Quick Take

- Judges, attorneys, and juries poorly evaluate the quality of expert evidence in legal proceedings.
- Cross-examination of experts and opposing testimony from another authority only sometimes help jurors see flaws in research presented as evidence.
- Scholars must continue to study how to make science more digestible to help judges and juries reach well-informed decisions.

Thirty years ago, I started my first faculty position, and I was petrified. The last year had been rough as I worked on my dissertation and searched for a job. I had failed to find a tenure-track position and now faced 2 years of a visiting position, after which I would have to begin the job search anew. My self-
confidence was shaken, and I was quite convinced that I would never have another idea worthy of research attention, in part because I believed that the research questions that I had been pursuing had reached a dead end.

As a graduate student, I had conducted several studies exploring whether expert testimony about psychological science influenced jurors’ decisions. Several of these studies examined whether information about how children react to being sexually abused influenced the verdicts that undergraduate students who were playing the role of jurors rendered, both individually and in group deliberations.

As I pondered my own research and the literature on expert testimony, it seemed to me that everyone had been asking the same fundamental question: Did expert testimony on psychological science have the intended effect on jurors’ verdicts? The answer was uniformly that it did, no matter the testimony’s topic:

- Experts who testified about the conditions that led eyewitnesses to make mistaken identifications caused mock jurors to render fewer guilty verdicts against defendants whom an eyewitness had identified under those conditions (Cutler et al., 1990).
- When a psychologist testified about how women who had been battered perceive their ability to leave a relationship, mock jurors were more likely to find that a woman had acted in self-defense when she killed her abusive partner and to acquit her of murder (Schuller, 1992).
- Learning from a psychological expert that both children and women commonly delay reporting sexual abuse increased the likelihood that mock jurors believed the claims of sexual abuse (Kovera et al., 1997) and rape made by complainants who had delayed their initial reports (Brekke & Borgida, 1988).

Was it necessary to test whether expert testimony describing yet another type of psychological science influenced jurors’ decisions? I didn’t think so, but I was not sure where to direct my research next.
Luckily for me (and my early career), the U.S. Supreme Court was simultaneously grappling with issues surrounding the admission of expert evidence at trial.

In June 1993, just as I was packing to make the move halfway across the country to start that first job, the Supreme Court handed down a landmark decision on the admissibility of expert testimony in court proceedings in the case of *Daubert v. Merrell Dow Pharmaceuticals*. Since the 1920s, many judges had been evaluating the admissibility of expert evidence by relying on what is known as the *Frye* standard—scientific evidence was admissible if it had been generally accepted by the relevant scientific community (*Frye v. United States*, 1923). But the *Federal Rules of Evidence* that Congress passed in 1975 created confusion among the lower courts about what standard judges should be using to evaluate the admissibility of scientific evidence.

In *Daubert*, the Supreme Court attempted to clarify the standards and the judges’ role in determining admissibility. *Daubert* held that judges were to be the gatekeepers responsible for evaluating whether the science that a party sought to introduce into evidence was reliable and relevant to understanding an issue in dispute. Essentially, the Supreme Court justices ruled that judges were to determine whether the scientific evidence was of sufficient quality to be admitted. The ruling provided judges with a nonexhaustive list of criteria by which they were to judge the reliability of the evidence, including the error rate associated with a particular test and whether

- the science was based on a falsifiable theory,
- the research had been peer-reviewed, and
- the science had been generally accepted by the scientific community (a criterion taken from the original *Frye* test).

The justices also argued that not only were judges up to the task of assessing the scientific reliability of research proffered for admission, but if they failed in some cases and admitted “junk” science, additional safeguards would limit the effects of this junk science on jurors’ decisions. These safeguards included judicial instruction on the burden of proof that jurors were to use to decide a case, cross-examination, and the presentation of contradictory evidence, perhaps through opposing expert testimony. Later Supreme Court decisions (*General Electric Co v. Joiner*, 1997; *Kumho Tire Co. v. Carmichael*, 1999) clarified that the justices had intended to apply their holding in *Daubert* to all expert testimony, not just scientific testimony.

The Supreme Court decision in *Daubert* raised a host of new research questions about how legal decision-makers evaluate expert evidence—questions that kept me busy for the better part of the next 30 years. No longer was the question whether psychological evidence did influence legal decision-makers, but whether it should. Were actors in the legal system up to the task of evaluating the quality of the expert evidence before them? After all, social psychologists had discovered that advanced law students were not nearly as capable of evaluating statistical claims and research methods as were advanced graduate students in psychology. Could judges accurately gauge the quality of scientific evidence, including psychological science, to determine whether it should be admitted at trial? If they weren’t, would attorneys know enough about the characteristics of good science to effectively cross-examine an expert who testified about flawed research? Would they know to call an opposing expert to challenge the unreliable science promoted by the opponent’s expert? And if jurors couldn’t grasp the variations in the reliability of the science underlying an expert’s testimony, would judicial instructions, cross-
examination, and opposing expert testimony help them make decisions that better reflected the quality of that evidence?

First, a colleague and I examined whether the quality of the psychological science underlying an expert’s planned testimony affected judges’ decisions to admit it (Kovera & McAuliff, 2000). Judges read about testimony that a social psychologist intended to introduce in a workplace sexual harassment case, relying on her own study exploring how exposure to sexually suggestive material affected men’s behavior toward women in the work environment. We varied whether the study was internally valid, included a control group, contained a confound, or allowed for experimenter expectancy effects. We also measured whether judges had any previous scientific training.

The quality of the research did not affect judges’ decisions about admitting the expert evidence at trial, suggesting that at least some judges would admit invalid science and that attorneys would need to help educate jurors about the flaws in the psychologist’s research methods.

Another study varied whether the assessment tool that an expert psychologist used to test the intelligence of a student was reliable (Chorn & Kovera, 2019). Once again, judges were likely to allow the testimony whether or not the test used by the psychologist was reliable.

My colleagues and I have also examined attorneys’ ability to spot problems with the research methods in the studies underlying experts’ opinions and to modify their litigation strategies accordingly (Kovera, Russano, & McAuliff, 2002). Presenting attorneys with the same materials that the judges received, we asked them whether they would file a motion to exclude the testimony from trial. Our research indicated that the study quality did not affect attorneys’ intentions to move for the testimony’s exclusion; 95% reported they would file such a motion. Attorneys’ self-reported strategies for cross-examination rarely mentioned problems with the internal validity of the expert’s research. The attorneys also questioned the internal validity of the high-quality study as frequently as the studies with research designs that contained significant threats to internal validity (e.g., missing control group, a confound).

**Jurors’ judgments**

Jurors do not perform much better than judges or attorneys when evaluating the quality of expert evidence. In one study, we examined jurors’ grasp of construct validity (the extent to which a test measures what it intends to measure; Kovera, McAuliff, & Hebert, 1999). We tested whether expert evidence based on research with poor construct validity had the same influence on jurors’ verdicts as did research with great construct validity. In several studies, jurors failed to recognize the same threats to internal validity that judges and attorneys had failed to appreciate: missing control groups, confounds, and the potential for bias from experimenter expectancy effects (McAuliff & Kovera, 2008; McAuliff, Kovera, & Nunez, 2009).

Two of the potential safeguards that the Supreme Court had argued would help jurors make more informed decisions, cross-examination and opposing expert testimony, were only sometimes effective. A scientifically informed cross-examination helped jurors recognize problems with a study with a missing control group (Austin & Kovera, 2015) but not studies with poor construct validity (Kovera et al., 1999) or psychological tests with poor reliability (Chorn & Kovera, 2019). Opposing expert testimony caused
It may be unsurprising that training laypeople to evaluate the quality of research in the context of mock trial proceedings has proven difficult. After all, many psychology doctoral programs require their students to take semester-long courses in research methods in the hope that they will master the material required to make sound judgments. In my own practice as an expert witness on eyewitness identification accuracy, I have tried to take lessons from research on what makes experts more informative. I use simple language and avoid jargon—or at least explain it when I do use it. In a recent case, I even persuaded a judge to allow me to use a PowerPoint presentation to explain how the base rate of guilty suspects in lineups affects the prevalence of mistaken identifications, recognizing that most people have difficulty understanding the importance of base rate information. If we want legal decision-makers to benefit from our psychological science, we must continue to study ways to make it more digestible for them.

References


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


