eyewitnesses were more unreliable compared to those conditions with just a defense expert. We also found that an opposing expert caused jurors to become more skeptical of the research presented by the defense expert, regardless of the variation in its validity or the content of the opposing testimony, extending the skepticism effect of opposing expert testimony to jurors' evaluations of evidence.

Perhaps jurors used the presence of opposing expert testimony to make inferences about the quality of the message presented by the expert. Jurors may have interpreted the battling experts as evidence that the research was not generally accepted in the field. That is, the jurors may have noted that two experts from the same field with similar degrees were disagreeing on a set of findings; therefore, the findings must not be generally accepted in the field. Thus, it is possible that jurors in our study inferred a lack of general acceptance from our opposing experts and used this lack of general acceptance rather than the actual content of the testimony in making their decisions. It is also possible that jurors saw the experts with opposing viewpoints as "hired guns," or experts hired specifically to give opposing opinions (Cooper and Neuhaus 2000). Thus, they may have used the presence of the opposing expert to conclude that the research presented was inconsequential because the experts were hired to give their opinions.

Limitations and Directions for Future Research

One limitation of our study is that because we did not include a condition in which no expert testimony was presented, we were unable to draw conclusions about the relative effect of the expert testimony proffered in this study on verdict. That is, we were able to determine if an opposing expert affected jurors' abilities to distinguish between varying levels of scientific validity, but we were unable to make any direct conclusions about the effect of the expert testimony, flawed or valid, on verdict. Previous research generally shows that expert testimony affects juror decision-making or knowledge about a particular subject (Greene et al. 1999; Griffith et al. 1998; Kovera et al.1994; Leippe 1995; McAuliff et al. 2006) but more recently, research has shown that some of these effects only occur in certain circumstances (e.g., Leippe et al. 2004) or not at all (Devenport and Cutler 2004). Thus, it may be possible that the reason we saw no sensitization effect in our dependent measure of verdict is because the expert testimony had no effect on verdict in the first place. In our findings regarding the skepticism effect, it could be possible that the defense expert testimony had no effect on verdict, and the opposing expert caused skepticism independent of the defense expert's testimony. We think that these possibilities are unlikely given that ratings of the quality of the expert evidence were well above the floor, and often well above the mid-point of the scale in conditions in which there was no opposing expert testimony.

We made a decision to focus our study on whether opposing experts influence jurors' evaluations of the quality of another expert's testimony rather than verdicts. Dichotomous verdict measures are relatively less sensitive to the influence of manipulations than are continuous ratings of evidence. To provide the strongest possible test of the ability of opposing experts to sensitize jurors to variations in expert evidence quality, we chose ratings of expert evidence quality as our primary dependent measure. As a result of this choice, a no expert testimony condition became less feasible as participants could not rate the quality of expert evidence that they did not hear. However, future research should expand the evaluation of the opposing expert safeguard to test not only its ability to assist jurors in distinguishing between flawed and valid science, but also to assess the effects of opposing experts on verdicts relative to conditions in which no expert testimony is presented.

It may also be possible to improve the ability of opposing experts to educate jurors about the relative quality of scientific evidence. First, using a stronger opposing expert including a more thorough evaluation of the initial testimony and explanation of the statistical concepts in that testimony may be a more effective safeguard against junk science. This extended discussion may enable jurors to more accurately understand the statistical concept behind the defense expert's testimony, and thus enable them to use the validity of that testimony when making a decision, eliminating the possibility that jurors may not understand the more difficult statistical concepts.

Second, perhaps a court appointed expert who evaluates the research presented at trial would be a more effective safeguard against junk science than an adversarial opposing expert. The experts in this study, consistent with normative trial practice, were called by either the defense or the prosecution. Thus, before the expert testifies, the jury already views the opposing expert as an adversarial witness called to disagree with the other side (perhaps as a hired gun), and therefore the jury may be less apt to listen to the content of the opposing expert's testimony and use it as a tool for better understanding the other expert's research. Thus, the adversarial nature of the opposing expert may have served as a cue for jurors that the information under debate was not generally accepted in the field, and therefore jurors became skeptical of the evidence. Indeed, the very process of cross-examining an expert's testimony may create an appearance of controversy even when the majority of the field agrees on the findings presented (Jasanoff 1993, 1995).

Perhaps an expert called to testify on behalf of the court may be a more effective safeguard against junk science than an expert called to testify for one of the parties. Currently, very few experts are court appointed, but research has demonstrated that this number is on the rise (Cecil and Willging 1993; Hooper et al. 2001). It is possible that the court appointed expert may overwhelm other expert testimony regardless of the content of the testimony, merely because the court appointed expert may be seen by the jury as a neutral party in an adversarial case (Cecil and Willging 1993; Hooper et al. 2001). In one study, however, a court appointed expert on eyewitness issues only caused skepticism in juror decisions (Cutler et al. 1990a). Future research testing the effectiveness of a stronger opposing expert and the court appointed expert is needed before we can strongly conclude that the opposing expert safeguard is ineffective in combating junk science in the courtroom.

Testing the safeguards established by the court is not only informative, but also necessary, as we have learned through research, that flawed expert testimony may enter the court system (Kovera and McAuliff 2000). Furthermore, jurors' ability to reason about scientific evidence is similar to that of a layperson, and therefore jurors may be unable to recognize methodological errors in research (Kovera et al. 1999; Lehman et al. 1988). It is of utmost importance to examine strategies that may increase juror ability to reason about scientific evidence. The results of this study demonstrate that opposing expert testimony may not be an effective safeguard against junk science, especially not in its current adversarial form. Thus, until we develop other methods of educating jurors about scientific validity issues at trial, revised judicial instruction may be our only available method of educating jurors about junk science that may be admitted at trial. However, more research is needed to develop a more effective opposing expert safeguard and to further explore the limitations and effectiveness of revised judicial instruction.

Acknowledgments This research was conducted in partial fulfillment of the requirements for the first author's master's degree at Florida International University, under the supervision of the second author, and was supported in part by a grant-in-aid from the American Psychology-Law Society. Portions of this research were presented at the 2004 joint meeting of the American Psychology-Law Society and the European Association of Psychology and Law, Edinburgh, Scotland. We are grateful for the helpful suggestions of committee members Aya Gruber and Janat Parker.

References

Barlow v. State, (1998). 270 Ga 54.

Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55*, 803–832.

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

Bruck, M., & Ceci, S. J. (1995). Amicus brief for the case of State of New Jersey v. Michaels presented by committee of concerned social scientists. *Psychology, Public Policy, and Law, 1*(2), 272–322.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cecil, J. S., & Willging, T. E. (1993). *Court appointed experts: Defining the role of experts appointed under Federal Rule of Evidence 706*. Washington, DC: Federal Judicial Center.

Chaiken, S., Liberman, A., & Eagly, A. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–251). New York: Guilford Press.

Cooper, J., Bennett, E. A., & Sukel, H. L. (1996). Complex scientific testimony: How do jurors make decisions? *Law and Human Behavior, 20*(4), 379–394.

Cooper, J., & Neuhaus I. (2000). The hired gun effect: Use of pay, credentials, and frequency of testifying as peripheral decisions. *Law and Human Behavior, 24*, 149–171.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1990a). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Psychology, 20*, 1197–1207.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990b). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*(2), 185–191.

Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993). 113 S. Ct. 2786.

Devenport, J. L., & Cutler, B. L. (2004). Impact of defense-only and opposing eyewitness experts on juror judgments. *Law and Human Behavior, 28*, 569–576.

DeWitt, J. S., Richardson, J. T., & Warner, L. G. (1997). Novel scientific evidence and controversial cases: A social psychological examination. *Law and Psychology Review, 21*, 1–23.

Frye v. United States, (1923). 54 App.D. C. 46, 293 F.1013.

General Electric Co. v. Joiner, (1997). 522 U. S. 136.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

Greene, E., Downey, C., & Goodman-Delahunty, J. (1999). Juror decisions about damages in employment discrimination cases. *Behavioral Sciences and the Law, 17*(1), 107–121.

Griffith, J. D., Libkuman, T. M., & Poole, D. A. (1998). Repressed memories: The effects of expert testimony on mock jurors' decision making. *American Journal of Forensic Psychology, 16*(1), 5–23.

Groscup, J., & Penrod, S. D. (2002). *Jury instructions effects on juror assessments of scientific reliability.* Paper presented at the biennial meeting of the American Psychology-Law Society, San Antonio, TX.

Hooper, L. L., Cecil, J. S., & Willging, T. E. (2001). Assessing causation in breast implant litigation: The role of science panels. *Law and Contemporary Problems, 64*, 139–189.

Jasanoff, S. (1993). What judges should know about the sociology of science. *Judicature, 77*(2), 77–82.

Jasanoff, S. (1995). *Science at the bar: Law, science, and technology in America*. Cambridge, MA: Harvard University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kovera, M. B., Levy, R. J., Borgida, E., & Penrod, S. D. (1994). Expert testimony in child sexual abuse cases: Effects of expert evidence type and cross examination. *Law and Human Behavior, 18*, 653–674.

Kovera, M. B., & McAuliff, B. D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: Are judges effective gatekeepers? *Journal of Applied Psychology, 85*(4), 574–586.

Kovera, M. B., McAuliff, B. D., & Hebert, K. S. (1999). Reasoning about scientific evidence: Effects of juror gender and evidence quality on juror decisions in a hostile work environment case. *Journal of Applied Psychology, 84*(3), 362–375.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 35–60). San Diego: Academic.

Kumho Tire Co. v. Carmichael, (1999). 526 U.S. 137.

Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology, 31*, 568–578.

Leippe, M. R. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy and Law, 1*, 909–959.

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Seib, H. M. (2004). Timing of eyewitness expert testimony, jurors' need for cognition, and case strength as determinants of trial verdict. *Journal of Applied Psychology, 89*, 524–541.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43*(6), 431–442.

Lyon, T. D. (1999). The new wave in children's suggestibility research: A critique. *Cornell Law Review, 84*, 1004–1087.

McAuliff, B. D., & Kovera, M. B. (in press-a). Estimating the effects of misleading information on witness accuracy: Can experts tell jurors something they don't already know? *Applied Cognitive Psychology*.

McAuliff, B. D., & Kovera, M. B. (in press-b). Juror sensitivity to methodological flaws in psychological science. *Journal of Applied Social Psychology*.

McAuliff, B. D., Kovera, M. B., & Viswesvaran, C. (1998). Methodological issues in child suggestibility research. In B. L. Bottoms & J. A. Quas (Co-Chairs), *Situational and individual sources of variability in children's suggestibility and false memories*. Symposium presented at the meeting of the American Psychology-Law Society, Redondo Beach, CA.

McAuliff, B. D., Nicholson, E., & Ravanshenas, D. (2007). Hypothetically speaking, can expert testimony improve jurors' understanding of developmental differences in suggestibility? In B. D. McAuliff & D. Ravanshenas (Co-Chairs), *Giving away developmental psychology in court: Can expert testimony on suggestibility and child abuse improve jurors' decisions?* Symposium presented at the meeting of the Society for Research in Child Development, Boston, MA.

McAuliff, B. D., Samis, A. Vargas, J., & Gonzalez, K. (2006). May we suggest a hypothetical? Expert testimony in witness suggestibility cases. In L. Levett (Chair), *Juror decision making about expert evidence.* Symposium presented at the meeting of the American Psychology-Law Society, St. Petersburg, FL.

Nisbett, R. E. (1993). *Rules for reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Pavel v. Hollins, (2001). 261 F. 3d 210.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–203). New York: Academic Press.

Petty, R. E., & Wegener, D. T. (1998) Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of social psychology* (pp. 323–390). Boston: McGraw-Hill.

Raitz, A., Greene, E., Goodman, J., & Loftus, E. F. (1990). Determining damages: The influence of expert testimony on jurors' decision making. *Law and Human Behavior, 14*(4), 385–395.

Schuller, R. A., & Cripps, J. (1998). Expert evidence pertaining to battered women: The impact of gender of expert and timing of testimony. *Law and Human Behavior, 22*, 17–31.

State of Ohio v. Gersin, (1996). 76 Ohio St.3d 491.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Welker, M. (2001). Email surveys: Non-response figures reflected. In U. D. Reips, & M. Bosnjak (Eds.), *Dimensions of internet science* (pp. 231–238). Lengerich, Germany: Pabst Science Publishers.

Wingate, P. H., & Thornton, G. C. III (2004). Industrial/organizational psychology and the federal judiciary: Expert witness testimony and the *Daubert* standards. *Law and Human Behavior, 28*, 97–114.