What Do the Gatekeepers See? Perceptions and Evaluations of Scientific Evidence Among State Court Judges

Esther Nir, Siyu Liu

\( ^a \) New Jersey City University
\( ^b \) Penn State Harrisburg

ABSTRACT AND ARTICLE INFORMATION

Judicial gatekeeping decisions impact criminal case processing (e.g., verdicts, pleas) in significant ways. Despite this influential role that judges play, little is known about the key factors that contribute to judicial perceptions of evidence and the evaluative processes that judges employ in making evidentiary rulings. Using qualitative interviews with 41 state court judges presiding in a U.S. state employing the Frye standard, we explore judicial perceptions of a variety of scientific evidentiary forms and the processes by which judges reach reliability and credibility determinations in performing their gatekeeping duties. Our analyses reveal that judges are highly influenced by the general reputation of forensic evidence, often lack understanding of the scientific methodologies involved, and tend to focus on factors external to reliability (e.g., expert’s communication skills and showmanship, weight of the evidence) in their evaluative processes. Further, credibility assessments of witnesses are fraught with subjective interpretations, potentially leading to disparate evidentiary rulings.

Article History:

Received January 15th, 2021
Received in revised form March 16th, 2021
Accepted March 16th, 2021

Keywords:
evidence, judicial discretion, forensic, law, procedural justice, qualitative methods, courts

© 2021 Criminology, Criminal Justice, Law & Society and The Western Society of Criminology
Hosting by Scholastica. All rights reserved.

Corresponding author: Esther Nir, Department of Criminal Justice, New Jersey City University – Jersey City, Professional Studies Building, Room 226, 2039 John F. Kennedy Boulevard, Jersey City, NJ, 07305, USA.
Email: enir@njcu.edu
“Evidence is a currency in which legal cases are transacted” (Cooney, 1994, p. 851).

The judicial role as “gatekeeper of the evidence” is a prominent and critical decision-making responsibility that is highly consequential to case outcomes (Daubert v. Merrell Dow Pharmaceuticals, 1993; Schweitzer & Saks, 2009). Indeed, judges’ decisions to admit or reject pieces of evidence at trial may contribute to the establishment of the defendant’s guilt, while rulings to allow favorable defense evidence or exclude damaging prosecution evidence may lead to an acquittal. Judicial gatekeeping decisions may also influence plea offers by the prosecution, as well as defense decisions to accept or reject offers (Kramer et al., 2007; Rossman et al., 1980); specifically, plea negotiations are often guided by prosecution and defense perceptions of the likelihood of prevailing at trial (based, in part, on their perceptions of the strength of the admissible evidence), sometimes referred to as “bargaining in the shadow of the trial” (Bushway et al., 2014; Dezember & Redlich, 2019; Kramer & Ulmer 2002; Landes, 1971; Redlich et al., 2016; Yan, 2019). In fact, mere projections that a particular judge may exclude or admit a specific piece of evidence (e.g., due to reliability or constitutional challenges) could profoundly impact the manner by which prosecutors and defense attorneys exercise their discretion in resolving cases within the court community (Nardulli, 1986; Nir & Liu, in press; Wilson, 2016). Lastly, the admission of faulty evidence has been linked with the occurrence of wrongful convictions (Derwin, 2018).

Beyond gatekeeping functions, judicial perceptions of evidence may influence other stages of criminal case processing. For example, in bench trials, judges play the role of triers of fact and determine whether the defendant is guilty based on their perceptions of the reliability, credibility, and overall strength of the admissible evidence (Blanck, 1992). Further, despite the theoretical understanding that evidence used to establish guilt should have little bearing in sentencing (since questions of guilt have been resolved by this point of case processing), recent empirical studies establish that judicial perceptions of evidentiary strength (as manifested by type and quantity) are significantly related to the length of custodial sentences judges impose (Nir & Griffiths, 2018, 2019; Peterson et al., 2010, 2013) and that judicial confidence level in guilt (based on perceptions of evidentiary strength) partially guides sentencing decisions (Nir & Griffiths, 2019).

The documented influence of evidentiary considerations in driving case outcomes underlines the determinative force of admissible evidence in criminal case processing. Among the various types of evidence, judges are uniquely challenged by the task of evaluating scientific evidence. As Beecher-Monas (1999) stated, “judges, traditionally triers of law, occasionally pressed into service as triers of fact, now must also be triers of science in cases where experts proffer scientific evidence” (p. 1047). These challenges prompt important questions: Do judges have sufficient knowledge and guidance to perform their gatekeeping function effectively? How do judges perceive the strength and reliability of different types of scientific evidence? What evaluative processes do judges employ in assessing the reliability of various forms of scientific evidence? Gaining insight on judicial perceptions of evidence and evaluative processes in determining admissibility is a crucial step to better understanding criminal case processing as a whole. Yet, despite the imperative role that judicial perceptions and evaluations of evidence play, only a handful of existing empirical studies specifically focus on this topic (Brodsky et al., 2010; Jurs, 2009; O’Brien, 2018; for studies using non-American data, see Canela et al., 2019; Connolly & Gordon, 2011; Moulin et al., 2018; Porter & Brinke, 2009). Using in-depth qualitative interviews of 41 state court judges presiding in a state in the United States that employs the Frye standard, we seek to help fill this gap by exploring judicial knowledge, perceptions, and opinions of the most common forms of scientific evidence used in criminal cases and the processes by which judges analyze evidentiary reliability, credibility, and value in performing their gatekeeping functions.

**Literature Review**

**Types of Evidence**

A broad range and scope of evidence may be introduced during criminal proceedings (Grimm, 2018) and can be classified in a variety of ways (Findley, 2013; Greenstein, 2009; Heller, 2006; Thompson, 2012). In this article, we discuss specific scientific evidentiary forms as points of focus to gain insight into the manner by which judges approach the task of evaluating the admissibility of scientific evidence, as well as their general perceptions of the evidence evaluated. We discuss DNA, latent prints, ballistic evidence, and chemist evidence, commonly introduced in criminal cases (for details, see Table 1). Our respondents most commonly referenced these types of evidence during our interviews, and they have been tasked with reaching admissibility decisions on
Judicial Gatekeeping Decisions

**The Standards and Processes of Gatekeeping**

While jurors consider evidentiary weight in a case, judges are responsible for deciding the admissibility of evidence (Faigman et al., 2016); evidence determined to be inadmissible by the trial court can be challenged on appeal. The gatekeeping process is an important component of judicial decision-making, as the judge acts as a filter to determine the admissibility of evidence. This process is designed to ensure that only relevant and reliable evidence is considered by the trier of fact.

**Table 1: Evidentiary Types in the Current Study**

<table>
<thead>
<tr>
<th>Type</th>
<th>Witness</th>
<th>Description</th>
<th>Method of Comparison</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Expert</td>
<td>Considered &quot;the most discriminating&quot; evidence, deoxyribonucleic acid (DNA) testing facilitates identifications of individuals; it is contained in body fluids, skin samples, bones, and hair (FBI, 2019).</td>
<td>By comparing the analytical result from the evidence sample to reference samples collected from individuals with known identities through polymerase chain reaction (PCR) Short Tandem repeat (STR) technology on matching 20 loci (FBI, 2020).</td>
<td>Forensic evidence garners a great deal of confidence from the general public and the courts. In fact, &quot;the trust that is laid upon the forensic sciences generally falls somewhere between uncritical faith and manufactured myth&quot; (Saks &amp; Faigman, 2008, p. 150; see also Edmond &amp; Roque, 2012).</td>
<td>Research shows that forensic evidence often includes speculative opinions, questionable/absent scientific techniques, and unreliable evaluative practices by technicians (Cole, 2009; Dror, 2013; Dror et al., 2011; Edmond &amp; Roque, 2012; Saks &amp; Faigman, 2008).</td>
</tr>
<tr>
<td>Latent print</td>
<td>Expert</td>
<td>Also called friction ridge examination, &quot;a latent print is an impression that can be left when a person touches an object with the unique friction skin of their hands or feet&quot; (FBI, 2019, p. 61). &quot;These prints consist of a combination of different chemicals that originate from natural secretions, blood, and contaminants&quot; (Trozzi et al., 2000, p. 1).</td>
<td>Depending on the type and condition of the surface where the print was retrieved and the type of residue creating the print, procedures may vary, but with visual and fluorescent examinations, and possible aid of powder and enhancing agents (Trozzi et al., 2000), the examiner can compare the print from the scene to reference prints from known individuals.</td>
<td>Judges are generally very receptive to expert opinions presented by the prosecution on behalf of the state (Edmond &amp; Roque, 2012).</td>
<td>Cole (2009) suggests that forensic identification evidence is actually incapable of accomplishing its goal of definitively establishing identifications, and that claims of individualization cannot be supported by the assumption of discernible uniqueness (Cole, 2009).</td>
</tr>
<tr>
<td>Ballistic evidence</td>
<td>Expert</td>
<td>This refers to any evidence related to the functionality and physical characteristics of suspect firearm or firearm-related items (FBI, 2019).</td>
<td>&quot;The observed physical characteristics of the questioned items are compared to reference firearms or other information sources to determine if they are consistent&quot; (FBI, 2019, p. 47). Often, using &quot;a traditional microscopic comparison of the actual evidence items,&quot; the digital image of a test-fired cartridge case/shotshell casing or evidence cartridge case/shotshell casing can be compared with records in the database (FBI, 2019, p. 70).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemist evidence</td>
<td>Expert</td>
<td>A broad category of evidence related to any chemical substance that may be of value to build connections between individuals and the crime scene (e.g., explosives, potentially illegal substances, lubricants, metallic objects, or other unknown items; FBI, 2019).</td>
<td>Depending on the general category of the substance (e.g., liquid, powder, or stains), the sample can be compared with a suspected source. Minimally, the chemical composition and consistency could be understood to be possibly compared with a likely source (FBI, 2019).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

these forms of scientific evidence. We further explore judicial perceptions/approaches to credibility determinations applicable to scientific expert witnesses. Below, we describe the processes and legal standards that guide and direct judges in their gatekeeping tasks. We then review challenges drawn from the literature on these evaluative processes.
judge is not heard by the jury.\textsuperscript{1} During hearings or trials, expert witnesses may provide explanations on procedures and scientific contexts for forensic evidence. Before an expert witness is permitted to present their opinion in court, the judge decides whether or not the witness possesses sufficient expertise on the subject matter in question. In contrast to lay witnesses, whose admissibility could rely largely on judicial common sense (Burns, 2016) and legal training, expert witnesses must possess the knowledge, experience, education, and skill set to reliably inform the factfinder about the evidence in the relevant and currently established field of scientific or professional practice. Since judges are in the “gatekeeping” position, criteria have been established to help them reach determinations on the admissibility of expert witness testimony (Shelton, 2009).

Currently, two legal standards are widely practiced in the United States. The Frye Rule, also known as the “general acceptance” test, was first established in Frye v. United States (1923, D.C. Circuit) and requires trial judges to determine whether the scientific methods employed are generally accepted in the scientific community. The newer standard is based on Daubert v. Merrell Dow Pharmaceuticals (1993) where the Supreme Court solidified the “gatekeeper” role of the trial courts and recommended four factors to facilitate the evaluative process: whether the methods used are generally accepted within the scientific community, whether they have been published and subjected to peer review, whether the methods have been or can be tested (i.e., falsifiability), and whether there is a known error rate. The Daubert standard replaces the prior Frye standard in federal courts and is aligned with the amended Federal Rules of Evidence 702 (i.e., expert’s specialized knowledge, reliance on sufficient facts and data, application of “reliable principles and methods,” and reliable application of the methods to case facts; Federal Rules of Evidence 702, 2021; Grimm, 2018). Over the past two decades, a majority of state jurisdictions became followers of the Daubert test or similar model, and only a handful of states remain in the Frye camp (Cheng & Yoon, 2005; DeMatteo et al., 2019; see also the Frye-Mack standard in Minnesota discussed by Alter, 2017).\textsuperscript{2} Fradera and colleagues (2004) reported that judges perform well under Daubert in excluding unqualified experts yet demonstrate inconsistencies in allowing or excluding certain types of evidence (e.g., those of forensic identification).

Both the Frye and Daubert standards are not without flaws. The Frye test is a single-prong criterion that lacks the flexibility and comprehensiveness to allow for constant scientific advancement (Allen & Nafisi, 2010). Beecher-Monas (1998) argued that many types of “junk” science evidence were admitted under Frye (e.g., hair comparison, bitemark analysis, and voice spectrography) even though there is a lack of sufficient support to establish that these evidentiary types could be of identification value. On the other hand, many commentators highlight that the Daubert standard shifted the authority of gatekeeping decisions from the relevant scientific community under Frye (i.e., fingerprint examiners in cases involving fingerprint evidence) to trial judges, who do not routinely receive scientific training and thus may not be the most appropriate gatekeepers (Allen & Nafisi, 2010; Cheng & Yoon, 2005; Grimm, 2018; Shelton, 2009). Nevertheless, some evidence show that judges have a heightened awareness of “junk” science and have adopted a more careful approach to evaluating expert testimonies since Daubert (Cheng & Yoon, 2005; Kraftka et al., 2002; see Dickson, 2018, for a critique that Daubert falls short in guiding judges to rule out “junk” science, using future dangerousness evidence as an example). Further, legal scholars have also pointed out instances where judges’ tolerated lab reports involving misconduct (Giannelli, 2007, on Joyce Gilchrist, forensic chemist in Oklahoma City Police Department and her abuse of scientific evidence).

In addition to reliability determinations, evidentiary gatekeeping requires credibility evaluations by judges. In evaluating credibility, judges make a host of assessments including whether the testimony sounds credible, whether the witness has a motive to lie, the witness’s demeanor as nonverbal “leakage cues” (Ekman & Friesen, 1969; O’Regan, 2017), surrounding circumstances, the witness’s facial expressions (Brodky et al., 2010; Connolly & Gordon, 2011; Porter & Brinke, 2009; Wessel et al., 2006), and the use of “uptalk”—rising final intonation (lower confidence rating by listeners in the expert witness context, Levon & Ye, 2020).

**The Evaluative Difficulties in Gatekeeping**

The ability of judges to ably perform their gatekeeping function is fraught with hardships (Derwin, 2018). First, judges must make evidentiary rulings on subjects that are outside of their legal expertise and general knowledge base. For example, judges have little training in medicine, methodology, and statistics, yet gatekeeping decisions often rely on the application of knowledge from these fields (Canela et al., 2019; McQuiston-Surrett & Saks, 2009; Saks & Faigman, 2008; Wojcikiewicz, 2013). Even guiding standards (i.e., Daubert or Frye) can only provide limited assistance to judges; in fact, scholars have found that it remains questionable whether the implementation of either the Frye or Daubert standard results in any practical difference in how judges reach
Further, instead of following the applicable standard of review, prior scholarship suggests that when the scientific processes are difficult to understand, judges resort to what they do understand (Beecher-Monas, 1998). For example, Gatowski and her colleagues (2001) found that judges have trouble understanding the concepts of falsifiability and error rates, and attach more weight to “general acceptance.” Saks and Faigman (2008) found that in jurisdictions that apply the Daubert standard, judges have “bent over backwards to evade the application of Daubert when conscientious application would lead to the exclusion of any of the nonscience forensic sciences” (p. 163). Other factors employed by judges to evaluate expert evidence include relevance and qualification of the expert, or factors external to their expertise—including likability, believability, trustworthiness, and intelligence (Brodsky et al., 2010).

The shortfalls in judicial scientific knowledge are exacerbated by several factors. First, criminal defense attorneys are generally unable to identify methodological weaknesses or to access expert witnesses to testify on behalf of their clients (Edmond & Roque 2012; Saks & Faigman, 2008). Second, there is a vulnerability to bias in the production of scientific evidence (Butt, 2013; Dror, 2013; Dror & Cole 2010; Elwell, 2020; Kassin et al., 2013; National Academy of Science [NAS], 2009; Smalarz et al., 2016; Stevenage & Bennett, 2017), even though forensic evidence is considered to be the most influential in guilt determinations (Smith et al., 2011).

Another major challenge to gatekeeping is the troubling vulnerability of perception to judicial bias (Burns, 2016). Using a sample of civil rights cases in U. S. district courts, O’Brien (2018) found that male judges are more likely than their female counterparts to exclude expert evidence. Other factors, such as judicial experience, may also be influential (Moulin et al., 2018). Judicial bias is also observed unevenly in civil and criminal cases. While defendants in civil cases are often able to successfully challenge plaintiffs’ evidence, criminal defendants face insurmountable challenges (Dioso-Villa, 2015; Groscup et al., 2002; Rissing, 2000; Rozelle, 2007; Shniderman, 2017; see Merlino et al., 2007 for a study of admissibility decisions under Daubert, comparing those of criminal and civil cases). Commentators like Judge Donald E. Shelton (2010) attribute the failure of judges as gatekeepers in criminal cases to precedent-based rationalizations and pro-prosecution bias, resulting in the admission of sub-par science into the courtroom, regardless of validity. Echoing Shelton (2010), Damon-Moore (2017) provides a deep and scathing explanation for this judicial failure, focusing on the lack of scientific training (see also Kovera & McAuliff, 2000), the lack of defense challenges to prosecutorial evidence (see also Neal et al., 2019), reliance on heuristics (i.e., mental shortcuts relying on the fallacy of tradition; see also Lidén et al., 2019), judicial instincts about the defense and prosecution, and the unspoken alliance between the bench and the prosecution (see Nir & Liu, in press), all of which disadvantage the criminal defendant’s case.

An additional challenge to judicial gatekeeping decisions involves the subjectivity inherent in these determinations (Walsh, 1999). Given the differences in attitudes and philosophies among judges toward this role, Judge Walsh (1999) raised doubt that the result of evidence earning entry in one courtroom would be able to do so in another. Despite Judge Walsh’s concern, little empirical evidence on judicial disparity in evidentiary rulings is available.

Method

Data for this study were drawn from qualitative interviews with 41 state court judges presiding in a Frye state. Interviewees preside over felony cases; among other functions, they make evidentiary rulings on a myriad of pre-trial motions and hearings, determine the admissibility of expert testimony at trial, preside over jury trials, conduct bench trials, and sentence convicted defendants. Judicial experience among our respondents ranges from less than 2 years and over 30 years (average = 17.6 years), with 36 judges having at minimum of 9 years of experience on the bench. Moreover, the vast majority of our subjects have prior legal experience in criminal law (e.g., about 95% of our sample are former prosecutors, defense attorneys, or both). Seven judges in our sample are female, and judges preside in urban (53.7%), suburban (19.5%), or rural (26.8%) locations and are either elected (over 60%) or appointed to the bench.

After receiving approval from the Internal Review Board (IRB), data collection took place during 2014 and 2015, and include judges from counties all over the state. Efforts to recruit judges to participate in our study involved various methods (Nir, 2018), including reaching out to personal contacts (n = 8), referrals from other respondents (n = 13), cold-calling methods (n = 15), and cold calling with reference to a general email that one of our interviewees sent to fellow judges in his courthouse introducing our study (n = 5). Our interviews were conducted in judges’ chambers and ranged from 45 minutes to 3 hours (90 minutes on average). An in-depth semi-structured interview design was used to collect the data to...
“ensure that the basic lines of inquiry [were] pursued with each person” (Patton, 2002, p. 343). As this study was part of a larger project evaluating the influence of evidentiary weight on sentencing determinations, our interview guide included questions related to a myriad of topics surrounding judicial decision-making, with an emphasis on judicial perceptions of different evidentiary types and processes. Other areas of focus include judicial perceptions of sentencing disparities, sentencing factors, juries, mandatory minimums, and perceptions of judicial discretion, among others. Regarding judicial perceptions of evidence, we asked judges to discuss their perceptions of different forms of evidence, including DNA, fingerprints, chemists, ballistic evidence, and other non-scientific evidence (eyewitnesses, character witnesses, alibi witnesses), among others. Judges were probed to discuss the strengths and weaknesses of each evidentiary form and to compare different evidentiary types to one another; they were also asked to discuss their evaluative processes in assessing reliability and credibility of evidence in their own courtrooms. We utilized an inductive grounded theory approach during data collection and analysis to develop themes that “place the data into a more general or abstract framework” (Maxwell, 2005, p. 97). Observed themes include judicial confidence in various forms of forensic evidence, judicial focus on the presentation skills of expert witnesses, and concerns regarding the reliability and credibility of witnesses, among others. We compared data from each new interview with prior interviews and refined developing themes to accurately reflect the complete body of data collected (Maxwell, 2005). In this paper, we focus on the gatekeeping of scientific evidence widely used in criminal cases.

Results

Overall, our respondents value scientific evidence and perceive it to be stronger and more reliable than non-scientific evidence. As one judge noted: “Science doesn’t lie. People make mistakes. People lie. People can deceive. The science doesn’t” (Judge 34). A second judge echoed a similar sentiment: “Well, scientific evidence, most of it, either it is or it isn’t whereas witness testimony, no matter what they say, there is a possibility that they are wrong, right, half right, half wrong” (Judge 40). As a result, judges generally feel more confident adjudicating cases when the evidentiary package includes scientific evidence (Nir & Griffiths, 2019). Yet, judicial perceptions of reliability vary considerably among scientific evidentiary types, with DNA ranking the most reliable, followed by fingerprint evidence (2), ballistics evidence (3), and forensic chemistry (4). In this section, we discuss these general impressions, followed by an analysis of the evaluative processes that judges employ in assessing reliability. We then discuss prevalent judicial concerns regarding scientific evidence and judicial credibility assessments. We conclude the results with a discussion of how judges’ personal experiences influence their evaluative processes.

General Impressions, By Evidentiary Type

DNA Evidence

The vast majority of our respondents perceive DNA evidence as the strongest, most objective form of evidence available: “DNA is 100% reliable . . . . It comes back so many billion to one” (Judge 9). Among a host of positive terms used by judges to portray their perceptions of DNA evidence, descriptors include “a rock crusher” (Judge 26), “infallible” (Judge 11), “bulletproof” (Judge 34), and “the golden child” (Judge 7). Given its perceived value, most judges expressed a strong preference to have DNA as part of the evidentiary package, when it is relevant and available: “I always prefer to have certainty, and you have that with DNA evidence. For truth seeking courts, it is always best to be sure” (Judge 11). Another judge described the value of DNA in criminal case processing this way: “When you get an expert up there who says that ‘with 99.7% certainty they can rule out A, B, and C,’ you have to give it weight” (Judge 18). Consistent with existing scholarship, the majority of our respondents not only praise DNA for its evidentiary strength, but also accept the applicable science and testing procedures without question (see Saks & Faigman, 2008), a few citing its strong reputation.

Fingerprint Evidence

Though generally ranked right below DNA in evidentiary strength, most judges view fingerprint evidence as a highly probative and reliable form of evidence (see Smith et al., 2011):

Fingerprint evidence, when presented by responsible witnesses who are experienced and it is something where the jury can actually see how the prints match up and the number of points of comparison and all of that, is very strong evidence. (Judge 26)

In describing their perceptions of its strength, many judges noted that fingerprint evidence is highly instrumental in resolving identification issues (particularly in circumstantial cases), is visually approachable and understandable (to juries and judges), and usually generates considerable juror interest.
Ballistics Evidence

Ranked third of all forensic evidence, most judges perceive ballistics evidence to be a relatively reliable evidentiary form. Judges tend to divide expert ballistic testimony into two general categories—gun operability and ammunitions/weapons matching. While the vast majority of judges perceive gun operability evidence to be highly reliable and objective (a couple describing it as “kosher” [Judge 2] or “based on some physical reality as opposed to statistical analysis” [Judge 36]), judges expressed far less confidence in the scientific procedures involved in matching shell casings to guns: “In terms of gun operability, ballistics evidence is purely scientific and reliable. But if you are trying to match gyrations on shells, it is somewhat subjective” (Judge 11).

Forensic Chemistry

Judicial opinions regarding the reliability of chemist testimony run a wide gamut. On one end of the spectrum, forensic chemistry is subject to far more criticism than DNA, fingerprint, or ballistic evidence, with about a quarter of the judges arguing that it is unreliable. Interestingly, perceptions of the general reliability of forensic chemistry evidence are more focused on the work quality of the particular expert than the other three evidentiary forms:

So many of the chemists are screwy. They have had issues in a lot of the labs. They haven’t met standards. Some of the chemists have fudged the numbers. It gets a little hairy. You have no way of knowing as you sit here that this is reliable or unreliable. I have seen them effectively impeached. (Judge 5)

Presenting a very different perspective, several judges in the jurisdiction expressed confidence in the overall scientific procedures employed by chemists appearing in their courtrooms:

The chemist comes in and they talk about the presumptive test and then about the gas test and how they confirm that it is cocaine or it’s heroin or whatever. That is very routine stuff and I have never seen anybody effectively challenge one of the police lab chemists on a simple drug analysis. (Judge 27)

Common Themes

Two prevalent themes (regarding these general perceptions) warrant mention. First, judicial notions of evidentiary reliability are heavily dependent on judges’ perceptions of the general reputation of the evidence, as opposed to their own courtroom experiences. For example, despite our request that judges discuss their evaluation of the evidence based on their own experiences, judges routinely referenced the strong reputation of DNA evidence in resolving identification issues, external to their experiences on the bench. Likewise, judges connected the discussion of chemist evidence with the dated scandal-tainted reputation of the work in the jurisdiction. In fact, while judges expressed outrage at past mistakes made by chemists, virtually all of these comments were based on information judges received outside of their courtrooms, even though they were reminded to comment on their direct experience.

Second, evidentiary rankings were influenced by the (perceived) degree that the evidence relies on human evaluations, with more required human assessments leading to a decrease in judicial confidence in the evidentiary form. For example, DNA was touted as the most reliable evidentiary type due partially to judicial perceptions that DNA evidence is largely based on computer generated results and, therefore, relatively free of the perils of human assessments. In contrast, most judges expressed less confidence in fingerprint evidence than DNA, partially because it requires subjective assessments by human experts (i.e., regarding whether there is a match; see Dror, 2013, for a related discussion). As one judge stated,

the DNA goes in the computer and it’s a match by the computer and it makes the match very accurately. With fingerprints, the computer says it is a match and a human being has to look with comparison microscopes and make a determination so there is more room for human error. (Judge 9)

Notably, while several judges explained that a computer’s ability to make accurate matches far exceeds that of human experts, these same judges conceded that they do not understand the computer processes involved.

Evaluative Processes

In assessing the reliability of different evidentiary forms, judges were encouraged and afforded the opportunity to freely discuss their evaluative processes. As our analyses reveal, certain evaluative patterns emerged among judges. First, we observed that increased understanding of the scientific methods commonly utilized for a given type of evidence was associated with judges being more critical of these processes, whereas a lack of understanding was often coupled with a higher degree of acceptance. For example, judges were more focused on the scientific procedures employed in fingerprint analyses than DNA; judges explained the various points of comparison and quality of the prints, were
aware of the comparative methods employed by experts in the field and noted the susceptibility of the evidence to subjective and potentially unreliable assessments. Likewise, judges were somewhat focused on the methodologies (e.g., visual comparisons) used to match shell casings to guns in evaluating ballistics evidence; this increased understanding was accompanied by greater critiques regarding the methodologies and conclusions rendered by ballistics experts. In contrast, despite their high regard for DNA evidence, almost a quarter of the judges openly admitted that they did not fully understand how DNA evidence works or the specific methodologies involved (for similar discussions, see McQuiston-Surrett & Saks, 2009; Saks & Faigman, 2008; Wojcikiewicz, 2013); while discussions regarding the reliability of DNA evidence were filled with praise, they involved virtually no critical analysis or even reference to scientific methods.

Second, in responding to questions regarding the reliability of different forms of evidence, assessments often centered around the showmanship, communication skills, and likeability of testifying experts rather than on the scientific basis of the testimony (see Brodsky et al., 2010). For example, in response to our question “what is your perception of DNA evidence?” one judge noted: “Some explain it better than others. Some just have a little bit more flair to their testimony because it is pretty dry stuff” (Judge 34). Another judge described the importance of reaching the jury: “I think that if the expert is hyper technical it goes right over everyone’s head. The best experts are those that make it comprehensible to the ordinary juror” (Judge 9). A third judge assessed the overall communication skills and protocols of most DNA experts appearing in his courtroom:

Most of them that I have seen have been from the Office of the Chief Medical Examiner. They have very clear English-speaking witnesses that can articulate what this is, how it works and what it means. . . . They’ve developed a script that is capable of explaining this very complicated stuff to the jury in a very simple way. (Judge 19)

Notably, none of our questions prompted our respondents to comment on the expert’s communication skills or showmanship; our inquiries were focused on reliability and credibility determinations applicable to admissibility decisions. Similarly, assessments regarding fingerprint testimony often led with critiques about the presentation skills of the expert witness; a few judges openly explained that this focus is due to their lack of knowledge about the scientific processes:

There is skill in the laboratory and there is skill in testifying. I can only judge their skill in testifying. There are some people who can communicate more effectively than others, but I don’t know if that means they’re less skilled at achieving the right result. (Judge 24)

Likewise, conversations pertaining to ballistics evidence focused on the testifying skill of the expert. In general, judges were highly complementary of ballistics experts, several noting less variation in quality among ballistics experts than fingerprint experts: “The same witnesses come in all of the time. It’s almost like you press play and they say what they say” (Judge 5). A second judge stated: “They testify very well because that’s all that they do. So, I think that their testimony is clean and succinct and persuasive” (Judge 28). This focus on presentation skills was also prevalent in analyses of the reliability of forensic chemistry, where judges were focused on the inability of expert chemists to effectively communicate with the jury.

We observed a third pattern in judges’ evaluative processes. Our respondents often focused on the weight of the evidence (a task for the factfinder) rather than its admissibility (a task for the gatekeeper). For example, some judges assessed the substantive limitations of DNA in criminal trials, extrinsic to the reliability of the scientific evidence itself. Specifically, about one third of the judges explained that DNA evidence is often only one piece of the puzzle and cannot establish all of the elements of the crime:

DNA doesn’t necessarily determine the case; it only determines defendant’s presence. What happened and everything else is still up for grabs. Sometimes you get DNA from a woman who says that she was raped; you still have to show that it’s not consensual. That’s not the end of the case. (Judge 9)

Similarly, judges noted the evidentiary limitations of fingerprint analyses in establishing the required elements of the charged crime. In particular, while judges perceive that fingerprint evidence is highly probative for identifying perpetrators in circumstantial cases (e.g., it can place a defendant at the scene of the crime), it doesn’t necessarily determine guilt: “Just because you find somebody’s fingerprints, that in and of itself doesn’t necessarily mean anything unless it ties in with other evidence” (Judge 40; i.e., there may be another explanation for defendant’s presence at the scene). Likewise, judges emphasized that ballistics evidence is only a small piece of the evidentiary puzzle. Specifically, while evidence relating to gun operability is usually highly reliable, weapon operability is rarely an issue at trial: “I have never
found it to determine guilt or innocence” (Judge 7). Finally, judges emphasized that chemist testimony is usually not dispositive of case outcome.

Inability of Attorneys

Discussions about properly preserving the chain of custody of fingerprints, ballistics evidence, and samples of drugs (e.g., by police and scientists) were noted during discussions about evidentiary reliability.

Reference to the Frye Standard

Only a minority of judges referenced the Frye standard when discussing how they evaluate the reliability of scientific evidence. Even among those judges who acknowledged that Frye is the applicable standard which guides reliability determinations of scientific evidence in the jurisdiction, their discussions regarding reliability were dominated by other areas of concern.

Prevalent Areas of Concern

Areas of expressed concerns varied by evidentiary type. As discussed below, judges focused on the inability of attorneys to challenge DNA experts, the variations in quality among fingerprint experts, and the poor communication skills of expert chemists. These concerns are expressed below.

Incompetency of Attorneys

Judges were highly critical of the inability of attorneys to challenge the reliability of DNA experts and lamented that this widespread deficiency limits the power of the court system to achieve its truth finding mission: “The adversarial system fails its purpose when the attorney is not skilled enough to challenge the expert” (Judge 14; see Damon-Moore, 2017; Edmond & Roque, 2012; Kovera & McAuliff, 2000; Saks & Faigman, 2008). Many judges attributed this inability to attorneys’ lack of scientific knowledge (see Edmond & Roque, 2012; Saks & Faigman, 2008): “DNA people almost always get a free pass because lawyers are intimidated by DNA and do not know how to attack it” (Judge 16).

Variations in Quality of Experts

With heightened concern regarding the subjective aspects of fingerprint evidence, judges focused on the quality of the fingerprint expert: “It all depends on the quality of the expert, which varies. I really think that there’s a strong bearing in who collected it, how they collected it, what their background and training is, and how they come across as a witness” (Judge 30). Several judges criticized certain police technicians who present fingerprint evidence in the jurisdiction due to lack of qualified experience:

In contrast, about a third of the judges perceive that the quality of fingerprint experts appearing in their courtrooms as relatively uniform: “Distinctions in the quality of these experts are minimal” (Judge 8).

Inability to Communicate Effectively

The overwhelming area of focused concern in response to questions regarding the reliability of chemist evidence is the inability of chemists to effectively communicate with the jury in English: “As far as the testimony is concerned, sometimes they have trouble being understood because English isn’t their first language” (Judge 24). Apart from language barriers, judges were critical of the overall ability of chemists to grasp the attention of jurors: “There is an art to testifying that some people do not get, like the ability to hold the jury’s attention and being up on the field” (Judge 17). Another judge simply stated: “Chemists don’t make the best witnesses” (Judge 6).

Credibility Assessments

Judges also discussed the factors they consider in evaluating witness credibility; consistent with the scholarly literature (see Brodsky et al., 2010; Connolly & Gordon, 2011; Porter & Brinke, 2009; Wessel et al., 2006), assessment tools and determinants vary widely and similar facts were interpreted differently by various judges. In fact, a high degree of subjectivity is evident at the core of judicial credibility assessments, highlighting the potential for vastly disparate conclusions. Factors considered to determine credibility include motive, body language, attitude, witness background (“you
Influence of Personal Experiences

At times, judges referred to their own out-of-court encounters in discussing their perceptions of evidence. For some, their personal life experiences might have influenced the manner in which they evaluated case evidence. One judge considered his own experience as a witness in evaluating the credibility of an “anxious looking” witness appearing in his courtroom: “If they were nervous, they probably had a good reason to be. Not because they were lying...because you should be uncomfortable, I was once called as a witness, I was uncomfortable” (Judge 21). A second judge was not disturbed by the prosecution’s failure to produce fingerprint evidence in a weapons case, based on his own experiences with guns: “People don’t understand that certain surfaces do not leave fingerprints so that’s a difficult hurdle for prosecutors... I... have handled guns so I know sometimes you leave one, sometimes you don’t” (Judge 1). Finally, a third judge was unconvinced by the defense argument that a boxcutter could not have caused an injury, based on his own past injury: “Having been cut by one in the past I know it can cause serious damage” (Judge 2). These examples demonstrate how judicial life experiences sometimes influence judicial assessments of evidence.

Discussion & Conclusions

Given the importance of judicial gatekeeping decisions, reliable and streamlined evaluative processes are critical to ensure that evidentiary decision-making proceeds fairly and that evidence admitted “through the gate” is reliable, relevant, and trustworthy. Despite the importance of fair and consistent gatekeeping processes, our study reveals that evidentiary rulings and evaluations are sometimes fraught with subjectivity and bias and made by uninformed judges using inadequate criteria.

First, consistent with existing scholarship (Allen & Nafisi, 2010; Canela et al., 2019; McQuiston-Surrett & Saks, 2009; Neal et al., 2019; Saks & Faigman, 2008), our data reveal that judges often do not understand the scientific processes involved in analyzing forensic evidence. While it is understandable that judicial expertise lies in legal matters and not science, those tasked to evaluate the admissibility of scientific evidence must be able to understand the processes that they are evaluating in order to effectively carry out their gatekeeping mission. This issue is exacerbated by current documented validity and reliability concerns regarding main feature-comparison methods in forensic science (e.g., fallibility of DNA evidence, lack of foundational validity of ballistic evidence, and the higher-than-believed false positive rates of fingerprint analysis; President’s Council of Advisors on Science and Technology, 2016). In Frye jurisdictions, how can judges possibly evaluate whether or not the scientific processes used are generally accepted within the scientific community if they cannot understand the methodologies employed by the testifying expert? This could be more alarming in Daubert jurisdictions where judges are further tasked with determining reliability using evaluative criteria foreign to their area of expertise. This lack of knowledge would be problematic on its own, even if defense attorneys were fully schooled in applicable scientific processes and able to competently challenge testifying experts. However, prior studies show—and our subjects confirm—that this is not the case (Edmond & Roque, 2012; Saks & Faigman, 2008). In fact, while the stakes involved in criminal cases (life, liberty) are greater than those in civil cases (predominantly financial), criminal cases are far less likely to have competing expert testimony, and defense attorneys are often less proficient in their ability to challenge scientific experts (Dioso-Villa, 2015; Groscup et al., 2002; Risinger, 2000; Shniderman, 2017) than civil court actors. Further, the ability of criminal defendants to obtain competing experts may depend on their socioeconomic positions; variances in financial resources may increase racial disparities in...
gatekeeping decisions. In the end, the combination of judges’ lack of scientific knowledge, the inability of many defense attorneys to challenge expert witnesses, and financial constraints preventing some defendants from securing competing experts, may lead judges to one-sidedly rely on prosecution experts in reaching gatekeeping decisions.

While legislative attempts and (jurisdiction-dependent) case precedents strive to provide guidance by delineating criteria and standards (such as the Frye and Daubert standards) for judges to follow, scholarship reveals that this guidance does not effectively direct judicial gatekeeping processes (Cheng & Yoon, 2005; Damon-Moore, 2017; Groscup et al., 2002; Jensen, 2003; Moreno, 2003; Neal et al., 2019; see Beecher-Monas, 1999 for a combined use of multiple mechanisms to facilitate judges’ decision making). In fact, only a few of our respondents even referenced the Frye standard in discussing reliability determinations, reinforcing prior scholarship that the applicable gatekeeping standards (i.e. Frye or Daubert, or other adapted variations) are not particularly influential in reliability assessments. Indeed, our respondents discussed many considerations they employ in evaluating the reliability of forensic evidence extrinsic to relevant factors, including the showmanship and communication skills of the testifying expert (see Brodsky et al., 2010; O’Regan, 2017). A few judges clearly expressed the reason for this focus: “There is skill in the laboratory and there is skill in testifying. I can only judge their skill in testifying” (Judge 24). While evaluations regarding showmanship and communication skills may touch on the witness’s effectiveness on the stand, it speaks little to the reliability of the substance of that testimony (O’Regan, 2017).

Extending beyond reliability assessments, credibility determinations play a role in gatekeeping decisions. While credibility assessments do not generally require specific expertise beyond judicial common sense, our data reveal that the determination process follows a highly subjective path and that different judges viewing similar witness aspects (e.g., body language) sometimes reach opposing conclusions regarding trustworthiness. Further, our data reveal that judges draw on their personal experiences in assessing credibility and reliability. Indeed, our respondents described a host of influential life experiences that guide their determinations. The influence of a judge’s life experiences on their perceptions and interpretations is natural; one might even argue that judges are elected or appointed to the bench because of their rich backgrounds that influence their world and courtroom views. Yet, distinct life experiences can influence judicial decisions in ways that lead to stark differences in the application of gatekeeping considerations. These differences highlight monumental challenges to equitable gatekeeping processes similar to those echoed by Judge Marvin Frankel and others in discussing sentencing disparities (Frankel, 1972; Lynch, 2009). Existing scholarship is relatively silent regarding unwarranted disparities in evidentiary rulings (in contrast to the abundance of scholarship on sentencing disparities). Nevertheless, evidentiary rulings may lead to unfair outcomes that not only influence the severity of punishment imposed post-conviction, but also affect plea offers, verdicts, and case strategies at earlier influential stages. Without effective advocates or knowledgeable gatekeepers, criminal defendants are vulnerable to damaging unreliable evidence that unfairly weaken their position during the adversarial process.

This study has several methodological limitations. First, our study directly explores judicial perceptions of four types of forensic evidence (DNA, fingerprints, ballistics, and forensic chemistry) and judicial credibility assessments of witnesses. Future studies should expand the types of forensic evidence and also study non-scientific evidence including confessions, eyewitnesses, and other physical and documentary evidence. Second, our respondents all preside in a single state that employs the Frye standard for admissibility of expert scientific evidence. Since the “general acceptance” test in Frye is considerably more general than the criteria established in Daubert, it would be helpful to sample judges from Daubert (or its variant) jurisdictions to further explore judicial understanding and application of the relevant standards in their jurisdictions. However, this limitation is mitigated based on existing literature that establishes that, regardless of the adoption of Frye or Daubert standards, “the practical results are essentially the same” (Cheng & Yoon, 2005: 474; see also Faigman et al., 2002; Giannelli, 2003). Finally, our study sample is limited to judges; future studies should explore the perceptions of other court actors regarding judicial gatekeeping decisions.

References


### About the Authors

**Esther Nir** is an Associate Professor in the Department of Criminal Justice at New Jersey City University. She received her doctoral degree from Rutgers University School of Criminal Justice and her law degree from Fordham Law School. She previously served as a prosecutor in Queens County, New York. Her research interests include sentencing disparities, police misconduct, qualitative methods, and community engaged learning initiatives. Her research can be found in various criminal justice and research methodology journals including *The British Journal of Criminology, Criminal Justice Policy Review, The International Journal of Social Research Methodology, International Journal of Police Science and Management*, and others.

**Siyu Liu** is an Assistant Professor of Criminal Justice in the School of Public Affairs at Penn State Harrisburg. Her research is currently focusing on police legitimacy, juvenile delinquency, and the Chinese death penalty. Her work has appeared in *Journal of Quantitative Criminology, Criminal Justice and Behavior, Psychology, Crime and Law, Asian Journal of Criminology*, and others. She received her doctoral degree in 2014 from the School of Criminal Justice, State University of New York at Albany.

### Endnotes

1. In contrast to jury trials, judges in bench trials are both gatekeepers of the evidence and triers of the fact.

2. Subsequently, the Supreme Court underlined the limited scope of appellate review in judicial gatekeeping decisions and emphasized the wide discretionary power of trial judges in the evaluation of methodology, as well as the expert’s reasoning (*General Electric Co. v. Joiner*, 1997); the Court further expanded the trial judge’s authority based on Daubert beyond scientific testimonies to all expert testimonies (*Kumho Tire Co v. Carmichael*, 1999).

3. While only 7 of our 41 respondents are female, this number roughly reflects the percentage of female judges practicing in the state at that time, by district.

4. Since our data were collected, the applicable rules and laws of evidence admission are substantially the same.