ON THE SUPPOSED EXPERTISE OF JUDGES IN EVALUATING EVIDENCE

BARBARA A. SPELLMAN[†]

In response to Frederick Schauer, *On the Supposed Jury-Dependence of Evidence Law*, 155 U. PA. L. REV. 165 (2006).

From way down in the dirty depths, where those of us who collect empirical data dwell unobserved and largely ignored by most legal academics,[1] it is refreshing to hear a call for "more data" from legal scholars such as Fred Schauer.[2] In asking whether it makes sense to follow the existing trend of discarding much of evidence law when judges—rather than juries—are the fact-finders at trial, he notes that the empirical literature involving judges' reasoning is sparse and the literature comparing judges with jurors is even sparser.[3] In order to determine whether judges are better than jurors at weighting[4] evidence and fact-finding, as many appear to believe,[5] he wishes for more and more focused research.[6]

[†] Associate Professor of Psychology, University of Virginia. Ph.D. 1993, University of California, Los Angeles. J.D. 1982, New York University School of Law. B.A. 1979 Wesleyan University.

[1] However, the times might be "a-changin'." Witness the new *Journal of Empirical Legal Studies*, its associated annual conference, and the increasing number of J.D./Ph.D.'s on the faculties of major law schools. Schauer, himself, seems to have turned the corner; see his call for the greater use of data in Frederick Schauer, *Foreword: The Court's Agenda—And the Nation's*, 120 HARV. L. REV. 4, 14 & n.31 (2006).

[2] *See* Frederick Schauer, *On the Supposed Jury-Dependence of Evidence Law*, 155 U. PA. L. REV. 165, 187-189 (2006) (expressing the need for more experiments on judge and jury behavior).

[3] *Id.*

[4] I use the term "weighting" evidence to indicate the consideration of how much value to place on an individual piece of evidence. I use the term "weighing" evidence to indicate the balancing of evidence from both sides when trying to reach a conclusion.

[5] *See* Schauer, *supra* note 2, at 188 (noting the popular belief). *But see* Paul H. Robinson & Barbara A. Spellman, *Sentencing Decisions: Matching the Decisionmaker to the Decision Nature*, 105 COLUM. L. REV. 1124, 1138-46 (2005) (arguing that given the un-

However, before I drag my computer, my research assistants, and my cognitive psychologist's bag of "heuristic and bias tricks" and "memory illusions" to the next judges' conference, I would want to ask myself several important questions that are all subsets of this one: Why would anyone expect judges to be different from jurors at evaluating evidence? Considering possible answers to this question is critical before beginning research because there are an infinite number of experiments that one might run comparing judges and jurors, but there is neither an infinite amount of time nor of judges' goodwill. Developing hypotheses will guide the selection of the experiments and, later, the generalization of the experimental results to situations broader than the specific experimental stimuli.

One subset of questions has to do with the task: What is it that people are asked to do mentally with evidence presented in the courtroom? The second subset has to do with comparing judges and jurors (or juries): How do they differ? It is only after considering those two types of questions—about the tasks and the reasoners—that we can offer testable hypotheses about whether judges and jurors are likely to differ on the tasks and, if so, when, why, and how. And it is only then that we should design our studies to investigate those hypotheses. This Response specifically focuses on whether judges have any expertise that might suggest that they should, in fact, be better at weighting evidence and, ultimately, at fact-finding, than jurors.

I. THE TASK

What is it that people believe judges might be able to do that jurors cannot? It is not the general task of weighting evidence appropriately. In their role as fact-finders, jurors are entrusted with that task, including, for example, weighting the testimony of less-than-credible witnesses, the value of circumstantial evidence and alibis, etc. But there are specific types of relevant evidence kept away from jurors[7] because it might be misused by them—either overweighted

certainty as to who is a *better* fact-finder, juries might well be the *preferred* fact-finders for policy reasons).

[6] Schauer, *supra* note 2, at 189.

[7] As Schauer points out, there are types of evidence that are excluded in court for extrinsic and non-epistemic policy reasons (e.g., privileged communications and evidence of subsequent remedial measures), and such evidence should be denied to judges as well as juries. *Id.* at 167-68. What about evidence that is disallowed based on the notion that juries would not be able to handle it properly? It seems to only make sense to permit judges to hear such evidence if it would either (a) result in better fact-

(whether supporting a correct or incorrect conclusion) or used in ways that the rules of evidence deem inappropriate or unfair (e.g., character evidence or evidence of prior bad acts).[8]

Thus, there is evidence that jurors are not supposed to hear at all (rather than hear and give appropriate weight). And there are many experiments demonstrating that when mock jurors[9] "accidentally" hear some of that forbidden information, and are told that they should disregard it and not allow it to affect their judgments, they often do not do so.[10] There are two distinctions that should be made among the explanations for why studies often show jurors' failure to disregard. The first is whether jurors *cannot* disregard or *will not* disregard—that is, whether they fail because they are unable or unwilling to follow the instructions. The second is whether the failure is due to cognitive or social psychological factors.[11] Social factors are those that involve external influences from other people on the reasoner; cognitive factors are internal to the reasoner. Table 1 (below) diagrams these factors. Consider the social/unwilling box. Psychological reactance theory states that people whose thoughts or behaviors are constrained by an external force will be motivated to reestablish their own freedom and act as if they were unconstrained. Thus, "reactance" suggests that jurors would be unwilling to disregard evidence precisely

---

finding, or (b) simplify the trial process—for example, by eliminating various pre-trial motions—without resulting in worse fact-finding.

[8] *Id.* at 178-79.

[9] "Mock jurors" are experimental participants asked to pretend to act as jurors in a research study. They are often college students but, increasingly, are a cross-section of jury-eligible adults from a community. Note that undergraduates and jury-eligible adults tend to perform similarly in these studies. *See* Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?* 23 L. & HUM. BEHAV. 75, 77-80 (1999). The experimental procedures in mock juror studies typically involve reading and responding to a short scenario or trial transcript or watching a videotape or trial enactment.

[10] For a recent review of the psychology research, see Nancy Steblay et al., *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: A Meta-Analysis*, 30 L. & HUM. BEHAV. 469 (2006).

[11] Molly J. Walker Wilson, Barbara A. Spellman, and Rachel M. York, Beyond Instructions to Disregard: When Objections Backfire and Interruptions Distract 4-8 (March 18, 2006) (unpublished manuscript, on file with the University of Pennsylvania Law Review); Molly. J. Walker Wilson & Barbara A. Spellman, Objection! The Unintended Consequences of Attorney Interruptions, Presentation to the American Psychology-Law Society, slide 2 (Mar. 5, 2004). A more common distinction is drawn between motivational and cognitive factors. *See, e.g.*, Steblay et al., *supra* note 10, at 487-88 (concluding that both jurors' motivation and information processing capacities may compromise their ability to disregard inadmissible evidence); Andrew J. Wistrich et al., *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 U. PA. L. REV. 1251, 1260-70 (2005) (same).

because a judge tells them that they *must* do so. "Attempt to hide" refers to the hypothesis that mock jurors believe that the fact that someone is trying to suppress information makes that information more valuable.[12] Although each of these factors might contribute to explaining why jurors might be unwilling to disregard information, they ought not be relevant to judges' failure to do so. The cognitive/unwilling box represents the idea that jurors might believe that the contested evidence is true and probative and that failing to use it would result in an unfair verdict. It is possible that judges might be more vigilant about this influence than jurors.[13] In the social/unable box, "ironic processes" refers to the finding that when people try to avoid thinking about something for a while (e.g., are told not to think of a white bear), at some later point thoughts about that suppressed object are likely to come back very strongly.[14] In the cognitive/unable box, "belief perseverance" refers to the finding that once people learn something they often continue to act and make judgments as if they believe it, even after learning that it is not true. And "story coherence" refers to the finding that when people learn information that helps make sense of other information—for example, motives, goals, or missing facts—the information that pulls other things together becomes particularly hard to forget.

Table 1: Taxonomy of Explanations for the Failure to Disregard

	Social	**Cognitive**
Unwilling	Reactance Attempt to hide	It's true (accurate/just verdict)
Unable	Ironic Processes	Belief Perseverance Story Coherence

II. DIFFERENCES BETWEEN JUDGES AND JURIES (OR JURORS)

Do judges differ from jurors in ability to disregard evidence? The few extant studies suggest not: in laboratory-type studies judges, just

---

[12] Wilson et al., *supra* note 11, at 15, 30.

[13] However, it is unclear whether such vigilance would show up in results. Often when people are conscious of possible biasing influences and try to correct for them, they end up over- or under-correcting.

[14] *See generally* Daniel M. Wegner, *Ironic Processes of Mental Control*, 101 PSYCHOL. REV. 34, 41-42 (1994) (explaining the ironic effects of thought suppression).

like mock jurors, fail to fully disregard evidence when instructed to do so.[15] But given the theories of why jurors don't disregard why would we expect that judges would? That is, in what relevant ways might judges differ from jurors?

The answers fall into three categories.[16] First, there might be relevant "individual differences": judges might differ from jurors *a priori* in relevant ways such as being smarter or less emotional. Second, institutional roles may matter: judges are elected or appointed whereas jurors are drafted; plus judges are "repeat players"[17] and have more accountability for their decisions. Finally, expertise may play a role: judges may differ from jurors due to their training in law, to their previous experience of judging, or both.[18]

That judges might be "smarter" than jurors, for example, have higher IQs or score higher on the "Need for Cognition Scale,"[19] might very well make them better at fact-finding given particular evidence but not necessarily better at disregarding evidence. Good fact-finding may require being able to consciously question the coherence of evi-

---

[15] *See* Wistrich et al., *supra* note 11, at 1323 ("Judges are indeed human; like jurors, they are often unable to 'close the [v]alves of [their] attention.'") (citation omitted).

[16] Schauer suggests that commonly believed advantages for judges include "possibly because judges are smarter, possibly because they are better educated, possibly because of their greater experience in hearing testimony and finding facts, and almost certainly because of their legal training and legal role-internalization." Schauer, *supra* note 2, at 188. Other authors have parsed the advantages differently. *See, e.g.*, Jennifer K. Robbennolt, *Evaluating Juries by Comparison to Judges: A Benchmark for Judging?*, 32 FLA. ST. U. L. REV. 469, 503-04 (2005) (identifying "training and expertise in the law," repeat player status, "better understanding of the law," "better memory for [] trial evidence," and being "better at identifying information presented at trial" as relevant factors); Wistrich et al., *supra* note 11, at 1277 (characterizing judges as potentially having better education, "legal training," a greater understanding of the "purpose, importance, and desirability of the exclusionary rules," and more experience making legal decisions).

[17] Gregory Mitchell, *Mapping Evidence Law*, 2003 MICH. ST. L. REV. 1065, 1122 (2003).

[18] Of course, another difference is that judges act individually whereas jurors act in groups; that distinction might create disparities in ability to disregard or appropriately weight evidence. *See* Kamala London & Narina Nunez, *The Effect of Jury Deliberations on Jurors' Propensity to Disregard Inadmissible Evidence*, 85 J. APPLIED PSYCHOL. 932, 938 (2000) (finding that mock jurors who deliberate are better at disregarding inadmissible information than those who do not). *See also* Robinson & Spellman, *supra* note 5, at 1143-45 (noting that groups are generally better fact-finders for problems where a "'right answer'" exists and suggesting that the larger information base possessed by a group may also assist juries in reaching a proper verdict).

[19] The Need for Cognition Scale is a measure of people's enjoyment of and motivation for effortful thinking. *See generally* John T. Cacioppo et al., *Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition*, 119 PSYCHOL. BULL. 197 (1996) (describing and reviewing the measure).

dence—a task for which cognitive skill or effort matters.[20] However, the infiltration of inadmissible evidence into the decision system might be, at least in part, automatic and not amenable to conscious manipulation and, therefore, not likely to be sensitive to cognitive skill or effort.[21] The institutional roles that judges play seem likely to affect their unwillingness to disregard evidence, but not the ability to do so. That is, by being an integral part of the legal system, and understanding and willing to comply with it, they might not react against rules that don't allow them to consider particular evidence, overreact to people who try to hide evidence from them, or decide that someone is guilty when the state needs inadmissible evidence to prove its case. But it is the third reason—that judges are "experts" and jurors are not—that underlies most arguments about why judges should be better at this task.[22] Thus, it is this issue—whether judges are experts, in what, and by virtue of what—that is my next and principal focus.

III. THE QUESTION OF JUDICIAL EXPERTISE

What *is* an expert? An expert is someone who performs consistently and reliably better on representative tasks than the overwhelming majority of other people.[23] What *makes* an expert? There are two types of answers to this question. One has to do with the qualities, training, and experience needed to develop expertise. The other has to do with the range of competences and abilities that we expect experts to demonstrate.

There is no good reason to conclude that, by virtue of qualities, training, or experience, trial judges should be considered experts at

---

[20] *See* Deanna Kuhn et al., *How Well Do Jurors Reason? Competence Dimensions of Individual Variation in a Juror Reasoning Task*, 5 PSYCHOL. SCI. 289, 295 (1994) (demonstrating that more sophisticated reasoners are better able to revise their opinions about a case in light of new facts).

[21] However, there is a suggestive finding in the deductive reasoning literature: people with higher IQs are less likely to be influenced by the content of syllogisms when evaluating their validity (which, of course, should be based purely on the syllogism's structure). *See* Keith E. Stanovich & Richard F. West, *Individual Differences in Reasoning: Implications for the Rationality Debate?*, 23 BEHAV. & BRAIN SCI. 645, 662-63 (2000) (noting that individuals with high analytic intelligence are more able to decontextualize analytical problems).

[22] *See supra* note 16 (listing judicial advantages). Although many of those authors suggest that judges might be better because they have more "training" or "experience," they finesse how those factors would create the presumed judicial expertise.

[23] Thus, although you, the reader, are no doubt expert at reading, you are unlikely to ever be called an expert reader.

weighting evidence or at fact-finding. Experts are not just "smarter" than non-experts.[24] Nor are large amounts of experience alone—such as thousands of hands of bridge or hundreds of rounds of golf—sufficient for developing expertise.[25] Rather, expertise develops out of "many thousands of hours of specific types of practice and training"—a process called "deliberate practice." Deliberate practice requires focused programmatic study. It includes appropriate feedback about performance. It includes identifying errors and working on procedures to eliminate them.[26] One might consider that law school training involves these qualities and so all lawyers (and even more so lawyers and judges working in the appellate system) would develop expertise in analyzing cases. That may be true, but analyzing cases—reading (truncated) text, considering an already-digested fact pattern, evaluating the justifications for a holding, looking for the real justification behind the stated ones, and evaluating a rule or principle in light of possible implications or future applications—is quite different from weighting evidence and finding facts. For the latter two tasks there is no prolonged training with feedback in law school; law students do not listen to trials as they unfold and learn to integrate—or not integrate—admissible and inadmissible evidence. And trial judges can sit through hundreds of cases and never do the focused study or have the fast reliable feedback necessary for developing expertise. With the exception of bench trials, it is not the trial judge's job to weight and evaluate evidence; that is, they need not "practice." Further, reliable feedback as to the appropriate weighting of evidence does not exist—the ground truth of cases is never known.[27]

---

[24] *See* K. Anders Ericsson, *An Introduction to* Cambridge Handbook of Expertise and Expert Performance*: Its Development, Organization, and Content*, *in* THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE 3, 10 (K. Anders Ericsson et al., eds., 2006) ("[M]easures of basic mental capacities are not valid predictors of attainment of expert performance in a domain . . . .") (citation omitted)).

[25] *See* K. Anders Ericsson, *The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance*, *in* THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE, supra note 24, at 683, 691 (noting that job performance is weakly correlated with experience for long-term employees) (citation omitted)).

[26] *Id.* at 691-700. *See also* John Horn & Hiromi Masunaga, *A Merging Theory of Expertise and Intelligence*, *in* THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE, *supra* note 24, at 587, 601 (describing the structured practice that is necessary to obtain expert performance).

[27] Worse yet, there is a sort of self-perpetuating reassurance built into the system: a judge might believe that she had done a good job weighting the evidence because the verdict seemed consistent with all of the evidence. But her weighting causes the verdict. This situation is reminiscent of that of police officers who believe that the confessions they elicit must be true because the suspect ends up being convicted. Such "feedback" is not reliable because the confession will, in part, cause the verdict.

And, even if trial judges were experts in something involving listening to and evaluating evidence, it is far from clear that being an expert would include the ability to disregard evidence appropriately. Experts excel because they "can detect and see [patterns of information] that novices cannot;" they are better at finding and selecting information to use; and they typically expend less cognitive effort at tasks (because they do more of the task "with greater automaticity").[28] Although these characteristics sound good, experts' prior knowledge and automated thinking can cause inflexibility.[29] Thus, it is also possible that once an expert is exposed to evidence that is then quickly integrated into her conception of the case, it might be even more difficult for her to exorcise that knowledge and its consequences.

IV. FINAL THOUGHTS

Of course, most of the above is stated in terms of disregarding evidence, not weighting evidence, and the latter may seem to be the more relevant (to Schauer's argument) and more common (in the courtroom) task. Disregarding, however, is just an extreme form of weighting: it is setting the weight of evidence to zero. It is also an important form of weighting in that even if most rules of evidence were relaxed for judges sitting as fact-finders, those judges might still have to rule on the admissibility of some evidence (e.g., illegally obtained evidence) and set the weight of that evidence at zero for themselves.[30] It is cleaner to study disregarding than weighting in general because there is a normative standard that can be used to determine whether the former was done: people exposed to the to-be-disregarded evidence should make judgments exactly like people who have not been exposed to that evidence at all. Research regarding the appropriate weighting of admissible evidence is trickier: we can tell whether two groups of people have differentially weighted a piece of evidence but we cannot tell who did it right.[31]

---

[28] Michelene T. H. Chi, *Two Approaches to the Study of Experts' Characteristics*, *in* THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE, *supra* note 24, at 21, 23-24.

[29] *See id.* at 26 (noting that experts have trouble adapting to changes in the structure of the domain of their expertise).

[30] *But cf.* Jennifer L. Mnookin, *Bifurcation and the Law of Evidence*, 155 U. PA. L. REV. PENNUMBRA 134, 136 (2006) (questioning a judge's ability to fully set the evidentiary value of an excluded piece of evidence at zero).

[31] The question of "who got it right" also emerges when studying how real judges and juries might differ in the disposition of actual cases, such as those in which a jury

Based on the analysis of judges' (non-)expertise, it is my guess that after more research is in, judges will not prove to be consistently better than jurors at disregarding evidence. However, that is just a hypothesis, and before I send some lucky research assistants off to Scottsdale or South Beach (or wherever judges hold conferences) to investigate the differences in laboratory-type studies, I would try to (a) create materials that could be used with both judges and mock jurors and (b) create materials that would isolate different explanations for the failure to disregard. Judges might be better than jurors at disregarding evidence when the reasons involve being unwilling to do so (e.g., are motivational) because of their roles in the legal system. However, nothing in their purported "expertise" suggests that they should be better when their reasons involve being unable to do so.

In the various lists comparing and contrasting the strengths and weakness of judges, jurors, and juries, one important factor is often forgotten: all are human. From earliest infancy, the human cognitive system is a sophisticated tool for detecting patterns, seeing relations, imagining causes, and creating coherent stories (even if sometimes they do not exist). In that we all could be called "experts." It is difficult to envision how a mere desire, or an admonition, to stop thinking like a human being could be effective. For anyone.

---

sat and the presiding judge later gave her opinion as to which was the correct verdict. When verdicts differ, as they will, we have no independent way of knowing which verdict is correct. *See* Robbennolt, *supra* note 16, at 502-03 (discussing the difficulty of normatively benchmarking decisions made by judges against those made by juries).