Psychology, Public Policy, and Law
2009, Vol. 15, No. 1, 1–18

1076-8971/09/$12.00 DOI: 10.1037/a0015290

THE GATEKEEPER EFFECT
The Impact of Judges' Admissibility Decisions on the Persuasiveness of Expert Testimony

N. J. Schweitzer and Michael J. Saks
Arizona State University

In a pair of mock-trial studies of a possible "gatekeeper" effect, our participants were presented with a summary of a trial that included a piece of expert scientific evidence. The judge's decision was manipulated to admit the scientific evidence, as well as the quality of the evidence and the credibility of the expert. Participants were found to be less critical of and more persuaded by expert evidence when it was presented within a trial, compared with the same evidence presented outside of a courtroom context. These findings suggest that, when judges allow expert testimony to reach the jury although the evidence is of low quality, they imbue it with undeserved credibility. Furthermore, no changes in participants' perceptions of the evidence were found if the mock jurors were explicitly informed that the judge had evaluated the evidence, suggesting that the participants assumed that judges normally review evidence before allowing it to reach the jury. In addition, implications for basic research are discussed, as the moderating effects of a gatekeeper have not previously been considered by established models of persuasion.

Keywords: scientific evidence, experts, persuasion, judges, jury decision-making

Judges in American courts have served as gatekeepers of evidence since the adversary process became dominant in the second half of the 18th century (Landsman, 1984). After the control of evidence gathering and presentation came firmly into the hands of the parties' attorneys, detailed rules of evidence developed to regulate the offerings of these persuasive adversaries (Gallanis, 1999). The principal purpose of the rules of evidence, of course, is to facilitate the exclusion of evidence from the jury's consideration. As even the most casual observer of courtroom procedures realizes, judges are the decision-makers responsible for that filtering.

In recent years, the subject of judicial gatekeeping has been rejuvenated by a series of cases in which the United States Supreme Court has emphasized the responsibility of judges to filter proffered expert testimony. In *Daubert v. Merrell Dow Pharmaceuticals* (1993), the Court held the test of admission in the federal courts to be essentially that of science: empirical verification of

N. J. Schweitzer, New College of Interdisciplinary Arts and Sciences, Arizona State University; Michael J. Saks, Sandra Day O'Connor College of Law, Arizona State University.

We thank Darwyn Linder and Robert Cialdini for their assistance with the planning of these experiments and their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to N. J. Schweitzer, New College of Interdisciplinary Arts and Sciences, Arizona State University, P.O. Box 37100, MC3051, Phoenix, AZ 85069-7100. E-mail: njs@asu.edu

claims of expertise using sound research methods.[1] The validity-based gatekeeping of *Daubert* was enlarged in *Kumho Tire v. Carmichael* (1999), which held that all fields of purported expertise, regardless of what they were called, had to pass the most appropriate tests of validity applicable to them or be refused admission. In *General Electric v. Joiner* (1997), the Court placed gatekeeping responsibility squarely on the trial court, subject to only deferential appellate review. Finally, *Weisgram v. Marley Co.* (2000) held that trial court gatekeeping was such serious business that an appellate court could terminate a case altogether (without remanding it for further proceedings) if a party failed in its first attempt to obtain admission of expert evidence necessary to its case. This heightened attention to the gatekeeping of expert testimony has, as of this writing, produced 559 law review articles with *Daubert* in their titles, an additional 5,782 law review articles that discuss or at least cite the case, and 5,219 news stories that mention the case by name (and presumably many others that discuss *Daubert's* existence, principles, or effects without naming it).

The psychological question asked by the experiments reported in this article is whether expert testimony acquires additional persuasive impact merely by virtue of the fact that it has been admitted through a judicial filter. This presumes, of course, that jurors are aware that judges serve as gatekeepers of evidence, something that appears to be common knowledge in American society, and perhaps made more salient in recent years by the *Daubert* quartet and continuing controversies over scientific evidence (Saks & Faigman, 2005).

The practical problem is that, if the hypothesis is correct, judges are inadvertently placing their thumbs on the fact-finding scale, and jurors are giving evidence more weight than it would receive on its own merits. Undesirable though this might be when the admitted evidence is sound, consider the situation when the evidence is unsound. Judges will, from time to time, admit expert evidence that is "shaky" or "expertise that is *fausse* [false] and science that is junky," to borrow from Justices Harry Blackmun and Antonin Scalia, respectively. What might be the consequence of letting "bad science" into the courtroom? Jurors might assume that all testimony put before them has passed the necessary tests established by the law. If false or misleading or unacceptably weak scientific expert testimony nevertheless passes through the legal filter, and if the fact that it has passed muster with the gatekeeper imbues it with unwarranted persuasive power, then at the end of the day, judicial gatekeepers will have inadvertently augmented the credibility of some evidence that ought not to have been admitted in the first place.

---

[1] More specifically, *Daubert* suggests that judges examine (a) whether the principles underlying proffered expert (scientific) testimony are testable and have been tested, (b) whether the methodology of that testing was sound (usually tagged as the "peer review and publication" factor), (c) what the results were of that testing (labeled as *error rates*), (d) maintenance of standards by the field using the proffered knowledge, (e) and whether the basis of the expert testimony was generally accepted, the central element of *Frye v. United States* (1923). In addition to the federal courts, about half the states have adopted some version of the *Daubert* package of admissibility rules, although a number of states retaining a version of the *Frye* test include several large and legally important states, among them California, Florida, and New York. (Faigman, Kaye, Saks, Sanders, & Cheng, 2008 [chap. 1]).

That judges are not altogether well prepared to serve as gatekeepers of scientific evidence, and that they do not always do the best job at it, has been documented and opined upon in a variety of ways and from a variety of directions. For present purposes, it should be sufficient to provide sources to which interested readers can refer (e.g., *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 1993 [Chief Justice William Rehnquist expressing doubt that judges can become the "amateur scientists" they need to become if they are to be reliable gatekeepers]; *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 1995 [Judge Alex Kozynski on remand expressing doubt that he or any judge can resolve disputes between experts]; "*Development in the Law*, 1995"; Faigman et al., 2008 [many chapters reviewing judicial opinions dealing with various scientific evidence topics]; Foster & Huber, 1997; Gatowski et al., 2001; Gross, 1991; Kovera & McAuliff, 2000; Moriarty & Saks, 2005; Risinger, 2000). A particularly interesting phenomenon is that judges have been far more active in excluding the relatively stronger evidence offered in civil cases than they have been in excluding the weaker science that is offered in criminal cases (Dixon & Gill, 2002; Groscup, 2004; Risinger, 2000; Saks & Faigman, 2005).

Related Research

We have found no research in the attitude change literature that addresses the persuasive impact of "gatekeeping" situations. Persuasion researchers, particularly those investigating the elaboration likelihood model (ELM) of persuasion (Petty & Cacioppo, 1986) have focused nearly exclusively on persuasive messages that are delivered directly from the source to the target. In the case of the courtroom, the judge acts as a moderator, allowing some messages to pass from source to target while blocking other messages. Thus, the considerable literature on source credibility effects speaks to a different issue. Moreover, the message itself is not altered by the judge making a judgment about it, although its meaning might be. The gatekeeper paradigm creates a unique situation: Although messages are still passed directly from the source to the target (i.e., from witness to juror), a third party has ostensibly evaluated the messages and dictated which messages are allowed to be presented.

Similarly, within the law and psychology literature, a body of research has developed on various aspects of expert evidence and its impact on jurors. Studies have investigated fact finders' sensitivity to the source characteristics of the expert (e.g., Cooper, Bennett, & Sukel, 1996; Cooper & Neuhaus, 2000; Greenberg & Wursten, 1988; Ivkovic & Hans, 2003; Saks & Wissler, 1984). Research has explored the impact of features of the substance of the expert evidence, including the quality of the science presented (e.g., Kovera, McAuliff, & Hebert, 1999; Krauss & Sales, 2001; McAuliff, Kovera, & Nunez, in press), the complexity of the testimony (e.g., Cecil, Hans, & Wiggins, 1990; Cooper, Bennett, & Sukel, 1996; Ellsworth, 1989; Hastie, Penrod, & Pennington, 1983; Horowitz, Bordens, Victor, Bourgeois, & Foster-Lee, 2001; Horowitz, Foster-Lee, & Brolly, 1996; Ivkovic and Hans, 2003; Krauss & Sales, 2001; Vidmar, 1998), its abstractness or concreteness in relation to the facts of the case at bar (e.g., Brekke & Borgida, 1988; Gabora, Spanos, & Joab, 1993; Kovera, Gresham, Borgida, Gray, & Regan, 1997), and other features. One of the most intensely and systematically studied

areas has been how laypersons handle statistical and especially probabilistic evidence (for reviews of this literature, see Sanders et al., 2008; Thompson & Cole, 2006; for several particularly illuminating experiments on this topic, see Koehler, 2001; Nance & Morris, 2005; Schklar & Diamond, 1999; Wells, 1992). Some research inquired into the use by jurors and judges of heuristics (e.g., Guthrie, Rachlinski, & Wistrich, 2001; Rachlinski, 2000; Saks & Kidd, 1981). The capability of cross-examination to undo the harm of erroneous or misleading evidence has also been studied (e.g., Diamond, Casper, Heiert, & Marshall, 1996; Kovera, Levy, Borgida, & Penrod, 1994; Kovera & McAuliff, 2000; Krauss & Sales, 2001).

This research has found (a) that lay fact finders generally are sensitive to differences in source characteristics; (b) that they have considerable difficulty evaluating the scientific quality of the evidence; (c) that scientific and other complex evidence presents special difficulty to fact finders; (d) that, as evidence becomes increasingly complex, fact finders turn to more superficial cues to validity (but they do not abandon the effort to make as much sense as they can of the substance, sometimes achieving good comprehension); (e) that the more concretely connected the testimony is to the specific case facts, the greater its impact; (f) that fact finders are more persuaded by storytelling and clinical testimony than by quantitative scientific evidence; (g) that they use heuristics in their decision making in the same way that others do in other decision contexts; and (h) that cross-examination is largely ineffective in undoing the impact of erroneous or misleading evidence.

Although none of those studies poses the question that is the concern of the present research, to the extent that they suggest that jurors have difficulty evaluating scientific evidence and that, faced with difficult evidence, they are alert to other cues to the validity of the evidence with which they are presented, the body of research tends to support the central hypothesis of the research reported in this article, namely that the admission of evidence provides a signal to jurors that the judge thought the expert evidence was sound and that, in reliance on that signal, jurors will give greater weight to the evidence than it inherently merits.

Testing for a Gatekeeper Effect

How, then, does one design a study to test the hypothesis of a possible gatekeeper effect? The fundamental challenge is to find appropriate comparison conditions. When proffered scientific evidence is challenged and a judge decides to admit it, the jury obviously has an opportunity to evaluate the evidence, perhaps being influenced by the fact that the judge admitted it, perhaps not. However, when the judge excludes, and the jury never even learns of the evidence, the jury has no occasion to think about it at all. In testing our hypothesis, it is necessary to find a way for the judge to exclude the evidence and for the fact finders to learn what the excluded evidence was.

Our solution was to create a condition in which our mock-juror participants are told that a judge had excluded an expert's study from the trial but that the evidence is nevertheless being presented to them for their evaluation and consideration with all of the other evidence in the case. Although this procedure tracks no normal trial scenario, it resembles the situation of appellate judges: When an

appeal challenges the exclusion of evidence, appellate judges know that the evidence was excluded, they have before them the proponent's offer of proof of that excluded evidence, and they can then evaluate that evidence. Circumstances exist when evidence (although not expert witness testimony) is admitted conditionally, subject to later determination of its relevance, and if that later determination is that it is not relevant, the jurors are expected to disregard the evidence that had been admitted. Our unavoidably peculiar condition is something of an amalgam of these procedures. It was important to the present study that this experimental condition contain the ingredients necessary to test the cognitive process that is at issue: whether jurors incorporate the signal of the judge's decision into their own evaluation of the evidence. Note also that the fact that the judge has made a screening decision is made explicit. If no effect can be found under these conditions, the likelihood that a gatekeeper effect exists is reduced. If an effect is found under these conditions, future research might try to pursue more ecologically valid strategies for testing the hypothesis.

The explicit-exclude condition just described is contrasted with the same evidence admitted in two different ways: explicit admission and implicit admission. In the former, participants are explicitly told that the judge decided to admit the expert's study. This kind of situation occurs when evidence is proffered in open court and challenged. The offers of proof and the judge's ruling would be made outside of the presence of the jury, but upon receiving the evidence, the jury would quickly see what the judge's decision was. This condition provides an "apples-to-apples" comparison with the explicit exclusion condition in that the judge's screening is made salient. In the implicit admission situation, the evidence is admitted but no mention is made that the court reviewed the study and decided that it was admissible; the scientific evidence study simply appeared as part of the case. This kind of situation occurs when anticipated evidence is challenged in a pretrial motion in limine, and the court rules before trial ever begins.

Overall, we expect mock jurors to be more persuaded by the scientific evidence when it is admitted into the trial, compared with when it has been excluded. More important, we expect the jurors to be more persuaded by the evidence when the judge has explicitly admitted it—making the gatekeeper's decision salient—than when the evidence is implicitly included, appearing without the explicit endorsement of the judge.

In addition, we varied the methodological quality of the research that the expert presented and of the expert's credibility. The purpose of varying these was simply to test the generality of any observed gatekeeper effect; that is, to test whether it operates under some conditions but not others, whether it operates when expert witnesses are of low as well as high credibility, and when the study admitted is of both low and high scientific validity. The absence of any interactions would confirm that a gatekeeper effect, if one is detected, operates at different levels of source credibility and of scientific evidence quality. Although other research suggests that trial fact finders are not sensitive to differences in quality between studies offered in evidence, a remote possibility exists that a discordance between a judge's decision to admit or exclude and the quality of the evidence (poor evidence admitted, good evidence excluded) might prompt jurors to work harder at evaluating the evidence.

In the first of our two experiments, we examine the impact of a judge's decision to admit or exclude scientific evidence on the perceived persuasiveness of that evidence. In the second experiment, we add a control condition to explore how the scientific research is perceived outside of the trial context, and we explore the impact of the admissibility decisions on verdicts.

Experiment 1

Method

Design. This first study used a 3 (gatekeeper decision: explicit-include, explicit-exclude, or implicit-include) × 2 (research validity: high vs. low) × 2 (source credibility: high vs. low) between-participants factorial design.

Participants. One hundred fifty-nine undergraduate students (100 female, 59 male) participated in this study; all volunteered in exchange for course credit. Although we did not collect information as to the ages of the participants, we required our participants to be of at least 18 years of age and U.S. citizens, making them jury eligible.

Materials. A set of documents that summarized the proceedings of a federal court trial was presented to the participants. This trial summary consisted of a four-page document designed to visually resemble the pleading format used for many court filings. The summary was written in paragraph form and did not include direct quotations from any party. Within this document was a one-page section that provided some background information on the case, a section that summarized the evidence presented during the trial, a section that very briefly described the testimony of two witnesses (one for the plaintiff and one for the defense), and a section that summarized the instructions that were presented to the jury. A fifth section, labeled "Judgment of the Court" was blacked out so as to suggest to participants that a judgment had been made but that they were not allowed to see what it was.

The case we developed involved a plaintiff who was seeking compensation from a large company for damages which, the plaintiff asserts, arose from use of that company's product: The parent of a child who died in a bicycle accident sued the manufacturer of the bicycle helmet that the child was wearing, claiming that the helmet was defective and did not offer adequate protection. In the course of the trial summary, within the section that summarized the evidence, a key piece of evidence was introduced to the mock jurors: a summary of a "research study." The trial summary stated that the research study was introduced by the plaintiff in support of his case. This research study supported the position of the plaintiff, ostensibly demonstrating that the defendant's bicycle helmets were, in fact, defective. All other information about the trial, including the testimony of the witnesses, was designed to be ambiguous so that it did not lend support to either party's case.

Procedure. Participants received a packet of test materials that consisted of a statement of informed consent, a summary of the aforementioned case, and the abstract of the scientific study that was to be entered into evidence. The participants completed all materials individually; however, they were tested in small groups of 5 to 8.

First, we manipulated the decision of the gatekeeper (i.e., evidentiary status of the evidence). In the explicit-include condition, we inserted a message in bold type:

> JUDICIAL REVIEW: The following research was subject to judicial review regarding admissibility as evidence. After thorough review, the study is admitted as Exhibit P3.

This was followed by the research abstract. In the explicit-exclude condition, a similar message was displayed, except it explained that the judge had excluded the evidence. Jurors in this condition were asked to turn to the last page of the trial summary, where the excluded research abstract was provided. They were asked to read the abstract before continuing on with the remainder of the trial summary. Finally, in the implicit-include condition, no mention was made of the judge's gatekeeping of the evidence; the "judicial review" paragraph did not appear; rather, the research study was simply presented along with the rest of the trial summary.

Next, we manipulated the credibility of the researcher who conducted the helmet study through a statement that preceded the research abstract. The high credibility version described the author as an engineering professor from a major university, whereas the low credibility version described the author as an investigative reporter from a local news agency.[2]

Finally, we manipulated the scientific merit of the research abstract: The low validity study used rudimentary methods, lacked a control group, and was conducted on a small sample, whereas the high validity study used experimental techniques, involved a control group, and used sophisticated technology with a representative sample. These two versions were created with the idea that the low validity study was objectively flawed and would likely fail a *Daubert* test, whereas the high validity study was more rigorous and would presumably pass a *Daubert* test and be admitted into evidence. The full abstracts and author descriptions are given in the Appendix.

After reading the trial materials, the participants were presented with a questionnaire that included five items designed to assess the persuasive impact of the study (e.g., "How convinced were you by [the research study]?" "Did the researchers do a good job designing the study?" "Do you think [the author of the study] is a good researcher?"). Once participants completed all of the materials, they were debriefed and released.

Results

The five items that were designed to measure the participants' perception of the scientific evidence were entered into a principal-axis factor analysis. The analysis extracted a single factor accounting for 65.2% of the overall variance. The internal reliability of these items (Cronbach's α) was 0.90. On the basis of the results of these analyses, we averaged those items into a composite measure of "research persuasiveness."

We then conducted a 3 (gatekeeper decision: explicit-include, implicit-include, or explicit-exclude) $\times$ 2 (research validity: high vs. low) $\times$ 2 (source credibility: high vs. low) between-participants analysis of variance (ANOVA) on

[2] In a pilot test involving 22 undergraduate students, the reported credibility of a professor as a source of scientific information ($M = 6.36$ on a 7-point scale) was significantly higher than the reported credibility of a television reporter ($M = 3.13$), $t(20) = 6.50$, $p < .001$.

the research persuasiveness index. A main effect of the gatekeeper's decision was discovered, $F(2, 151) = 5.08$, $p = .01$; $\eta^2_{\text{partial}} = .063$. However, no effect of research validity or source credibility emerged, nor did any interactions (all ps $> .17$), suggesting that the admissibility status of the helmet research as determined by the gatekeeper was the sole predictor of the perceived quality of the research. To examine our specific hypothesis that our participants would be less persuaded by the research in the exclusion condition than either of the inclusion conditions, we conducted paired-comparison tests between each of the three gatekeeper decision conditions. Consistent with expectations, participants' ratings of the study in the exclusion condition ($M = 4.23$) were significantly lower than their ratings in the implicit-include condition ($M = 4.91$), $t(151) = 2.93$, $p < .01$; and marginally lower than their ratings in the explicit-include condition ($M = 4.62$), $t(151) = 1.70$, $p = .09$. Contrary to expectations, however, a second simple contrast test revealed that the ratings of the research study did not significantly differ between the explicit- and implicit-include conditions, $t(151) = 1.17$, $p = .24$. (See Figure 1.)

Brief Discussion

When participants were presented with research offered as evidence in a trial, the evidentiary status of the research was the primary predictor of how the research was perceived by mock jurors. When the trial judge ruled the evidence inadmissible, participants found the helmet research to be less persuasive than when the evidence was admitted into the trial, although it did not matter whether the included evidence came with the explicit endorsement of the gatekeeper. Although the expert evidence research was manipulated to vary in its validity and the source of the research was manipulated to vary in credibility, these factors did not affect the persuasiveness of the message. This finding lends initial support to the concern that jurors, who ideally are to be objective and independent evaluators of evidence, can be influenced by the decision of the gatekeeper when they assess the quality of scientific evidence.

Several questions are raised by these results. Jurors saw no difference between scientific evidence that was admitted, whether it was admitted after a careful review by the judge (explicit-include) or there was no mention of such a

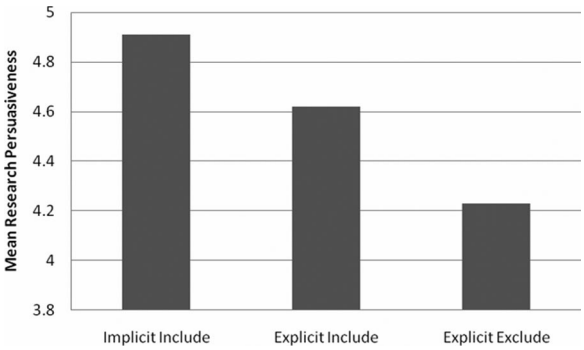

Figure 1. Mean Persuasiveness Rating × Condition, Experiment 1.

review (implicit-include). What does this suggest about how jurors view trial evidence? Does something about the trial setting cause people to view evidence differently than they would view the same evidence in a setting other than a trial? That is, when the evidence is implicitly admitted at trial, might jurors already be assuming that it has been scrutinized and approved by the court?

The second experiment attempts to replicate the findings in the first experiment, adds an additional condition in which participants view the research outside of a trial context, and examines how verdicts are affected by the changes in the persuasiveness of the scientific evidence caused by the manipulations. We expected again to find that included evidence is viewed as more persuasive than excluded evidence but, furthermore, that included evidence is viewed as more persuasive than that same evidence if viewed outside of a trial context. Finally, if our first hypothesis was correct, we expected to find that the difference in perceived persuasiveness would in turn affect our participants' subsequent verdicts.

Experiment 2

Method

Design. The second study used a 4 (gatekeeper decision: explicit-include, explicit-exclude, implicit-include, or control) × 2 (research validity: high vs. low) × 2 (source credibility: high vs. low) between-participants factorial design.

Participants. A total of 196 undergraduate students (131 female, 65 male) participated in this investigation; all volunteered in exchange for course credit. Although we did not collect information as to the ages of the participants, we required our participants to be of at least 18 years of age and U.S. citizens, making them jury eligible.

Materials and procedure. The second experiment's materials and procedures were similar to those used in the first experiment: We used the same trial summary containing the same set of research studies and the same manipulations of research validity and source credibility. The primary differences involved the inclusion of a fourth gatekeeper condition and an additional set of dependent measures. To address the main question raised by the first study, one set of participants viewed the research abstract but not any of the trial materials; that is, these control participants simply viewed the various forms of the research abstract and then answered the questions pertaining to the quality and persuasiveness of the research. The participants in the other conditions, however, read the full trial materials and then played the role of juror by deciding for either the plaintiff or defendant. In the exclusion condition, in which the judge excluded the research from the trial, the participants were again supplied with (and instructed to read) the research abstract, knowing that the judge had excluded it from evidence. If the judge's exclusion of the testimony had an effect on the weight the jurors gave to the evidence, that impact would be reflected in lower ratings of the evidence compared wth other conditions. At the conclusion of the trial, materials were the summarized jury instructions that read:

> Plaintiff must prove through the balance of evidence that [the defendant] was responsible for the death of [the plaintiff's son], particularly:

> 1. [The defendant] was negligent in the manufacturing of the bicycle helmet worn by the deceased.
> 2. [The defendant's] product was not sufficient protection against injuries that might result from a bicycle accident.
>
> If the jury can find that both of the above items have been shown to be true by the evidence presented, judgment should be for the plaintiff. If one or both of those of the above items has not been shown, judgment should be for the defendant.

Participants were then asked to provide a verdict and rate their confidence in their verdicts, and they were given the opportunity to award damages to the plaintiff when they had found the defendant's product defective and the defendant therefore liable for the plaintiff's losses. This set of items was omitted for participants in the control condition. Following these items were the measures of research persuasiveness used previously.

Results

As with the previous study, the five items that were designed to measure the participants' perception of the scientific evidence were entered into a principal-axis factor analysis. The analysis extracted a single factor accounting for 71.2% of the overall variance. The internal reliability of these items (Cronbach's α) was 0.92. On the basis of the results of these analyses, we averaged the five items into a composite measure of research persuasiveness.

Effects of gatekeeper decision. A 4 (gatekeeper decision: explicit-include, explicit-exclude, implicit-include, or control) $\times$ 2 (research validity: high vs. low) $\times$ 2 (source credibility: high vs. low) between-participants ANOVA was conducted on the mock jurors' judgments of the persuasiveness of the bicycle helmet research.

We found several main effects but no significant interactions. Unlike our first study, the research validity manipulation produced a significant main effect: Those participants who viewed the more methodologically valid research found it to be more persuasive ($M = 4.73$) than those who viewed the low validity research ($M = 4.29$), $F(1, 179) = 6.59$, $p < .01$, $\eta^2_{\text{partial}} = .037$; and although the source credibility manipulation did not produce a significant main effect, $F(1, 179) = 2.89$, $p = .16$, the means were in the expected direction ($M_{\text{high}} = 4.64$, $M_{\text{low}} = 4.39$). Most important, the gatekeeper decision produced a significant main effect, $F(3, 179) = 4.00$, $p < .01$; $\eta^2_{\text{partial}} = .063$. Confirming the findings of the previous experiment, a contrast analysis demonstrated that the implicit-include ($M = 4.80$) and explicit-include ($M = 4.79$) conditions did not differ, $t(179) = 0.09$, *ns*; and that the explicit-exclude condition ($M = 4.26$) produced significantly lower ratings of the evidence than the two inclusion conditions did combined ($M = 4.79$), $t(179) = 2.47$, $p < .02$, $\eta^2_{\text{partial}} = .033$. Although the jurors' reported confidence in their verdicts did not differ among the various conditions, participants who found the research to be more persuasive reported having more confidence in their verdicts, $r(143) = .39$, $p < .001$.

Finally, we compared the control condition, in which the research was presented without any mention of a trial or gatekeeper, with the other conditions, using a simple contrast. Confirming our hypothesis, the mean research persua-

siveness score in the control condition ($M = 4.19$) significantly differed from that in the implicit-include condition ($M = 4.80$), $t(179) = 2.67$, $p < .01$, $\eta^2_{\text{partial}} = .04$; and that in the explicit-include condition, ($M = 4.79$), $t(179) = 2.68$, $p < .01$, $\eta^2_{\text{partial}} = .04$; but did not differ from that in the explicit-exclude condition ($M = 4.26$), $t(179) = .43$, *ns*. These means are presented in Figure 2.

Damage awards, which were capped at $5 million, were given by 110 of the participants; the mean award size was $2.66 million, with a standard deviation of $1.61 million. The distribution of awards was not significantly skewed; skewness = .296, $SE_{\text{skew}} = .230$. No significant relationships were found between damage awards and any of the manipulated variables; however, damage awards were marginally correlated with the research persuasiveness measure, so that those participants who found the research to be more persuasive tended to give larger awards, $r(107) = .15$, $p = .13$.

Mediation analysis. If the gatekeeper's decision is affecting the persuasiveness of the expert evidence, then the decision might be affecting verdicts as well. To test for this, we conducted a mediation analysis using the product of coefficients method[3] to test the "gatekeeper decision (X) affects research persuasiveness (M) affects verdict (Y)" path. To simplify the analysis, the gatekeeper decision variable was collapsed into a new variable so that the two inclusion conditions were combined into a single condition (because these two conditions did not differ on any of the analyses in either of the experiments), leaving a dichotomous variable (include vs. exclude). This new condition variable was significantly predictive of the research persuasiveness measure, $b = .536$, $t(143) = 2.41$, $p < .02$. In addition, the research persuasiveness measure was significantly predictive of the verdict when the variable was included in the equation, $b = .948$, odds ratio = 2.58, Wald = 23.21, $p < .001$. The mediated effect was found to be significant; $ab = .508$, $s_{ab} = .236$, 95% confidence interval = ($.0455 < ab < .9705$).[4] In other words, the gatekeeper's decision to include the evidence increased the persuasiveness of the evidence, which then increased the odds of a verdict for the plaintiff.[5]

General Discussion

Experiment 2 replicated the results of Experiment 1: When the judge excluded the proffered scientific evidence, jurors found that evidence to be of lower quality (and less persuasive) than when the evidence had been admitted. Furthermore,

---

[3] The product of coefficients method was chosen because it is a more powerful test than the causal steps (popularized by Baron & Kenny, 1986) and difference of coefficients (e.g., Sobel test) methods (Fritz & MacKinnon, 2007). In addition, because of the mixture of logistic and ordinary least squares (OLS) regression in this model, tests that rely on difference of coefficients result in grossly inaccurate estimates without substantial rescaling of the coefficients (MacKinnon, 2008).

[4] Coefficient a is calculated by an OLS regression equation: $M = aX + i$; $a = .536$, $s_a = .223$. Coefficient b is calculated in a logistic regression equation: $Y = c'X + bM + i$; $b = .948$, $s_b = .197$. The mediated effect, ab, is the product of a and b; the standard error of ab is calculated as $\sqrt{(a^2 s_b^2 + b^2 s_a^2)}$.

[5] Although we are most concerned with the indirect path of the gatekeeper decision affecting verdicts through a change in perceived persuasiveness, the direct path was nonsignificant, $\chi^2(1, N = 196) = 2.28$, $p = .15$.

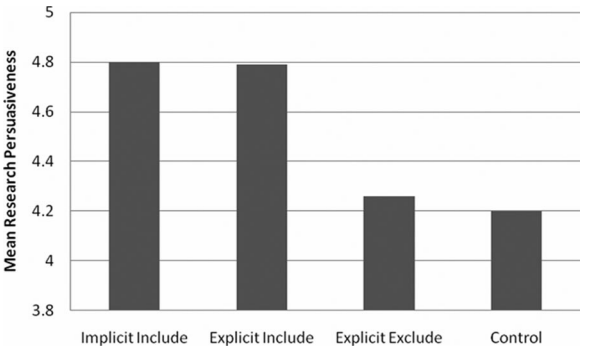

Figure 2. Mean Persuasiveness Rating × Condition, Experiment 2.

there was no difference between the implicit-include and explicit-include conditions, again suggesting that jurors operate under the assumption that judges review scientific evidence (perhaps all evidence) before its presentation at the trial.

In addition, two new findings emerged from the second experiment: First, when we presented the scientific evidence outside of the courtroom context, the jurors' perceptions of that evidence were similar to the perceptions of the jurors in the exclusion condition. This is a key finding. It demonstrates that jurors' impressions of the scientific evidence were inflated when the judge admitted the evidence into the trial; when that same evidence was excluded from the trial, the jurors held it in lower regard (gave it lower ratings)—nearly identical to how they would have rated it had they come across the evidence in a noncourtroom setting. This reinforces our interpretation that jurors, perhaps nonconsciously, assume that all expert evidence admitted into a trial (the implicit-include condition) has been "approved" by a competent gatekeeper (because the very same evidence encountered outside of the trial context received significantly lower ratings).

Second, as demonstrated by our mediation analysis, not only did the judge's decision to include or exclude the scientific evidence affect the jurors' perceptions of that evidence, but also the judge's admission decision then indirectly affected the jurors' verdicts.

Taken together, the results of these experiments suggest that jurors assume too much about the quality of scientific evidence presented at trials. Specifically, jurors assume that judges review scientific evidence before it is presented to them, and that any evidence used in a trial must be above some threshold of quality. Because of these assumptions, jurors seem to be less critical of scientific evidence used in trials and are more persuaded by it. If trial judges are adhering to the *Daubert* standard, the jurors' assumption would make sense; however, not every jurisdiction uses the *Daubert* standard, and even within *Daubert* jurisdictions there is reason to believe that many judges do a poor job of screening expert evidence (Gatowski et al., 2001; Kovera & McAuliff, 2000), especially in criminal cases (Moriarty & Saks, 2005). Thus, the sort of evaluation that is necessary to prevent bad science from reaching the jury is not being done; at least sometimes and perhaps often. At the same time, jurors credit the scientific evidence unduly because of their apparent assumption about judicial filtering.

Implications for the Legal System

When judges allow expert testimony to reach the jury, they are implicitly lending credence to the testimony, increasing its persuasiveness. This tips the scales toward the party offering the expert witness, perhaps affecting the jury's verdict. For example, in the two inclusion conditions, 79% of our participants found for the plaintiff, whereas in the exclusion condition, only 67% found for the plaintiff, a difference that was mediated by the participants' evaluations of the research abstract. Moreover, as found in Experiment 2, the ratings in the exclusion condition were nearly identical to those in the control condition. If we assume that the control condition provides an estimate of what participants thought of the research unaided and uninfluenced by the views of others, then we can infer that the exclusion condition returned the evidence to what might be regarded as its basic (pregatekeeping) value in the eyes of jurors. This leads to the further inferences that, by admitting the evidence into the trial, the judge has unavoidably inflated the persuasiveness of that evidence and (at least in our second experiment) altered the proportion of plaintiffs' verdicts.

Ironically, a landmark Supreme Court decision motivated in large part by a desire to shield jurors from "junk science" could serve to heighten the impact of false or misleading scientific evidence when judges allow it through the courtroom gates. If a judge's decision to admit evidence endows that evidence with additional weight, and if that phenomenon is exacerbated by a *Daubert* ethos, then the burden on judges to make the correct gatekeeping decision is that much greater.

If the gatekeeper effect influences much or all evidence included in a trial, what could be done to ameliorate its effect? One solution would be to find ways to ensure that courts (whether through judges or special masters selected for their specific expertise in the subject matter) exercise their gatekeeping effectively—minimizing occasions when evidence that should be admitted is excluded or evidence that should be excluded is admitted. Another approach that would permit judges to perform their gatekeeping duties without simultaneously undermining the ability of jurors to independently exercise their critical judgment might be to fashion an instruction that emphasizes to jurors that the judge is not an unfailing authority on evidence, which is one important reason why we have jurors and why they must ably and independently evaluate all expert evidence that is presented to them. Implementing either of the approaches outlined in this paragraph would be quite a challenge (see Sanders, Saks, & Schweitzer, 2007).

Implications for Persuasion Research

Although not the focus of this research, a gap in the persuasion literature has been exposed in the present work. Persuasion researchers have focused nearly exclusively on messages that are delivered directly from the source to the target; however, in a courtroom, the judge acts as a moderator, allowing some messages to pass from source to target while precluding delivery of other messages. The same moderator structure can be found in other settings, such as those involving teachers (selecting the materials for a course) and editors (accepting or rejecting news stories, journal articles, etc.). This gatekeeper paradigm creates a special situation: Although messages are still passed directly from the source to the target

(i.e., from witness to juror), a third party has ostensibly evaluated the messages and decided which messages are allowed to be presented.[6]

The persuasion literature has ignored situations such as those occurring in the gatekeeper paradigm, but related research on multiple sources and source characteristics (e.g., Harkins & Petty, 1981a, 1981b, 1987) suggests that audiences receiving information that involves a third party would be pushed toward increased central processing (i.e., they should pay more attention to the content of the message). This does not, however, appear to be occurring in our data. In Experiment 2, both source and message attributes were manipulated. If the presence of a gatekeeper had nudged participants to engage in more central processing, we should have seen differential effects of the manipulation between the control condition (in which there was no gatekeeper) and the courtroom situation (in which there was a gatekeeper). This effect would have been manifested as a Message Quality × Gatekeeper Condition interaction, but no such interactions were found.[7] Furthermore, the research on multiple source characteristics (a situation in which there are multiple peripheral cues that affect the persuasiveness of a message; Ziegler, Diehl, & Ruther, 2002) suggests that conflicting peripheral cues (e.g., low credibility source, high credibility gatekeeper) lead to an increase in central processing. Assuming the participants considered the judge to be a credible gatekeeper, participants presented with research from the low-credibility witness might have been more attuned to the quality of the message, compared with those participants who heard from the high-credibility witness. Again, no such interaction was found.[8] Unfortunately, this study may lack the statistical power to detect small three-way interactions that could potentially add texture to our primary finding and allow for more direct comparisons with traditional ELM experiments. We leave it to future studies to determine how—if at all—this moderated persuasion paradigm fits into existing persuasion theories.

Limitations and Future Directions

One of the most common concerns about paper-and-pencil mock trial studies is that the lack of ecological validity will translate into a lack of external validity—that the findings of such studies will not generalize to actual jurors. That is, of course, a concern with the findings in this article. However, we see little reason to suggest that a selection of actual jurors would be less affected by the

---

[6] In the courtroom, the gatekeeper serves as a *moderator*—filtering messages but allowing those messages to be passed directly from the source to the target. An alternate version of the gatekeeper paradigm could involve *mediation*, in which the gatekeeper receives the message from the original source, evaluates the message, and then personally presents the message to the target. Both such situations have gone unnoticed by the persuasion literature.

[7] The Message Quality × Gatekeeper Condition interaction was nonsignificant, $F(3, 179) = 0.66$, $p = .56$.

[8] To specifically look for this possibility, we conducted a 2 (credibility) × 2 (quality) ANOVA within the two gatekeeper inclusion conditions combined. We would have expected to find an interaction such that the quality manipulation was more powerful when the source credibility was low (and thus differed from the judge's credibility) rather than high (in line with the judge's credibility). No interaction was found, $F(1, 96) = 0.07$, $p = .79$.

persuasive impact of a gatekeeper, and perhaps the formality and pomp of an actual trial may reinforce the jurors' notion of the court-as-gatekeeper and thus increase the size of our effect.

In addition, one could question whether our statistically significant effects hold any real-world significance. For example, in our second experiment, the composite persuasiveness measures differ by only 0.6 points on a 7-point scale. Nevertheless, we note that our partial eta-squared effect size of .04 approaches a moderate size as suggested by Cohen (1988), and furthermore, Rosenthal (1991) demonstrated that even very small effects can have a substantial real-world impact, especially when extrapolated across a large population (e.g., the thousands of cases involving scientific evidence heard annually).

Although far from conclusive, our data suggest both that a gatekeeper phenomenon exists and that there may be more to the gatekeeper paradigm (and moderated persuasion situations in general) than what established models of persuasion can explain. Further study of such situations ought to be of interest to both applied and basic researchers.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Brekke, N., & Borgida, E. (1998). Expert psychological testimony in rape trials: A social–cognitive analysis. *Journal of Personality and Social Psychology, 55,* 372–386.

Cecil, J., Hans, V., & Wiggins, E. (1990). Citizen comprehension of difficult issues: lessons from civil jury trials. *American University Law Review, 40,* 727–774.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Cooper, J., & Neuhaus, I. (2000). The hired gun effect: Assessing the effect of pay, frequency of testifying, and credentials on the perception of expert testimony. *Law and Human Behavior, 24,* 149–172.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 587, 113, S. Ct. 2786, 125 L. Ed. 2d 469 (1993).

Daubert v. Merrell Dow Pharmaceuticals, Inc., 43 F. 3d 1311, 1316 (9th Cir. 1995).

*Development in the law: Confronting the new challenges of scientific evidence. (*1995). *Harvard Law Review, 108,* 1481–1605.

Diamond, S., Casper, J., Heiert, C., & Marshall, A. (1996). Juror reactions to attorneys at trial. *Journal of Criminal Law and Criminology, 87,* 17–47.

Dixon, L., & Gill, B. (2002). Changes in the standards for admitting expert evidence in federal civil cases since the *Daubert* decision. *Psychology, Public Policy, & Law, 8,* 251–308.

Ellsworth, P. (1989). Are twelve heads better than one? *Law and Contemporary Problems, 52,* 205–224.

Faigman, D., Kaye, D., Saks, M., Sanders, J., & Cheng, E. (2008). *Modern scientific evidence: The law and science of expert testimony, Vol. 1.* St. Paul, MN: West Group.

Foster, K., & Huber, P. (1997). *Judging science: Scientific knowledge and the federal courts.* Cambridge: MIT Press.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18,* 233–239.

Frye v. United States, 293 F. 1013 (D. C. Cir. 1923).

Gabora, N., Spanos, N., & Joab, A. (1993). The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial. *Law and Human Behavior, 17,* 103–119.

Gallanis, T. P. (1999). The rise of modern evidence law. *Iowa Law Review, 84,* 499–560.

Gatowski, S. I., Dobbin, S. A., Richardson, J. T., Ginsburg, G. P., Merlino, M. L., & Dahir, V. (2001). Asking the gatekeepers: A national survey of judges on judging expert evidence in a post-*Daubert* world. *Law and Human Behavior, 25,* 433–458.

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

Greenberg, J., & Wursten, A. (1988). The psychologist and psychiatrist as expert witnesses: perceived credibility and influence. *Professional Psychology: Research and Practice, 19,* 373–378.

Groscup, J. (2004, March 4–7*). Legalized gambling, beekeeping or science? Judicial decision-making about scientific evidence in the aftermath of Daubert/Kumho.* Poster presented at the American Psychology-Law Conference, Scottsdale, AZ.

Gross, S. (1991). Expert evidence. *Wisconsin Law Review, 1991,* 1113–1232.

Guthrie, C., Rachlinski, J., & Wistrich, A. (2001). Inside the judicial mind. *Cornell Law Review, 86,* 777–830.

Harkins, S., & Petty, R. (1981a). Effects of source magnification of cognitive effort on attitudes: An information-processing view. *Journal of Personality and Social Psychology, 40,* 401–413.

Harkins, S., & Petty, R. (1981b). The multiple source effect in persuasion: The effects of distraction. *Personality and Social Psychology Bulletin, 7,* 627–635.

Harkins, S., & Petty, R. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology, 52,* 260–268.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury.* Cambridge, MA: Harvard University Press.

Horowitz, I., Bordens, K., Victor, E., Bourgeois, M., & Foster-Lee, L. (2001). The effects of complexity on jurors' verdicts and construction of evidence. *Journal of Applied Psychology, 86,* 641–652.

Horowitz, I., Foster-Lee, L., & Brolly, B. (1996). Effects of trial complexity on decision making. *Journal of Applied Psychology, 81,* 757–768.

Ivkovic, S., & Hans, V. (2003). Jurors' evaluations of expert testimony: Judging the messenger and the message. *Law and Social Inquiry, 28,* 441–482.

Koehler, J. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior, 25,* 493–574.

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.

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

Kovera, M., McAuliff, B., & Hebert, K. (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,* 362–375.

Krauss, D. A., & Sales, B. D. (2001). The effects of clinical and scientific expert testimony

on juror decision making in capital sentencing. *Psychology, Public Policy, & Law, 7,* 267–310.
Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).
Landsman, S. (1984). *The adversary system: A description and defense.* Washington, DC: American Enterprise Institute for Public Policy Research.
MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Mahwah, NJ: Erlbaum.
McAuliff, B., Kovera, M., & Nunez, G. (in press). Can jurors recognize missing control groups, confounds, and experimenter bias in psychological science? *Law and Human Behavior.*
Moriarty, J. C., & Saks, M. J. (2005, Fall). Forensic science: Grand goals, tragic flaws, and judicial gatekeeping, *Judges Journal, 44,* 16–33.
Nance, D., & Morris, S. (2005). Juror understanding of DNA evidence: An empirical assessment of presentation formats for trace evidence with a relatively small random-match probability. *Journal of Legal Studies, 34,* 395–444.
Petty, R., & Cacioppo, J. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.
Rachlinski, J. (2000). Heuristics and biases in the courts: Ignorance or adaptation? *Oregon Law Review, 79,* 61–102.
Risinger, D. M. (2000). Navigating expert reliability: Are criminal standards of certainty being left on the dock? *Albany Law Review, 64,* 99–152.
Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist, 46,* 1086–1087.
Saks, M. J., & Faigman, D. L. (2005). Expert evidence after *Daubert*. *Annual Review of Law and Social Science, 1,* 105–130.
Saks, M. J., & Kidd, R. F. (1981). Human information processing and adjudication: Trial by heuristics. *Law & Society Review, 15,* 124–160.
Saks, M. J., & Wissler, R. L. (1984). Legal and psychological bases of expert testimony. *Behavioral Sciences and the Law, 2,* 435–449.
Sanders, J., Saks, M., & Schweitzer, N. (2008). Trial factfinders and expert evidence. In D. Faigman, D. Kaye, M. Saks, J. Sanders, & E. Cheng (Eds.), *Modern scientific evidence (2007–2008 ed.).* St. Paul, MN: Thompson/West.
Schklar, J., & Diamond, S. S. (1999). Juror reactions to DNA evidence: Errors and expectancies. *Law and Human Behavior, 23,* 159–184.
Thompson, W., & Cole, S. (2006). Psychological aspects of forensic identification evidence. In M. Costanzo et al. (Eds.), *Expert Psychological Testimony for the Courts.* Mahwah, NJ: Erlbaum.
Vidmar, N. (1998). The performance of the American civil jury: An empirical perspective. *Arizona Law Review, 40,* 849–900.
Weisgram v. Marley Co. 528 U.S. 440 (2000).
Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62,* 739–752.
Ziegler, R., Diehl, M., & Ruther, A. (2002). Multiple source characteristics and persuasion: Source inconsistency as a determinant of message scrutiny. *Personality and Social Psychology Bulletin, 28,* 496–508.

(*Appendix follows*)

Appendix

Research Abstracts and Author Descriptions

High Validity (With High-Credibility Source)

Product evaluation of HEADSTRONG-brand bicycle helmets.

Howard Spearman, PhD, Associate Professor of Engineering, University of Wisconsin.

To evaluate the effectiveness and overall safety provided by HEADSTRONG-brand bicycle helmets, electronic crash-test dummies specifically designed to simulate an 8-year-old child were used. The dummies were fitted with a helmet from either HEADSTRONG or other manufacturers, positioned on stools 3 ft. (0.914 m) high, and forcefully thrown to the ground to simulate a bicycle accident. Data from sensors placed inside the dummies' heads revealed that HEADSTRONG helmets were significantly less effective than their competitors at absorbing the impact from the accidents, making it more likely that head injuries would occur. It is concluded that HEADSTRONG-brand helmets do not provide adequate protection for children.

Low Validity (With Low-Credibility Source)

Product evaluation of HEADSTRONG-brand bicycle helmets.

Howard Spearman, investigative reporter, WTMJ Broadcasting Corporation.

To evaluate the effectiveness and overall safety provided by HEADSTRONG-brand bicycle helmets, five of these helmets were purchased from stores in and around Milwaukee, WI. A 60-lb. (~27-22-kg) piece of concrete was obtained to simulate the weight of the average 8-year-old child. The helmets were placed on a solid surface, and the concrete weight was dropped from a height of 4 feet (1.219 m) onto each helmet. In all cases, the helmet suffered severe damage, with large cracks appearing in four of the five helmets. Two of the five helmets were destroyed, splitting completely in half. It is concluded that HEADSTRONG-brand helmets do not provide adequate protection for children.

Received September 30, 2008
Revision received January 2, 2009
Accepted January 5, 2009 ■