On The Supposed Expertise of Judges in Evaluating Evidence
Barbara A. Spellman


From way down in the dirty depths, where those of us who collect empirical data dwell unobserved and largely ignored by most legal academics, it is refreshing to hear a call for “more data” from legal scholars such as Fred Schauer. In asking whether it makes sense to follow the existing trend of disposing of 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. In order to determine whether judges are as much better than jurors at weighting evidence and fact finding as many appear to believe, he wishes for more and more focused research.

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


1 Although 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 Harvard Law Review, 4, 14, and footnote 31.


3 Id. at 189-192.


5 Schauer at 189.
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. Such hypotheses will guide in the selection of the experiments and, later, in the generalization of the experimental results to situations beyond the specific experiment stimuli used.

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 comment specifically focuses on whether judges have any expertise that might suggest that they should, in fact, be better at weighting\(^6\) evidence and, ultimately, at fact finding, than jurors.

The Task

What is it that people believe that 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 (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

---

\(^6\) 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 on both sides when trying to reach a conclusion.

\(^7\) As Schauer points out at 167-68 there are types of evidence that are disallowed 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. What about evidence that is disallowed based on the notion that juries wouldn’t be able to handle it properly? It only makes sense to permit judges to hear such evidence if it would (a) result in better fact finding or (b) simplify the trial process – for example, by eliminating various pre-trial motions -- without resulting in worse fact finding.
Thus, there is evidence that jurors are not supposed to hear at all (rather than hear and give appropriate weight). And there are many psychology laboratory studies 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. \(^{11}\) The first is whether jurors cannot disregard or will not disregard – that is, whether they fail because they are unwilling or unable to follow the instructions. The second is whether the failure is due to cognitive or social psychological factors. Social factors are those that involve external influences from other people on the reasoner; cognitive factors are all internal to the reasoner. Table 1 diagrams those 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 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 hide information (e.g.,

---

\(^8\) Id. at 179.

\(^9\) “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 college students and jury-eligible adults tend to perform similarly in these studies. Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?* 23 Law & Human Behavior 75, 77-80 (1999). The experimental procedures may involve reading and responding to a short scenario or trial transcript or watching a videotape or trial enactment.


by objecting to its use) makes that information more valuable.\(^{12}\) Although both of these might contribute to explaining why jurors might be unwilling to disregard information, they ought not be relevant to judges’ failure to do so. In the cognitive/unwillingness box is the point that jurors might believe that the contested evidence is true 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 suppress thinking about something for a while (e.g., are told not 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 (e.g., 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, or goals, or missing facts – the information that pulls other things together becomes particularly hard to forget.

<table>
<thead>
<tr>
<th></th>
<th>Social</th>
<th>Cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unwilling</td>
<td>Reactance</td>
<td>It’s true</td>
</tr>
<tr>
<td></td>
<td>Attempt to hide</td>
<td>(accurate/just verdict)</td>
</tr>
<tr>
<td>Unable</td>
<td>Ironic Processes</td>
<td>Belief Perseverance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Story Coherence</td>
</tr>
</tbody>
</table>

**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 like mock jurors, fail to fully disregard evidence when instructed to do so.\(^{15}\) But given the theories of why jurors don’t why would we expect that judges would? That is, in what relevant ways might judges differ from jurors?

\(^{12}\) Molly J. Walker Wilson, Barbara A. Spellman, and Rachel M. York at p. ___.

\(^{13}\) But it’s not clear that 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.


The answers fall into three categories: 16 (1) “individual differences” – judges might differ from jurors a priori in ways that might be relevant (e.g., they might be smarter or less emotional); (2) institutional roles – judges are elected or appointed whereas jurors are drafted; judges are “repeat players” and have more accountability for their decisions 17; and (3) “expertise” -- judges may differ from jurors due to their training in law or 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” 19 scale 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 evidence – a task for which cognitive skill or effort matters. 20 However, the infiltration of inadmissible evidence into the belief system

\[\text{Schauer, at 188, suggests these as commonly believed advantages for judges: “possibly because judges are smarter, possibly because they are better educated, possibly because of their greater experience in hearing testimony and finding facts, and almost certainly because of their legal training and legal-role internalization”}.\]

\[\text{Other authors have parsed the advantages differently. Jennifer K. Robbennolt,} \]


\[\text{Of course another difference is that judges act individually and jurors act in groups and that might create a difference in ability to disregard or appropriately weight evidence. See Kamala London and Narina Nunez, The Effect of Jury Deliberations on Jurors’ Propensity to Disregard Inadmissible Evidence, 85 Journal of Applied Psychology, 932 (2000) (finding that mock jurors who deliberate are better at disregarding inadmissible information than those who do not). See Robinson and Spellman, id at 1143-44, for a discussion on whether individuals or groups are likely to be better fact finders.}\]

\[\text{The Need for Cognition scale is a measure of people’s enjoyment of and motivation for effortful thinking. See John T. Cacioppo et al., Disposition Difference in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition, 199 Psychological Bulletin 197 (1996) (for a description and review of the measure).}\]

\[\text{Deanna Kuhn et al., How Well Do Jurors Reason? Competence Dimensions of Individual Variation in a Juror Reasoning Task, 5 Psychological Science, 289, 295 (1994) (showing that}\]
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. The institutional roles that judges play seem likely to affect their unwillingness to disregard evidence (i.e., eliminating the reasons across the top row) 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, over-react 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. Thus, it is this issue -- whether judges are experts, in what, and by virtue of what -- that is my next and principal focus.

The Question of Judicial Expertise

What is an expert? An expert is someone who performs consistently and reliably better on representative tasks (from the domain of expertise) than the overwhelming majority of other people. 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 weighting evidence or at fact finding. Experts are more sophisticated reasoners were better able to revise their opinions about a case in light of new facts).

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). Keith E. Stanovich and Richard F. West, Individual Differences in Reasoning: Implications for the Rationality Debate? Behavioral and Brain Sciences 645 (2000).

See footnote 16 (eds: the one after “three categories”). 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.

Thus, although you, the reader, are no doubt expert at reading, you are unlikely to ever be called an expert reader.
not just “smarter” overall than non-experts.\(^{24}\) Nor are large amounts of experience alone (e.g., thousands of hands of bridge, 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 a different thing from weighting evidence and finding facts. For the latter two 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 and never have the fast reliable feedback necessary for developing expertise. In part that is because that except for bench trials, it is not the trial judge’s job to weight and evaluate evidence (that is, they need not “practice”), and in part that is because, in fact, reliable feedback as to the appropriate weighting of evidence does not exist – the ground truth of cases is never known.\(^{27}\)


\(^{27}\) 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. Among the “Ways in Which Experts Excel”,\(^\text{28}\) Michelene T. H. Chi lists: that they can detect and see patterns of information that non-experts do not, that they are better in finding and selecting information to use, and that they typically expend less cognitive effort at tasks (because they do more of the task “automatically”). Although those all sound good, experts’ prior knowledge and automated thinking can cause inflexibility.\(^\text{29}\) 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.

**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. 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.\(^\text{30}\)

---


\(^\text{29}\) Id. at 24-27 (see “Ways in Which Experts Fall Short” numbers 5 and 7).

\(^\text{30}\) The question of “who got it right” also emerges when studying how real judges and juries might differ in the disposition of actual cases (e.g., one in which a jury sat and the presiding
Based on the analysis of judges’ (non-) expertise, it is my guess that after more research is in, judges will not prove to be across-the-board 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 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.

judge later gives 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 Jennifer K. Robbennolt, *Jury Decisionmaking: Evaluating Juries by Comparison to Judges: A Benchmark for Judging*, 32 Fla. St. U. L. Rev. 469 (2005) (for a discussion).