

‘Is your accuser me, or is it the software?’ Ambiguity and contested expertise in probabilistic DNA profiling

Social Studies of Science

2024, Vol. 54(1) 30–58

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03063127231186646

journals.sagepub.com/home/sss

Hannah Pullen-Blasnik ,
Gil Eyal, and Amy Weissenbach

Abstract

What happens when an algorithm is added to the work of an expert group? This study explores how algorithms pose a practical problem for experts. We study the introduction of a Probabilistic DNA Profiling (PDP) software into a forensics lab through interviews and court admissibility hearings. While meant to support experts’ decision-making, in practice it has destabilized their authority. They respond to this destabilization by producing alternating and often conflicting accounts of the agency and significance of the software. The algorithm gets constructed alternately either as merely a tool or as indispensable statistical backing; the analysts’ authority as either independent of the algorithm or reliant upon it to resolve conflict and create a final decision; and forensic expertise as resting either with the analysts or with the software. These tensions reflect the forensic ‘culture of anticipation’, specifically the experts’ anticipation of ongoing litigation that destabilizes their control over the deployment and interpretation of expertise in the courtroom. The software highlights tensions between the analysts’ supposed impartiality and their role in the courtroom, exposing legal and narrative implications of the changing nature of expertise and technology in the criminal legal system.

Keywords

expertise, algorithms, probabilistic DNA profiling, DNA analysis, forensics

Department of Sociology, Columbia University, New York, NY, USA

Correspondence to:

Hannah Pullen-Blasnik, Columbia University—Sociology, 606 W. 122nd St. Knox Hall, Suite 501, New York, NY 10027, USA.

Email: hannah.pullen-blasnik@columbia.edu

What happens when an algorithm is added to the work process of an expert group? An emerging sociology of algorithms has moved beyond predicting the wholesale replacement of human experts by AI that renders human input superfluous (Aerts, 2018; Agrawal et al., 2018; Domingos, 2015; Pasquale, 2015; Siegel, 2013). Instead, sociologists have documented hybrid arrangements where human experts, machines, and algorithms collaborate (van den Broek et al., 2021), ranging from the emergence of new forms of ‘algorithmic expertise’ (Sachs, 2020; Waardenburg et al., 2018) to ‘loose coupling’ where algorithms are merely ‘shiny new toys’ used for legitimization (Christin, 2017). By the same token, some legal scholars suggest that the ‘true declarant of any machine conveyance’ is neither a human nor a machine, but ‘the result of distributed cognition between technology and humans’ (Dror & Mnookin, 2010; Roth, 2017, p. 1978). This focus on collaboration and distribution— however important as a counterpoint to the simplistic pitting of humans against machines (Stark, 2022)— risks underplaying the tensions and contradictions that result from incorporating an algorithm into an expert’s work process.

Forensic expertise provides an especially useful case study. DNA analysis has come to play an essential role in the prosecution and defense of a variety of cases. To introduce DNA evidence in court proceedings, DNA analysts, or criminalists, analyze a sample and formulate a judgment about the likelihood that a specific person of interest contributed DNA to the mixture. The credibility of this probabilistic judgment depends upon a credentialed expert vouching for the soundness and objectivity of the process.

In recent years, forensic science laboratories have incorporated the use of algorithmic software as decision support tools to analyze cases previously deemed too complicated. While in the past analysts inspected graphs visually and limited themselves to cases with suspected DNA from one or two contributors, it is fast becoming standard practice to use Probabilistic DNA Profiling (PDP) software to arrive at the judgment in cases with minute amounts of DNA or with multiple potential contributors. The question becomes: What is the likelihood that a specific person of interest is one of the two to four people whose DNA seems to be present in the sample? As a result, cracks have opened in what has been a settled field of practice and jurisprudence since the end of the ‘DNA wars’ (Aronson, 2007, pp. 173–202). An increasing number of contentious admissibility hearings raise fundamental questions about what has been dubbed ‘machine testimony’ (Kwong, 2017; Roth, 2017).

To explore the tensions that arise or reemerge with the introduction of PDP software, we conducted interviews with DNA analysts at a large forensics laboratory that is among the leaders in its adoption. Additionally, we analyzed court hearings conducted between 2015 and 2022 that debate the admissibility of evidence produced using PDP software and supplemented this research with an interview with a defense attorney who specializes in forensics litigation. We found that DNA analysts, to draw boundaries around their work and expertise, balance competing stories about the software and their relation to it. Some stories perform ‘boundary work’ (Gieryn, 1983) between human decision-makers and machines that merely execute these decisions. Others jettison boundary-work in favor of hybridization, where the algorithm looms large as co-creator of forensic judgment and an indispensable corrective to the subjectivity of human expertise. These narratives are reflected in work practices within the lab and in expert testimonies within the courtroom.

What is the purpose of simultaneously telling two different, potentially contradictory stories? We argue that, while the algorithm has made their performance of expert testimony more defensible against accusations of subjectivity, it also creates distinct vulnerabilities that destabilize analysts' authority. An algorithm's introduction typically is part of an 'objectivity campaign' (Eyal, 2019, pp. 115–118), premised on problematizing the subjectivity of human judgment as biased, error-prone, and inconsistent (Daston & Galison, 1992; Porter, 1995). Yet, algorithms do not eliminate reliance on human judgment, but merely background it (Eyal, 2019, pp. 118–120). Thus, the introduction of algorithms both undermines trust in trained experts and relies on that expertise, creating a source of contention.

PDP software, championed as a solution to the problem of subjective and unreliable forensic expert judgment, relies on assumptions and judgment calls built into the source code. This has led courts to question analysts' authority and whether fundamental changes in the relevant jurisprudence are necessary to incorporate competing experts, such as computer scientists and software engineers, into the relevant scientific community (Kwong, 2017, pp. 286, 297–301; Roth, 2017, pp. 2017, 2027, 2035). These legal debates impinge upon DNA analysts through the 'anticipatory culture' of the forensic laboratory (Bechky, 2021), leading them to alternate between boundary-work and hybridization.

We present a series of tensions in how analysts discuss the software within the legal and anticipatory context. The first tension addresses the novelty, or lack thereof, of PDP software compared to previous analytical techniques. The second contrasts how analysts present themselves as the sole authors of expert judgment but also describe ways the software co-creates this judgment. The third regards the strict boundaries analysts draw around their work as impartial scientists and the results of the case to 'distribute responsibility between the lab and the court' (Kruse, 2012). The radical opacity of the software threatens this careful orchestration of boundary work, introducing new sources of bias and error. As a result, analysts struggle to maintain control over their expert status as the software's sole interpreter.

Algorithms in expert work settings

Collins's (1990) prescient *Artificial Experts* sketched the sociological response to algorithms in expert work settings and has since been expounded on in a variety of cases (e.g. Autor, 2015; Bessen, 2015; Davenport & Kirby, 2016; Mindell, 2015; Shestakofsky, 2017). Intelligent machines, argues Collins, are 'social prostheses' that must fit into pre-existing social arrangements. People digitize their input and repair their output—not only engineers designing the technology, but also workers creating training data (Irani, 2015), maintaining and updating digital technology (Ekbia & Nardi, 2017), and occupying intermediary interpreter positions (Waardenburg et al., 2018). The resulting 'expert system' is a collective composed of people and machines working in concert.

Within this framework, one could still anticipate that the introduction of algorithms would deskil and downgrade human labor, as seems to be the case with image-tagging performed by poorly paid, contingent laborers across the globe (Irani, 2015). In such cases, companies manage, surveil, and make contingent work that previously demanded

more secure employment (Delfanti, 2021; Griesbach et al., 2019; Vertesi et al., 2020). In other cases, workers adapt to new technology by modifying their expertise, developing new skills, and abandoning old ones (Nelson & Irwin, 2014; Walker, 1958; Zuboff, 1988). These changes produce tensions between positions in their ability to embrace new technology (Barley, 1986; Barrett et al., 2012). Workers redefine their roles and knowledge to highlight their necessity, and in so doing impose their interpretations on the technology's use (Faraj et al., 2018; Liker et al., 1999; Schultze, 2000). Organizations, too, dynamically reconfigure teams around algorithms to provide labor complimentary to the technology (Shestakofsky, 2017).

Adding technology to a workplace can quickly and dramatically undermine the legitimacy of previous expertise and destabilize pre-existing power dynamics. Waardenburg et al. (2018) show how a predictive policing tool supplanted the judgment of Dutch police officers in determining where and when to patrol. At the same time, newly created 'intelligence officers', initially low status, became powerful authorities interpreting the algorithmic output. Despite the subjectivity of their recommendations, others took their decision-making as objective because of its algorithmic backing. Sachs (2020) finds similar patterns in her ethnography of art experts, where some non-technical experts acquired significant influence by honing skills in interpreting and repairing image similarity match scores.

These newly empowered expert roles emerge in response to the opacity (or 'black box-ness') of the algorithm and its source code (Pasquale, 2015). Opacity creates the need for interpreters and spokespeople, on the one hand, but also creates distinct legitimation problems. Lebovitz et al. (2022) study the use of AI tools in three radiology departments and find that the tools' opacity caused perplexity for radiologists, especially when diagnoses conflicted. While in two departments this led radiologists to ignore the output, in the third they developed 'rich interrogation practices of the opaque AI results' and incorporated them into their final judgments. As van den Broek et al. (2021) find in their study of hiring algorithms, success often depends on a willingness for mutual learning between software developers and subject experts. Under such conditions, technology may lead to a 'new hybrid practice that relie[s] on a combination of ML and domain expertise'.

Finally, in some cases experts successfully resist changes to their work practices. Christin (2017) interviewed and observed legal professionals working with risk-assessment algorithms. They developed a range of strategies—'foot-dragging', 'gaming', and 'open critique'—to buffer themselves from the algorithm's impact. Consequently, the risk-assessment algorithm created only minimal changes in the practices and status of the professionals involved. Yet, its existence still provided the appearance of objectivity and rationality that the criminal legal system lauds. Similarly, doctors in a large hospital resisted the introduction of a machine learning algorithm meant to shorten hospital stays, reducing it from a decision-support tool to a surveillance and management one (Galper & Kellogg, unpublished).

The literature on algorithms continues to evolve, but when an expert work setting incorporates an algorithm, it creates a series of perturbations like what Abbott (1988) suggests is the consequence of jurisdictional struggles between groups of human experts. Rarely do algorithms completely replace experts, and even then, as Collins (1990, p. 2018) indicates, they still must fit into a human-machine collective.

Between human expertise and mechanical objectivity

Developers and management often problematize the work practices of experts as subjective to highlight the benefit of algorithmic input as more trustworthy ‘mechanical objectivity’ (Daston & Galison, 1992). They cite cost, error rate, and lengthy turnaround time to claim that algorithms would outperform human experts. While pressures from without can inspire ‘objectivity campaigns’, as with the introduction of RCTs into the field of development economics (de Souza Leao & Eyal, 2019), the impetus to use algorithms may come also from within to fortify themselves against public distrust and accusations of bias.

The turn to mechanical objectivity in a field, as Porter (1995) argues, can be attributed to the relative weakness of experts in that field, and their need to shore up public trust. Cole (2001, p. 266) shows that highly scrutinized British fingerprint examiners opted to limit their own discretion with a uniform rule requiring a minimum of 16 ‘ridge characteristics’ to declare a match, while the more autonomous US examiners did not feel the need to do the same. In examining the introduction of DNA analysis in the 1980s and 1990s, Aronson (2007, pp. 196–200) notes that, despite increasing automation, lingering reliance ‘on human intervention and judgment’ remains a point of challenge. Expert judgment was portrayed as subjective and potentially biased in favor of the prosecution, leading to increased reliance on DNA software.

This sets up a tension. The introduction of algorithms into the workplace obscures reliance on human judgment in at least three distinct ways: Certain judgments are ‘black-boxed’ in the source code¹ (Faraj et al., 2018; O’Neil, 2016). Second, the decisions in what is included and excluded are presented as purely conventional and unproblematic (Aronson, 2007, pp. 198–200; Pine & Liboiron, 2015; Ribes & Jackson, 2013). Third, interpreting and repairing the output are folded into the rhetorical device *prosopopoeia*, speaking in the name of an absent entity, ‘an ensemble that is made to exist by the fact of speaking in its name’ (Bourdieu, 2014, p. 45). To introduce mechanical objectivity, it is necessary to problematize human judgment as subjective and biased, especially forms of judgment that previously enjoyed ‘disciplinary objectivity’ (Porter, 1995, pp. 3–4). Yet algorithms still rely on human input and interpretation, creating instability. Those introducing algorithms to an expert work process may then swing between undermining human judgment as subjective and untrustworthy, on the one hand, to justify the reliance on mechanical objectivity, and reasserting its impartiality and usefulness, on the other, as attention shifts to those judgments that have been backgrounded.

Forensics faces constant pressure to appear objective. Bechky (2021) shows that forensic laboratories are dominated by a ‘culture of anticipation’: Every aspect of work is designed to withstand scrutiny during expert testimony. The mechanical objectivity of procedures, protocols, numbers, and machines protects the testifying expert from accusations of subjectivity and bias. At the same time, it requires the testifying expert to appear as a credible spokesperson for the mechanically objective instrument, technique, or number. That an expert witness can speak for this evidence, can deploy the device of *prosopopoeia*, is not a given. Judges in the U.S. have alternated between deferring to the jury’s judgment of evidence or permitting experts to interpret evidence for the jury. When US courts first weighed x-rays’ admissibility, they declared x-rays

identical to photographs and subject to jury interpretation. Over time, however, courts began to permit the newly emerging profession of radiology to interpret x-rays for the jury (Golan, 1995, pp. 206–207). The higher the probative value of such mechanically objective evidence, the more likely it comes under such scrutiny. Forensic DNA holds a very high probative value, a proverbial ‘smoking gun’. As we shall see below, labs respond to this potential vulnerability by performing ‘boundary work’ (Gieryn, 1983) between scientific and legal facts. Precisely this boundary work, however, is destabilized by the algorithm.

Probabilistic DNA profiling

DNA profiling entered U.S. courtrooms in 1987 and was subject to heated admissibility debates through the early 1990s (Aronson, 2007). Big biotechnology companies began offering and patenting DNA profiling products and services, and quickly DNA evidence became the gold standard in criminal cases. DNA profiling allowed prosecution teams to report a random match probability—the probability of a match between a subject’s DNA and that of an unrelated person from the population—that sounded like staggering evidence of guilt. For example, in a 1987 Florida criminal case, lawyers told the jury that the probability of a random match was one in ten billion (Aronson, 2007, p. 37).

Meanwhile defense teams, usually in the position of challenging DNA forensics results, have had to play an ongoing game of catch-up to challenge the admissibility of DNA evidence. The prosecution had access not only to the biotechnology companies’ proprietary databases and genetic probes, but also to expert communities ready to testify to the reliability of the technology and knowledge base upon which it was built, mostly in medical research (Aronson, 2007, p. 55). Defense teams made headway by arguing that DNA profiling in medical research was not a reliable reference for forensics work, since samples in forensics were degraded and in lower quantities, with different stakes to reporting a match (pp. 58–9). They also challenged population assumptions built into forensic labs’ models, for example that databases constituted representative samples by race and interracial reproduction (pp. 29–31). They further argued that forensics work was unscientific because there were no standards across labs (p. 89).

However, by the mid 1990s, most of these controversies had been laid to rest by the courts. The National Research Committee accepted the ‘ceiling principle’ to handle population substructure, taking the most conservative estimate based on empirical data on racial/ethnic populations. Meanwhile, the Federal Bureau of Investigation (FBI) took over DNA forensics, developed a DNA typing regime, built public laboratories, and established their own voluntary guidelines and national DNA database, which were not subjected to outside review or regulation (Aronson, pp. 91–103). Their sway over congressional expert commissions meant that they were able to focus regulatory bodies’ attention on the issue of population genetics and dismiss other questions about representing error in expert testimony. Aronson writes that by 1993, the courts viewed ‘DNA profiling as a technological system could finally be considered fundamentally valid and reliable, because population genetic and statistical issues were no longer a major concern and protocols existed to prevent problems arising from the damaged and degraded nature of test samples’ (p. 175). This despite the fact that there was still ‘no generally accepted

or nationally recognized standard for declaring a match between two profiles', and still no third party validation or regulation of proprietary DNA analysis systems.

How is the situation today different from the earlier period of the DNA wars? First, there has been a distinct change in the type of cases undertaken by forensic laboratories. For years, criminalists analyzed electropherograms visually, processing only large single-source DNA samples and two-person mixtures (President's Council of Advisors on Science and Technology, 2016, pp. 69–70). As DNA quantification methods have become more sensitive, forensics laboratories have taken on cases involving more complex mixtures with lower amounts of DNA and multiple contributors, which are more difficult to interpret by visual assessment and human calculation alone (Butler, 2021, pp. 12–16). These complex samples introduce additional uncertainty, as it becomes harder to identify missing or misleading information and, given that DNA is very easily transferred, more pertinent to discern the relevance of the DNA to the case.²

Second, analysts no longer report their findings using the language of a 'match' as random match probabilities are prone to misinterpretation by 'the prosecutor's fallacy' and align the forensic expert with the adversarial argument of lawyers (Thompson & Shumann, 1987). Instead, analysts report their findings as Bayesian likelihood ratios (LRs), or the conditional probability of finding the specific combination of alleles present in the sample if a known person of interest 'contributed' DNA to the sample, compared to if the DNA came from another, unknown person in the population. They bound their expertise to prohibit testifying to how or why the DNA may be present, leaving the final decision about culpability to the court (Kruse, 2012, pp. 671–672). However, PDP software interferes with the strict boundary around expertise and responsibility.

Almost all major laboratories now rely upon proprietary software to analyze complex mixtures, reigniting the legal debate through a series of contentious admissibility hearings. Legal scholars have zeroed in on the 'black box' quality of algorithms to argue for the need to adapt to a new era of 'machine testimony' (Kwong, 2017; Roth, 2017). They argue that the 'true declarant of any machine conveyance' is neither a human nor machine, but should be seen as 'the result of distributed cognition between technology and humans' (Dror & Mnookin, 2010; Roth, 2017, p. 1978). Just as the 'hearsay dangers' of human witnesses 'are believed more likely to arise and remain undetected when the human source is not subject to the oath, physical confrontation, and cross-examination, black box dangers are more likely to arise and remain undetected when a machine utterance is the output of an 'inscrutable black box'" (Roth, 2017, p. 1976). Such black box dangers include 'human error at the programming, input, or operation state, or because of a machine error due to degradation and environmental forces'. Roth ends her argument by calling for the algorithms and source code to be made available to defense experts as a routine matter, something that at least a few courts have recently sought to compel.

Many of the questions raised by the litigation and the legal literature are identical to the questions that are of interest to sociologists and STS scholars: Does the addition of the algorithm represent a significant change in the work practices of an expert group? Nonetheless, this literature has very little to say about how the current legal impasse impinges upon criminalists through the 'culture of anticipation' (Bechky, 2021) of the forensic laboratory.

Data and methods

To understand how analysts perceive the role of algorithms and expertise in their work we conducted ten in-depth, semi-structured interviews with analysts employed in a large urban medical examiner lab who work directly with the software. Though a relatively small sample size, our interview subjects' cumulative experience covered hundreds of cases at all stages of the analysis process and over 62 cumulative years of experience in the lab. We believe that they provide reasonable representation of the experience of DNA analysts at the levels that interact with the software. The interviews proceeded through the steps involved in the lifecycle of a DNA sample, covering the handling of samples from admission through to the final report, points of ambiguity and contention in the process and interpretation, courtroom testimony, and reflections on automation. We were not permitted to record the interviews, so members of the research team transcribed content in real time. We coded transcripts thematically and developed a composite chronology of the stories, judgments, and ambiguities arising at each stage of the process.

Through our interviews it became clear that there was a set of tensions that appeared repeatedly in analysts' accounts. In some cases, these tensions became apparent in the responses of a single analyst, especially as they anticipated possible third-party challenges to their story. At other times, however, the variation appeared as a discrepancy between the answers given by different analysts to the same question. Our argument is less about the multiple stories told by certain analysts than it is about a collective story, the lab's story, that we have reconstructed from our various interviews. While we do not focus on the individual analysts, we do provide numbers for each interview based on the order they appear in the text. The overlapping stories and points of contention that these interviews expose provide fruitful data for understanding this specialized and often inaccessible expert community. While we conducted interviews over zoom during 2020, we focused on the analysts' day-to-day experiences prior to COVID-19.

Additionally, we read and coded ten court opinions on PDP software, seven of which are admissibility hearings, two are appeals of previous admissibility hearings, and one is a hearing about whether an admissibility hearing is necessary (Table 1). The cases were selected based on their significance to the evolving controversy and to represent the full gamut of litigation caused by the software's introduction. These hearings took place from 2015 to 2022 in a diversity of jurisdictions and debate whether evidence produced by PDP software—FST, TrueAllele or STRmix—is reliable and can be admitted as evidence. While admissibility hearings on PDP software began in 2010, the first decision against admissibility dates from 2015 and is in our sample. The New York State hearings assess whether the software is 'generally acceptable within the scientific community' (the *Frye* criterion), the appeal in Michigan considers whether it satisfies the more stringent *Daubert* criteria, and the New Jersey hearings question the 'right to face your accuser'. We draw on the hearings to contextualize the analyst interviews and provide supplementary evidence for our arguments. We also use quotes and anecdotes from an interview with a defense lawyer who specializes in forensic litigation and identify such quotes with '(DL)'.

What analysts say about their relationship to the algorithm must be understood in context. We precede our findings with a section describing how work is organized at the

Table 1. Summary of court cases relevant to deciding admissibility of PDP software for DNA analysis, 2015–2022.

Court case	Decided	Jurisdiction	Software	Holdings
People v. Collins	7/2/2015	NY Kings Supreme Court	FST	High sensitivity DNA analysis and FST are not generally accepted as reliable (<i>Frye</i> hearing)
People v. Carter	1/12/2016	NY Queens Supreme Court	FST	Denies defense motion to remove DNA evidence generated by FST from testimony, conduct a <i>Frye</i> hearing, and/or disclose FST source code
People v. Bullard-Daniel	3/10/2016	NY Niagara County Court	STRmix	DNA evidence is sufficiently reliable to be admissible (<i>Frye</i> hearing)
United States v. Gissantaner	10/16/2019	Western District of Michigan District Court	STRmix	Probabilistic genotyping software program used to determine the probability of DNA profile in a mixture of DNA from multiple contributors is not sufficiently reliable, as a prerequisite to admission of DNA report.
United States v. Lewis	3/3/2020	MN US District Court	STRmix	Absence of error rates and any lack of compliance with software development and testing standards did not render program unreliable; validation testing established foundational reliability
People v. Williams	3/3/2020	NY Court of Appeals	FST	Trial court abused its discretion in permitting admission of LCN DNA evidence and FST without holding <i>Frye</i> hearings
State v. Pickett	2/3/2021	NJ Superior Court	TrueAllele	Defendant is entitled to access the software's source code, supporting software development, and related documentation with proper protections
Appeal of US v. Gissantaner	3/5/2021	US Court of Appeal, 6th Circuit	STRmix	Reversed earlier decision by district court. DNA evidence satisfies <i>Daubert</i> criteria
State v. Rochat	1/28/2022	NJ Superior Court	FST	FST and LCN DNA evidence not generally accepted as reliable and trial court's error in admitting them was not harmless
People v. Wakefield	4/26/2022	NY Court of Appeals	TrueAllele	PDP generally accepted as admissible; report created by software is testimonial; source code is not in itself testimonial and is not entity that can be cross-examined, its disclosure not required

forensic laboratory based on our analyst interviews, depicting the lifecycle of a DNA sample from crime scene to court testimony. We then outline three major tensions in both the interviews and court hearings: How material are the software's contributions? Who has final decision-making authority? How are boundaries drawn between scientific facts ('data') and legal facts ('context')?

The setting

DNA analysis occurs over four rotations: evidence exam, pre-and post-amplification, report writing, and technical review. An analyst starts with one or more pieces of evidence—say, a t-shirt—that has suspected DNA of interest on it. They swab the evidence for DNA based on information from the police report. On the pre/post-amplification rotation, analysts extract DNA from the resulting sample and amplify its volume. A gel electrophoresis separates the DNA into alleles, short segments of DNA that vary from person to person. The resulting image translates alleles into spikes at 15 preset identifying locations where higher peaks indicate greater quantity of those alleles in the sample.

The most junior analysts split their time evenly between the first two rotations, running amplification and electrophoresis in large batches. Second-level analysts assume new responsibilities including interpreting DNA profiles and writing reports summarizing results. Analysts “deconvolute” the results, contending with the stochastic errors from the amplification process, including ‘drop-out’ (no allele at one of the 15 points), ‘drop-in’ (contamination by an extraneous allele), and ‘stutter’ (slippage of the DNA strand while being copied) (Butler et al., 2021, p. 26). From this information they determine the number of contributors to the sample. Deconvolution is especially difficult with complex mixtures of DNA containing low copy number (LCN) DNA contributed by more than two persons.

For this reason, analysts increasingly rely on proprietary software in the post-amplification stage. STRmix is the most popular such software, with TrueAllele its main competitor. Both are commercial software packages, while an earlier option, FST, developed by New York City's Office of Chief Medical Examiner (OCME), has been tied up in litigation and phased out of usage around 2017. Given an assumption about the number of contributors to the sample, a particular distribution of peaks in the electropherogram, a hypothesis about whether a known individual is a contributor or not, and a large database of genomic information about the general population, the software uses a Bayesian algorithm to calculate the LR. This calculation must also consider the noise caused by the amplification process, yet it is debated whether a human analyst can meaningfully verify the software's process with LCN DNA at levels below the standard 100 picogram threshold employed by human analysts. With the output from the software, the analyst then writes up a report summarizing the findings that undergoes technical review by a fourth level criminalist before it is considered complete.

Finally, an analyst may be assigned to testify in court about the report. The work process is organized such that no single analyst conducts an analysis from start to finish. Rather than follow a particular case, analysts work on rotations at particular stages in the analysis and may handle three to four cases per day. As a result, each DNA sample is analyzed through collaboration among several analysts, so one analyst must stand in for the lab process as a whole during testimony. We refer to composite stories of the lab's

identity and practices, just as ethnographers often create ‘composite characters’ to shield the identity of their subjects (Arjomand, 2022). The organization of the work process by rotations, rather than a single analyst ‘owning’ a case, supports our reasoning.

Contested novelty: Software is ‘just a tool’, but has become required ‘statistical backing’

Nothing new: The software is ‘just a tool’

The introduction of PDP software into DNA analysis raises questions about the novelty of algorithmically generated results. Is the output from the software meaningfully distinct from the output that an analyst could produce by hand or using older methods? In our interviews, many analysts claim that the answer is no: the algorithm is ‘just a tool we use to help us interpret the sample. We run the evidence through, the algorithm will give us the results, but a trained DNA analyst has to go in and make sure it conforms to our intuitive explanation’ (1). By this interpretation, analysts control the analysis of DNA evidence; the algorithm simply makes the process more manageable and efficient. One analyst told us, ‘STRmix just makes our job easier. It does tons of calculations that it would take us an extremely long time to do’ (2). Another explained, ‘We don’t let STRmix become this big, overarching—it doesn’t rule in our lab. It’s just software’ (3). The algorithm makes no decisions but merely facilitates the experts’ analytic process: ‘We equate STRmix to being a calculator—we’re putting something in and seeing if it makes sense what we get out of it, based on parameters we’ve set and things like that’ (4).

A New York judge who denied the need for an admissibility hearing made essentially the same argument. The algorithm, he said, was ‘not new, novel, or experimental ... [it] is merely the [lab]’s way of computerizing the analytical process in order to make it feasible’ (*People v. Carter*, 2016). Another judge repeated analysts’ mantra: ‘The FST software does calculations of Bayesian probabilities like a simple calculator’ (*People v. Collins*, 2015).

Analysts often implied, and sometimes explicitly said, that they could perform the job without the algorithm. One emphasized that analysts do not ‘blindly follow’ the output and have the technical capacity to analyze manually: ‘We could do the calculations by hand, but it would take thousands of hours’ (5). A judge also noted that ‘the mathematics have been repeated by hand by the U.S. Army and the California Department of Justice’ (*United States v. Lewis*, 2020). The message is that the algorithm’s mechanics are nothing conceptually novel: it alters neither the process nor the analysts’ ultimate judgment. In other words, ‘[We] use [the software] as a tool the way people use Excel: You still need to take the information, absorb it yourself, and create a conclusion’ (1). The software provides a new tool, but the work process and results are unchanged.

Admissibility and algorithmic opacity

However, some criminalists, judges and lawyers also articulated the opposite view, namely that ‘probabilistic genotyping software marks a profound shift in DNA forensics ... There is a substantial difference between testing DNA utilizing traditional DNA

methods and analyzing low levels or complex mixtures of DNA relying on probabilistic genotyping software' (*State v. Pickett*, 2021). A defense lawyer responded thus to the idea that PDP algorithms have not substantively changed DNA analysis: 'My immediate reaction is that that's wrong. And it's wrong legally and it's wrong about how a jury hears evidence, it's wrong about how analysts testify evidence, it's wrong about how we as defenders defend against the evidence, it's just wrong in every way' (DL).

The resurgence of admissibility hearings reinforces this claim. If increased speed and higher throughput were all that the addition of the algorithm changed about the work of forensic labs, an admissibility hearing would hardly be necessary. In *United States v. Gissantaner* (2019), the court ruled that STRmix did not fulfill the *Daubert* criteria for 'a finding of reliability':

The DNA evidence sought to be admitted in this case ... is not really evidence at all. It is a combination of forensic DNA techniques, mathematical theory, statistical methods ... decisional theory, computer algorithms, interpretation, and subjective opinions that cannot in the circumstances of this case be said to be a reliable sum of its parts. Our system of justice requires more.³

This objection was raised in earlier cases where the prosecution sought to estimate LRs by combining different techniques in a single formula.⁴ However, it acquires new significance when the techniques, assumptions, and methods are blackboxed within proprietary software. This opacity introduces new sources of error and alters the lab process. An audit of FST's source code revealed that certain loci were removed from the LR calculation without the analyst's knowledge and, as a result of this bug, the software was overestimating the likelihood of guilt (Bellovin et al., 2021; Lacambra et al., 2018). This realization played a major role in the discontinuation of FST and the switch by most labs to STRmix and TrueAllele. In response to concerns about errors found in STRmix source code in 2015 and 2017, its developer testified that 'numerous diagnostics render STRmix's internal processes transparent and reviewable' (*United States v. Lewis*, 2020). Despite these assurances, the matter remains at the heart of current litigation.

Though in our sample the judges who held admissibility hearings split fairly evenly on ruling the algorithm admissible (Table 1), they overwhelmingly agreed that the introduction of the software made a significant change to the work practices of the lab, requiring additional scrutiny. One judge considered that a trial court 'abused its discretion' by denying a *Frye* hearing in a case involving FST (*People v. Williams*, 2020). Another, who found algorithmically produced results to be reliable and admissible, nonetheless viewed litigation as still very much ongoing and undecided: 'It may be among the first words in New York courts on the admissibility of STRmix, but the Court certainly does not expect it to be the last' (*People v. Bullard-Daniel*, 2016).

Changes to lab evidence, complexity, and required 'statistical backing'

Analysts and court opinions identify three common threads to how PDP software alters lab analysis. First, the software allows labs to analyze cases they would not have been able to in the past. 'As technology has changed,' one analyst said, 'we have been able to

get DNA from more and more types of samples' (5). A defense lawyer summarized the novelty of the cases as the reason behind the software's introduction:

We're moving to touch samples with mixtures of DNA, degraded, relatives. I mean, you always had those problems, but they didn't dominate the evidence the way they dominate now. So, the nature of the evidence being tested is radically different. Even if you could have mixtures back then, maybe they just weren't interpreted. But now every envelope has been pushed in terms of how low you can go, how many contributors to the mixture. And that's why this software was invented. (DL)

Cases often involve LCN DNA samples below the standard 100 picogram threshold employed by human analysts, at 'the outer limits of a complex mixture of low-template, low level DNA, which has very limited comparable validation, if any' (*United States v. Gissantaner*, 2019). The difficulty of detecting stochastic errors at such low levels turned attention to backgrounded human judgment, namely to the interpretive protocols laboratories adopted to correct for such imperfections. Defense experts argued that there is no scientific consensus about their reliability in high sensitivity (LCN DNA) mixtures (*People v. Collins*, 2015) and 'no controlling standards' for the auditing process (*United States v. Gissantaner*, 2019).

Second, the algorithm was not merely a tool, because the new complexity of samples often means that analysts cannot form judgments independent of the software's output—an independent calculation may take thousands of hours. A defense lawyer with whom we spoke made a similar point: 'DNA analysts pride themselves on the STRmix training, that they have us replicate certain formulas by hand. But it's ..., You couldn't do it by hand, the whole thing' (DL). Practically, the complexity of the statistical calculations would not be feasible without algorithmic assistance, leading analysts to conclude that 'Most cases could not be completed without STRmix at this point' (5, 2).

As a result, the software alters the lab's case flow and transforms work practices: 'If you have a comparison and thought the person wasn't in there, [before STRmix] you wouldn't even run the program. You'd just say they're excluded. Now, you run every comparison and let the likelihood ratio support what is stronger' (6). The algorithm is not only a tool but performs analyses human experts would not otherwise conduct. A judge made the point more forcefully: 'Computer analysis can thus go beyond human review of a sample using methods and calculations that a human would be unable to replicate. Essentially, this testimony revealed that TrueAllele extrapolates and processes data to make judgments that supplant human choice, leading to conclusions based on assumptions generated by artificial intelligence' (*People v. Wakefield*, 2022). An expert witness for the prosecution in that case simply said that TrueAllele's 'increased sensitivity' and 'stronger statistical measure' meant it 'can do what a human cannot or will not'.

Finally, the software at times provides the resulting judgments, which otherwise would have appeared as speculative and tendentious, with 'statistical backing'. Cases with touch samples and LCN DNA are 'more difficult observationally', one analyst explained, but the software gives 'statistical backing' to judgments about such samples (5). Similarly, a judge opined that 'rather than merely presenting evidence that a defendant could be a contributor or could not be excluded as a contributor to a DNA mixture,

using the FST the People are able to provide a quantitative statistical value to the result of the DNA testing' (*People v. Carter*, 2016). Another criminalist implied that with the presumed objectivity that this 'quantitative statistical value' lends to their judgments, they no longer needed to be as conservative as in the past:

[I]t gives us a better picture and more confidence behind our results. When we were doing manual decon, we always erred on the side of caution, and would say 'Hey, can we determine something at this location.' But we have mathematical backing vs. the information we'd have in the past. (7)

There is little doubt that without the mathematical backing provided by the algorithm, such less conservative judgments would have seemed subjective, biased, or lacking basis. This vacillation between depicting the algorithm as merely a tool and much more than a tool has ramifications also for analysts' decision-making authority.

Contested authority: Analysts claim decision-making authority, but must rely on algorithm for final judgment

Analysts claim sole decision-making authority

When asked about how they create a final determination, analysts emphasize that their decision-making is a process guided by their expert training, not the algorithm. 'There's no part in STRmix that we're just letting go', one criminalist told us. '[The result is] automated in the sense that it's calculated for us, but in the final determination of the genotypic assignment, that's on the analyst' (2). Another affirmed their ultimate authority, stating, 'I think it's important with STRmix—although it's a very advanced forensic software—the role the analyst plays is just as vital. In the end, the analyst is the one interpreting the data and making sure results conform to our expectations of the data generated from the process. It's our decision whether the data generated makes sense' (8). While the software output is part of the result creation, the analyst remains responsible. A third similarly voiced, '[You're] not just taking [the software output] for what it is. Each analyst has a final say about what they're seeing in the data, how they're drawing upon STRmix. If we see something that STRmix is seeing that we don't see, there is a determination made by the analyst. It's "the analyst says", not "STRmix says"' (7).

Without experts speaking for the algorithm, they claim that the results do not mean anything and have no force. While some acknowledge collaboration akin to Roth's 'distributed cognition', they insist that the responsibility for the judgment ultimately rests with the analyst, not the algorithm. Judges repeat this view:

A human analyst tells the computer what to download and under what conditions to analyze the data, the analyst tells the computer what questions to ask when interpreting the data and the analyst ... determines how many 'runs', or cycles, of the data the system will complete and the analyst then makes comparisons to form the likelihood ratios ... the analyst certified to more than a machine-generated number. (*People v. Wakefield*, 2022)

Analysts appeal to the algorithm to resolve conflict in the lab

Yet, analysts also acknowledge that their own interpretation would lack force without the algorithm's backing, as becomes apparent when analysts defer to the software to resolve conflict in the lab. Facing internal conflict over an interpretation, analysts use the software output to defuse interpersonal tension. One explained, '[STRmix] cut[s] out a little bit of subjectivity, a little bit of thinking you're right when someone else is wrong because you now have this unbiased opinion of this is a 98% match' (6, 7). Assigning authority to the algorithm helps resolve ill-feeling when correcting another's analysis. STRmix functions as a neutral third party:

Someone can't get offended when you disagree because STRmix said it. ... They see it as a more scientific-based thing than a personal thing. ... It's really all the same kinds of points that you're checking [as before], it's maybe just a little less on a particular person and more on the software. (6)

The algorithm provides both a way to distance the individual analyst from the decision to overrule another's interpretation and a potential source of expertise or blame in the dispute. In such cases, analysts frame it as providing an unbiased opinion; here it *is*, in fact, 'STRmix says' and not 'the analyst says'.

This displacement of authority is also essential for helping the lab to speak in a single, unified voice, crucial to delivering testimony. Since work at the lab is organized in distinct stages, each conducted by a different analyst, the algorithm smooths transitions from one analyst to another. It helps the lab create a cohesive identity, whereby one testifying analyst can speak for the whole. This cohesion addresses a vulnerability that defense lawyers may seek to exploit in determining who must testify in court. The defense lawyer with whom we spoke explained,

We say you can say that [you independently verified the results] until you're blue in the face: Were you the person that looked at the data before somebody edited it? ... Because it's not truly independent, if you just get a piece of paper that someone else has done all the work on and you're like, oh, yeah, that looks right. (DL)

The algorithm's ability to lend objectivity and independent authority to a 'piece of paper' handed from one analyst to another can thus play a crucial role in anticipating such potential challenges.

When the software fails, analysts must fix it rather than decide without it

Paradoxically, it is precisely when analysts explain how they repair software output that their dependence on it becomes most glaring. 'Repair' is shorthand for analysts' use of more and less conscious heuristics to interpret ambiguous or unexpected output. Repair aligns an algorithmic result with the expert's judgment, but the persistent need for it also makes evident that analysts cannot proceed without it. Repair demonstrates that analysts *must* recruit the algorithm as co-creator of expert judgment and (re)construct its objectivity or face challenges in court.

When the algorithm provides results that are ambiguous or don't make sense to them, analysts run through a series of checks to problem-solve the discrepancy. One analyst told us,

If I think there's something in the data that shows this was not the best scientific interpretation, we have primary diagnostics: Do the results make sense scientifically? Do these results make sense given experience? Then we have secondary diagnostics: the statistical calculations that the software puts out, criteria that it has to fall within. (2)

If the output fails one or more diagnostic check, then the analysts might do any of the following. They might simply run the analysis again, or even several times. 'DNA isn't always pristine when collected at the crime scene', they explained. 'Sometimes [STRmix] has to run samples several times' (8). Additionally, they might 'reevaluate the number conditioning', namely the number of contributors to the sample, or they might 'do more PCR amplification because the algorithm likes more data' (5). Analysts anthropomorphize or personify the software—what it 'sees', what it 'likes', when it's 'having a hard time'. Even as they speak of its failures, analysts draw on the rhetorical device of *prosopopoeia*, speaking in the name of an absent entity (Bourdieu, 2014, p. 45). The entity is thereby recruited to support what otherwise would seem like mere subjective judgment: 'our primary diagnostic is "does this make sense"' (9); 'once you see something that looks correct 1000 times, when it doesn't look correct it's like, "woah"' (3).

This was the case even though, by and large, analysts described repair as needed only when the algorithm was not getting what they believed should be 'obvious'. The software might stumble over a task that appears straightforward for an analyst, like identifying a major contributor in a 2-person mixture or identifying the number of contributors from an easily legible electropherogram. It might take an inordinately long time to converge on an answer or provide incorrect results for control samples. Analysts described instances where they immediately saw the correct answer, but rather than make the call themselves, undertook elaborate procedures to ensure the software arrived there as well. One analyst described how, 'if there's one thing that's really getting under my skin that I can't figure out, I'll look at each location with the [output] next to me, I'll ask if this makes sense. I'll even make the calls on my own, does this make sense' (4). Some complained that these processes hindered their work. In one example, an analyst judged that a case involved a 2-person mixture with one major contributor, but had to reconstruct why the algorithm couldn't 'see' the same thing:

I had to talk to my supervisor, consult the manual, see what are the next steps I can take so that STRmix can find that major contributor. ... Instead of [starting the mixture proportions at] 50-50, I had it start at 90-10 ... it's not protocol to manually deconv mixtures, but once I was able to set the mixture proportions to what it actually looks like, it was able to start. (10)

Instead of deconvoluting the mixture manually, the analyst tinkered with the input settings until the software could recognize the unbalanced sample contributions. Other analysts recounted similar stories of repair work, where the analyst could 'have deconvoluted [the mixture] very easily, but STRmix could not':

I'm looking at this and realizing that STRmix is having a hard time. ... If I wasn't paying attention or it wasn't my decision, I would have been like great ... and left it at that. But I'm like no, this doesn't make sense, so I made the decision to give STRmix more time so that it can [finish analyzing the sample]. ... Sometimes it just needs more time. That's what we're trained to recognize. It can take many iterations to figure this out. I can allow STRmix to have more time. If it's still not getting it, I can tell it's this mixture's proportions. (3)

Here again, the analyst does not disregard the algorithmic output but adjusts conditions until the algorithm and analyst align.

Criminalists described other occasions when the algorithm overlooks missing information or misidentifies artifacts of the laboratory process as DNA. One recalled a time when a peak 'dropped out' of the DNA profile, leading to a mismatch, 'so we did a replicate of the sample and the allele showed up' (6). Another described removing artifacts resulting from 'an electrical surge or a piece of dust that gets detected by the laser' so as to simplify analysis for the algorithm: 'Because they're well-characterized, we know they're not DNA and can take them out. Keeping them in can confuse STRmix' (10). However, in other cases an artifact may be less clear: 'sometimes there are four peaks involved in the threshold that we set up. But you see all these other little peaks at that threshold. Is that instrument noise? A real peak? An artifact? Why is it so low that it hasn't been called as a number by the instrument?' (4). In each of these examples, the analyst disagrees with something in the technology's output, and in most, the analyst notes how they might have deconvoluted the mixture by hand more quickly. But in each case, they repair the output.

Why bother to fix it? One analyst shrugged and responded: 'I'm not in charge, I don't know' (3). A more senior analyst explained that changes to the FBI standard guidelines have made it so that legally, criminalists must have 'statistical backing' to present their findings in court; otherwise, they must say a mixture is inconclusive. Before the technology, the analyst explained, criminalists in this office 'really did not do statistical comparisons':

We just looked for inclusion/exclusion. Around 2012, [the FBI] standard guidelines changed. ... Best practices became that you need statistical analysis to do the comparison. ... That's what prompted the development of [a previous program]. Otherwise, we would have to call the mixture inconclusive, we would have been unable to say anything. (5)

In other words, the algorithm appears useful and necessary not only to address interpersonal lab disputes, but also to resist legal contestation.

Mechanical objectivity and human subjectivity

The FBI's change in guidelines is a symptom of a larger trend, which explains why algorithmic statistical backing has become obligatory. The software's introduction relies on a prior problematization of human judgment as subjective and biased. Whether judges ultimately ruled PDP software admissible or not, all asserted that the adoption of proprietary software was motivated by the potential for subjective bias in expert judgment. The

very qualities that constituted analysts' claim to disciplinary objectivity became evidence for their subjectivity: 'the determination of the number of contributors relies upon the analyst's knowledge, experience, and expertise to provide his or her best estimate; it is, by definition, subjective' (*United States v. Lewis*, 2020).⁵ The software, in contrast, 'increas[es] the objectivity of the analysis, given the concern that prematurely exposing a human analyst to a suspect's profile could introduce observer bias' (*State v. Pickett*, 2021). Compared to 'subjective interpretation', STRmix mitigates 'the risks of human error, processes more potential interpretations of the mixture in less time ... [and] also mitigates the effect of cognitive bias, as the software does not know the other facets of the case' (*United States v. Gissantaner*, 2021). Once the software has been added to the organization of expert labor, once the objectivity campaign has cast the analysts as 'subjective', relying on it is no longer optional. The algorithm becomes a co-creator of the judgment.

Contested impartiality: Analysts construct themselves as neutral scientists, but the software undermines this self-presentation

Analysts construct themselves as impartial scientists through boundary-work between 'data' and 'context'

Analysts were keenly aware that their testimony weighs as a deciding factor in trial proceedings: 'The prosecution always takes DNA as the gold star, they put us on last. ... Literally, you're the smoking gun. They view it that way' (9). In one analyst's experience, jury members often care little about the scientific processes experts outline:

[The jury] definitely perks up when the DNA expert is here. They're like 'what do you have to say?' They're more engaged. Most of the time, they're interested in the end of the story. ... Not so much the testing, the quality control, or the competency testing or efficiency testing—they don't care about that. ... They want to know: Did he match? That's it. (3)

Echoing previous findings (Bechky, 2021; Kruse, 2012), analysts anticipate the adversarial value assigned to their court testimony by constructing themselves as impartial scientists. Presenting the analyst as 'an unbiased person, just trying to present scientific results' (5) relies on boundary-work between scientific facts and legal facts, between 'data' and 'context'. As one analyst said, 'We really don't care if it matches John Doe or not. We want to analyze the sample and come up with the conclusion best supported by the data' (1). Another said: 'For me personally, I'm here [in court] as a statement of a fact re: the results, not accusing an individual' (9). A third expressed the same sentiment even more forcefully:

I mostly, in my mind, try to hold steadfast to our data, this is what makes sense. ... I cannot be anything but impartial—I don't have any other feeling except for the test result. I don't care about making [the lab] look good, I don't want to be the savior. I just want to produce good science. (3)

Context, in this case, consists of the information and related judgment connecting data to culpability. Analysts insisted that understanding and digesting the context was not their job but the job of the jury:

The DNA have a context. You have to ask what is the sample, how it ties into the crime. Having no DNA doesn't mean they didn't commit a crime. Ultimately, that's for the jury to [decide]. My job is to interpret the science as best I can. I help them understand, help understand the weight of it. (5)

Another explained how they restrained their testimony when asked to speak to a sample's context:

We can say whether something is more *likely*—we wouldn't expect a major profile on a gun to be from someone who never touched the gun just based on common sense on how it works. Mostly I try to stay away from those types of hypotheticals, because if you're going to say it's possible for this then you have to say it's possible for other things as well. (6)

One even described how they intentionally try not to learn any context about the case at any point so that they can't possibly be biased by that information (9).

At other times, analysts admitted that this boundary-work between data and context did not reflect workplace realities but cited pragmatic reasons for maintaining it. One explained, 'Context does matter, but there are degrees to which it matters. We try to look at it with an unbiased eye, but sometimes that's just not possible because it would take forever' (1). For example, even a basic parameter such as number of contributors is a matter of context. Analysts acknowledged, therefore, that they depend on police reports to interpret the evidence. They expressed frustration at how often reports were incomplete, biased, missed relevant information, or differed from police testimonies in the courtroom (5, 8, 10). They also mentioned occasional pressure from prosecutors to word their testimony in specific ways (8) or acknowledged that they are implicitly trained to view the defense as their 'nemesis' (3). None of this, however, caused them to revise their self-image as impartial scientists or their claim that their 'job is to explain what the data says regardless of who is there. The data is what's telling the story' (2).

The algorithm threatens boundary-work and impartiality

The algorithm's introduction destabilizes analysts' self-presentation and the boundary-work upon which it rests. When analysts calculated LRs by hand, they could explain why they made particular decisions and how they chose the assumptions underlying the calculation. When done with software, the analyst cannot explain many of the decisions blackboxed in the code, undermining their authority as an impartial data spokesperson. As one analyst lamented, 'I almost feel like if anything, [the lawyers] try to make us seem like we're just technicians, and not really experts because we're not doing the manual labor of deconvolution and the locus by locus calculating of the ratio' (6). To reassert their expertise, analysts refer to lab protocol:

It's not that we ever doubt what we do. But if you have someone coming at you and they're so intense, their sole intent is to try to discredit us We have a huge, long list of why we do—, why we use the methods that we do, why we have the protocols to figure out the procedures. (4)

Yet, to convey this 'huge, long list' under cross-examination is not only difficult, but opens up the testifying analyst to new challenges.

For example, lab protocol dictates that analysts present only one ratio in the courtroom, despite that each run will produce a slightly different LR. This especially irked the defense lawyer we interviewed: 'Only one number gets reported in court under one set of assumptions. And that just seems absolutely scientifically wrong ... if you're gonna make the factfinder be the Bayesian decision-maker or whatever, you have to provide more information' (DL). While one judge decided that this had no bearing on the matter of admissibility, he expected that it 'would add another arrow in the quiver of defense counsel that is going to be used to undermine the STRmix results when the issue is presented to the trial jury' (*People v. Bullard-Daniel*, 2016). The analyst's *prosopopoeia* of 'the data is what's telling the story' thereby becomes more tenuous. Even though LRs were introduced long before PDP software, they have suddenly become 'a thing that everyone's attention is focused on' to a larger extent than in previous litigation (DL).

Even more fundamentally, analysts run the sample under only one set of assumptions chosen by the lab. In one case, a defense expert disagreed on the number of contributors to the sample, but the lab refused to rerun the analysis based on scientific disagreement: 'Even the software developers say that you cannot know the number of contributors with certainty, even [the lab] will admit that. So if that's true, then why is the next tack not to run under a reasonable different number of contributors?' (DL). Indeed, some judges, determining that the software involved is a 'black box' that 'is not open to the public, or to defense counsel', have notified forensic laboratories that they may rule analysts' testimonies inadmissible if labs report algorithmically-calculated LRs without permitting the defense to run the analysis under an alternative set of assumptions (*People v. Collins*, 2015). Prosecutors, too, have challenged the analysts' *prosopopoeia*, and in one case recounted to us 'shopped around' between competing software programs to obtain an LR that best supported their case (DL).

Forensic labs seek to anticipate these challenges by carefully controlling the language that analysts use when presenting LRs. But this creates its own problems, as was dramatized in another episode the defense lawyer recounted. An otherwise personable and competent criminalist, testifying for the prosecution, froze on the stand and was reduced to reading 'the results from start to finish on the report. And it was the most boring thing you have ever heard. ... And she would never answer the prosecutor's questions simply. Which I'm a little sympathetic to, you don't want to be inaccurate. But she was so terrified of going off script that ... she just read the same script 20 times' (DL). The lawyer was clear that this situation was caused by the addition of the algorithm. While in the past a testifying analyst might 'have to look at the report to get the number', now, however, there is increased scrutiny on conveying uncertainty. 'The specific wording of the likelihood ratio, its like an incantation. ... They're so terrified of getting it wrong that they just read it', and as a result, 'sound like an automaton when they testify' (DL).

Analysts further avoid discussing the algorithm directly and, wary that it may bring what they see as unwarranted doubt, explain that: ‘the models are based on many studies done before probabilistic genotyping even existed. It already existed before but now it’s just in a software’ (6). However, the LR calculation is also shaped by assumptions and decisions built into the software and source code that analysts may be unaware of or unable to explain.

Contested expertise: ‘Is your accuser me, or is it the software?’

In the opinion of TrueAllele’s developer, the ‘main barrier’ to the algorithm’s broad acceptance is ‘the difficulty in educating analysts on the system to the degree they would be able to testify in court’ (*People v. Wakefield*, 2022). At the laboratory we studied, analysts are trained in forensics through a combination of undergraduate and/or graduate training and a mandatory in-house training program, which is run by rotating teams of experienced analysts. The lab’s training includes oral examinations about output interpretations, as well as mock court testimonies where analysts prepare for what’s ‘intended to be the hardest testimony that [you ever do]’ (4). Even so, ‘it’s kind of hard because you can never really cover the circumstances you might see. For STRmix specifically, it’ll give you the info if something is wrong with your results, something doesn’t make intuitive sense, but you cannot cover every possibility [in training]. But you are trained to be like, “well this is not what I qualitatively expected”’ (10). While training develops the analysts’ intuition, they are not trained as statisticians or software engineers.

Attention to backgrounded human judgments brings attention to the ‘numerous subtle choices made by programming developers regarding how to interpret input data’ (*State v. Pickett*, 2021). Forensic laboratories struggle to account for algorithmic results that seem to involve systematic error, bias or discrepancy. In one example, a change to the software’s source code increased the LR by a ‘10 orders of magnitude, 15 orders of magnitude, difference’. The defense lawyer we spoke to said that when questioned by the defense, the software company explained: “‘It’s not a software bug! Don’t worry, it’s NOT a bug.” And they’re like, “It’s a modeling decision.” But when I heard that, I’m like, if it’s a modeling decision, why are you discarding it?’ (DL).

Concerns about lack of transparency have opened a dispute over whether the relevant scientific community for a *Frye* hearing should also include computer scientists and software engineers. The previous DNA wars resolved debates over expert witnesses in favor of strict boundary-work that limited the circle of relevant expertise mostly to forensic science (Aronson, 2007, pp. 121, 171–172). The introduction of PDP software reignited this struggle. While one court did not consider computer scientists or software developers to have relevant expertise (*United States v. Lewis*, 2020), another formulated the *Frye* question as about general acceptability in ‘the computer science community, which is one of the disciplines’ (*State v. Pickett*, 2021).

The potential widening of relevant expertise inevitably impinges on criminalists. One analyst discussed how these tensions lead to challenges in court:

I've been challenged for expert status on statistics. I'm not a statistician, I'm testifying to how they relate to PDP software. So we've had challenges of where my expertise ends. ... "You're presenting stats, how are you an expert in statistics?" "You're presenting statistics from a software, how are you an expert in the software when you didn't develop it?" (9)

An analyst must not only display a command of mathematics but also of programming, and of all the prior decisions and judgments encapsulated in the software's source code. Meant as a decision-support tool lending mechanical objectivity, the software ends up destabilizing analysts' performance of expertise.

All these issues—the fact that the algorithm performs the calculation, presenting only one number under one set of assumptions, the refocused attention on backgrounded design choices, the software design expertise needed to audit blackboxed algorithms—were encapsulated for analysts in their greatest concern, challenges to their testimony based on the Constitution's Confrontation Clause. The 'black box' nature of proprietary software has implications for understanding analyst decision-making regarding the 'right to face your accuser':

Is your accuser me, or is it the software?... I make all the final calls and I make the final decision. It's something that I hammer home because this is the issue, but it's a truth and fact that it's not just the software doing it. There's a misconception—I'm not going to say misconception—there's a push for saying the software tells us everything because it's helpful re: right to face your accuser. (9)

The prospect of being challenged in this fashion leads this analyst to reaffirm their final decision-making authority. At least some judges, however, were not convinced. The Confrontation Clause requires, they argued, that the code be open for public scrutiny:

A human analyst can be asked why they decided a peak was or was not present, or why they believed they could not make a reliable determination of the existence of a peak based on the available data, TrueAllele uses statistical modeling to 'infer[] the probability that the peak is present'. Without the source code, no independent third party or defendant could challenge TrueAllele's assumptions for what is essentially a mathematical guess—a computer-run, theoretically-based conclusion, but still a guess. (*People v. Wakefield*, 2022)

At least some courts have ruled that the admissibility of PDP evidence must be conditioned on the software and source code being made available to defense experts' scrutiny (*State v. Pickett*, 2021). Another referred to this analysis as 'sound, illuminating and persuasive' in justifying its decision to deny admissibility (*State v. Rochat*, 2022). A two-judge majority of the New York Court of Appeals, on the other hand, refused to consider the source code as a declarant and ruled for admissibility based on the developer's testimony. The third judge on the panel, however, vehemently dissented, arguing: 'Although a computer cannot be cross-examined ... the computer does the work, not the humans, and TrueAllele's artificial intelligence provided "testimonial" statements against the defendant as surely as any human.' He concluded in terms that precisely captured the analysts' fears: 'The algorithm makes the critical decisions and a human being is an "analyst" in name only' (*People v. Wakefield*, 2022).

The courts' language in making such determinations emphasizes the extent to which the software's incorporation into analysts' work processes undermines their performance of impartiality and expertise. While PDP software lends mechanical objectivity to analysts' interpretation of complex DNA mixtures, it also leaves them exposed to the challenge that they are analysts 'only in name'.

Discussion and conclusion

Reignited controversy around what was previously settled jurisprudence on PDP evidence's admissibility indicates that the courts are grappling with a question very much related to our own: What happens when an algorithm is added to the work of an expert group? The evolving literature on algorithms in expert work settings led us to expect that PDP software would not replace DNA experts, but would likely lead to new distributions of work, authority and cognition. However, this process of re-distribution is likely to be shot through with unresolved tensions. Algorithms are typically introduced based on a problematization of human judgment as subjective and biased. Yet, algorithms do not eliminate such judgment but merely background it. By the same token, however, if what is backgrounded becomes a focus of attention once again, it is harder to defend. We expected analysts, expert witnesses, judges, and lawyers, therefore, to be grappling with how to frame these tensions and contradictions. This is indeed what we found. PDP software allows analysts to deconvolute samples that are too low-level or complex for previous methods and it provides 'statistical backing' that legitimates their judgments. In this sense, PDP software has buttressed experts' authority, allowing them to weigh in on cases that would have been impossible in the past. Yet, it was introduced as part of an objectivity campaign that problematized analysts' judgment. When judicial attention turns to the judgments blackboxed in the software, analysts find themselves in a dilemma. They can no longer proceed without the statistical backing it provides, but they also cannot fully explicate and defend the choices built into the software.

We show how Bechky's (2021) culture of anticipation dictates that criminalists anticipate judicial scrutiny throughout the analysis process. This is why, given the tensions we describe, analysts kept telling two somewhat distinct stories about their relationship to the algorithm. In one, analysts maintain full control over the decision-making process, performing boundary work to assert their position as neutral scientists. In the other, they laud the software's ability to expand the types of cases they can analyze, defer to it as a third party in disputes, repair its output to enhance objectivity for the courtroom, yet constantly worry that its opacity opens them up to new challenges that contest their authority.

Though early DNA forensics debates temporarily resolved by attributing expertise strictly to forensic scientists, PDP software has resurfaced arguments over the relevant expert communities. Many admissibility hearings argue for the inclusion of computer scientists and software engineers, who are better prepared to testify to the backgrounded human judgments embedded in the software interpretations. Meanwhile, forensics experts direct the jury's focus away from the algorithm and toward the uncertain connection between the DNA and the context—that is, the uncertainty not in forensic analysis but in legal argumentation.

Finally, we would like to underscore the significance of analysts repairing algorithmic output and performing *prosopopoeia*, or the act of speaking in the name of the algorithm. This phenomenon reveals that just as human analysts are constructed as subjective and biased by objectivity campaigns, so the algorithm too must be endowed with the property of objectivity. Algorithms do not come into the world assuming the mantle of objectivity: They must be constructed as such. When the software does not process evidence correctly or when the algorithm and analyst disagree, the analyst must repair it. The anticipatory culture of the lab has incorporated the critique of subjective expert judgment, so failure to repair algorithmic output is impermissible: A clash between analyst and algorithm can open a Pandora's box of questions regarding the subjective choices incorporated into the software. Defense experts testified, for example, that allele dropout rates used to set the software's parameters were 'arbitrarily lowered ... below empirically observed rates' (*State v. Rochat*, 2022), that the 'numerous subtle choices ... regarding how to interpret data' incorporated into the software could skew the results (*State v. Pickett*, 2021), and that the protocol created to compensate for such potential errors does not 'yield reliable DNA profiles' (*People v. Collins*, 2015). Just as analysts require the algorithm's backing, so the algorithm requires analysts' repair to become this unassailable backing. Positioning PDP software as co-creator of the judgment, then, requires analysts to juggle maintaining their own claims to scientific objectivity with defending the required backing that the software provides.

There is ongoing debate around the role of algorithms in DNA analysis and criminal legal procedures more broadly. We expect that the tensions exposed in this article may have implications for the use of algorithms in the work of DNA analysts. The validity of algorithm-aided DNA analysis has yet to be challenged in federal court, and the ongoing legal debates hung over the analysts we interviewed.

We expect, too, that analogous tensions may arise in other fields where automation promises to supplement experts' work. This data illustrates a broader struggle for authority during the introduction of algorithmic decision-making in specialized workplaces beyond the case studied here. The ways that algorithms may work simultaneously to support and contest their authority poses a problem for experts. Their awareness of this tension and the narratives they develop to deal with it complicates arguments about the challenges of introducing algorithms into the workplace and evokes an under-examined set of questions about the relationship between algorithmic decision-making and automation. While this article investigates how these processes play out in a highly specialized environment, it draws attention to a dynamic occurring in other settings as well, and especially other fields reliant on expert judgment. Such contexts challenge us to rethink the scope conditions of expertise, human and otherwise.

Acknowledgements

We would like to thank Kiran Samuel for her work during conceptualization and conducting interviews with the authors. We would also like to thank fellow panelists and participants in the 2021 4S panel *Forensic Devices* for thoughtful comments and feedback. We are indebted to participants in the Columbia Institute for Social and Economic Research and Policy (ISERP) Sociology of

Algorithms workshop for reading through and providing comments on an early draft, and to the anonymous reviewers at *Social Studies of Science* for their detailed feedback. Special thanks are due to Jessica Goldthwaite and Tamara Giwa from the Legal Aid Society for initially sparking our interest in this project. Most importantly, we would like to thank the DNA analysts and defense lawyer who shared their experiences and knowledge with us.

Data use agreement

Court opinions are publicly available and can be found online. Per the IRB used to conduct the interviews, interview transcripts are restricted from being shared publicly. An anonymized version of the interviews may be provided by request pending IRB approval. Please contact the authors for further information.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is in part supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2036197 and the ISERP Sociology of Algorithms Workshop Grant No. 40000768. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or ISERP.

ORCID iD

Hannah Pullen-Blasnik  <https://orcid.org/0000-0002-1978-6598>

Notes

1. See *People v. Collins* (2015), where the defense points out that set thresholds for peak heights may cause information loss in cases of LCN DNA. Roth (2017, p. 2026) discusses PDP in her argument for making algorithmic judgments public: ‘programmers of probabilistic genotyping software should not be the ones to choose the level of uncertainty that prompts a system to declare a mixture of DNA ‘inconclusive’ as opposed to declaring someone a potential contributor ... or to choose their own estimate related to the frequency of certain phenomena, such as genetic markers or allelic drop-out.’
2. Whereas two-person mixtures tend to be rape and assault cases, three- or four-person mixtures are often property cases and touch samples, where DNA transfer is a higher concern.
3. Upon appeal, however, this decision was reversed (*United States v. Gissantaner* 2021). Applying the *Daubert* criteria, the court decided that STRmix was testable, even if unclear whether it was adequately tested; subject to peer review, even though only two of the fifty studies were conducted by people uninvolved in the algorithm’s development; generally accepted; and produced a sufficiently low error rate. This decision, however, was vehemently disputed by researchers affiliated with the National Institute of Standards and Technology (Butler et al., 2021), who argue that reporting an LR does not sufficiently convey the information from the evidence and is highly influenced by choices during the process (Lund & Iyer, 2017). Indeed, even as it confirmed reliability, the court did warn that reporting a single LR value could confuse the jury.
4. We thank an SSS anonymous reviewer for this point.
5. *People v. Wakefield* (2022) similarly emphasized analyst subjectivity and bias in justifying the algorithm’s adoption, citing Erin Murphy, *The Art in the Science of DNA: A Layperson’s Guide to the Subjectivity Inherent in Forensic DNA Typing*, 58 Emory LJ 489, 501–508 [2008]; Itiel

E. Dror & Greg Hampikian, *Subjectivity and Bias in Forensic DNA Mixture Interpretation*, 51 *Sci & Just* 204 [2011]; Saul M. Kassin et al., *The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions*, 2 *J Applied Rsch Memory & Cognition* 42 [2013]).

References

- Abbott, A. (1988). *The system of professions: An essay on the division of expert labor*. University of Chicago Press.
- Aerts, H. J. (2018). Data science in radiology: A path forward. *Clinical Cancer Research*, 24(3), 532–534.
- Agrawal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.
- Arjomand, N. A. (2022). Figuring ethnographic fiction. *American Anthropologist*. Retrieved 19 March 2022, from <https://www.americananthropologist.org/online-content/figuring-ethnographic-fiction>
- Aronson, J. (2007). *Genetic witness; science, law and controversy in the making of DNA profiling*. Rutgers University Press.
- Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3), 3–30.
- Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, 31(1), 78–108.
- Barrett, M., Oborn, E., Orlikowski, W. J., et al. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science*, 23(5), 1448–1466.
- Bechky, B. A. (2021). *Blood, powder, and residue: How crime labs translate evidence into proof*. Princeton University Press.
- Bellovin, S. M., Blaze, M., Landau, S., et al. (2021). Seeking the source: Criminal defendants’ constitutional right to source code, 17 *Ohio State Tech. L.J.* 1, 38.
- Bessen, J. (2015). *Learning by doing: The real connection between innovation, wages, and wealth*. Yale University Press.
- Bourdieu, P. (2014). *On the state: Lectures at the Collège de France, 1989–1992*. Polity Press.
- Butler, J. M. (2021). *Webinar on the NIST DNA Mixture Interpretation Report*. National Institute of Standards and Technology. Retrieved July 21, 2021, from <https://www.nist.gov/news-events/events/2021/07/webinar-dna-mixtures-nist-scientific-foundation-review>
- Butler, J. M., Iyer, H., Press, R., Taylor, M. K., Vallone, P. M., & Willis, S. (2021). *DNA mixture interpretation: A NIST scientific foundation review. NISTIR 8351-DRAFT*. National Institute of Standards and Technology, US Department of Commerce. <https://doi.org/10.6028/NIST.IR.8351-draft>
- Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2), 1–14. <https://doi.org/10.1177/2053951717718855>
- Cole, S. A. (2001). *Suspect identities: A history of fingerprinting and criminal identification*. Harvard University Press.
- Collins, H. M. (1990). *Artificial experts: Social knowledge and intelligent machines*. MIT Press.
- Daston, L., & Galison, P. (1992). The image of objectivity. *Representations*, 40, 81–128. <https://doi.org/10.2307/2928741>
- Davenport, T. H., & Kirby, J. (2016). *Only humans need apply: Winners and losers in the age of smart machines*. Harper Business.
- Delfanti, A. (2021). Machinic dispossession and augmented despotism: Digital work in an Amazon warehouse.’ *New Media & Society*, 23(1), 39–55. <https://doi.org/10.1177/1461444819891613>

- de Souza Leao, L., & Eyal, G. (2019). The rise of randomized control trials (RCTs) in international development in historical perspective. *Theory and Society*, 48(3), 383–418. <https://doi.org/10.1007/s11186-019-09352-6>
- Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.
- Dror, I. E., & Mnookin, J. L. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk*, 9(1), 47–67. <https://doi.org/10.1093/lpr/mgp031>
- Ekbja, H. R., & Nardi, B. A. (2017). *Heteromation and other stories of computing and capitalism*. MIT Press.
- Eyal, G. (2019). *The crisis of expertise*. Polity Press.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018) Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70.
- Galper, A. B., & Kellogg, K. (unpublished). *Local vs. central: Negotiating the boundaries between regimes of knowing* [Unpublished Manuscript].
- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48(6), 781–795. <https://doi.org/10.2307/2095325>
- Golan, T. (1995). *Laws of men and laws of nature: The history of scientific expert testimony in England and America*. Harvard University Press.
- Griesbach, K., Reich, A., Elliott-Negri, L., & Milkman, R. (2019). Algorithmic control in platform food delivery work. *Socius*, 5, 1–5. <https://doi.org/10.1177/2378023119870041>
- Irani, L. (2015, January 15). *Justice for 'data janitors'*. Public Books. <https://www.publicbooks.org/nonfiction/justice-for-data-janitors/>
- Kruse, C. (2012). The Bayesian approach to forensic evidence: Evaluating, communicating, and distributing responsibility. *Social Studies of Science*, 43(5), 657–680. <https://doi.org/10.1177/0306312712472572>
- Kwong, K. (2017). The algorithm says you did it: The use of black box algorithms to analyze complex DNA evidence. *Harvard Journal of Law & Technology*, 31(1), 275.
- Lacambra, S. J., Matthews, J., & Walsh, K. (2018). Opening the Black Box: Defendants' rights to confront forensic software. *Champion*, 28(30), 28–68.
- Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science*, 33(1), 126–148. <https://doi.org/10.1287/orsc.2021.1549>
- Liker, J. K., Haddad, C. J., & Karlin, J. (1999). Perspectives on technology and work organization. *Annual Review of Sociology*, 25, 575–596. <https://doi.org/10.1146/annurev.soc.25.1.575>
- Lund, S. P., & Iyer, H. (2017). Likelihood ratio as weight of forensic evidence: A closer look. *Journal of Research of National Institute of Standards and Technology*, 122(27), 1–32. <https://doi.org/10.6028/jres.122.027>
- Mindell, D. A. (2015). *Our robots, ourselves: Robotics and the myths of autonomy*. Viking.
- Nelson, A. J., & Irwin, J. (2014). Defining what we do—All over again: Occupational identity, technological change, and the librarian/internet-search relationship. *Academy of Management Journal*, 57(3), 892–928.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pasquale, F. A. (2015). *The Black box society: The secret algorithms that control money and information*. Harvard University Press.
- President's Council of Advisors on Science and Technology. (2016). *Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Report to the President.

- https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf
- People v. Bullard-Daniel, 54 Misc.3d 177 (2016).
- People v. Carter, 50 Misc.3d 1210(A) (2016).
- People v. Collins, 49 Misc.3d 595 (2015).
- People v. Wakefield. N.E.3d — (2022).
- People v. Williams, 35 N.Y.3d 24 (2020).
- Pine, K. H., & Liboiron, M. (2015). The politics of measurement and action. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, New York, NY, pp. 3147–3156. New York, NY: ACM.
- Porter, T. M. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton University Press.
- Ribes, D., & Jackson, S. J. (2013). Data bite man: The work of sustaining a long-term study. In L. Gitelman (Ed.), *Raw data is an oxymoron* (pp. 147–166). MIT Press.
- Roth, A. (2017). Machine testimony. *The Yale Law Journal*, 126(7), 1972–2259.
- Sachs, S. E. (2020). The algorithm at work? Explanation and repair in the enactment of similarity in art data. *Information, Communication & Society*, 23(11), 1689–1705. <https://doi.org/10.1080/1369118X.2019.1612933>
- Schultze, U. (2000). A confessional account of an ethnography about knowledge work. *MIS Quarterly*, 24(1), 3–41.
- Shestakofsky, B. (2017). Working algorithms: Software automation and the future of work. *Work and Occupations*, 44(4), 376–423. <https://doi.org/10.1177/0730888417726119>
- Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Wiley.
- Stark, D. (2022). Questioning humans versus machines: Artificial intelligence in class conflict. Review of Frank Pasquale's new laws of robotics. *Administrative Science Quarterly*, 67(3), NP42–NP46. <https://doi.org/10.1177/00018392221095682>
- State v. Pickett, 466 N.J.Super. 270 (2021).
- State v. Rochat, 470 N.J.Super. 392 (2022).
- Thompson, W. C., & Shumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, 2(3), 167. <https://doi.org/10.1007/BF01044641>
- United States v. Gissantaner, 417 F. Supp. 3d 857 (W.D. Mich. 2019).
- United States v. Gissantaner, No. 19-2305 (6th Cir. Mar. 5, 2021).
- United States v. Lewis, 442 F.Supp.3d 1122 (2020).
- van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, 45(3), 1557–1580.
- Vertesi, J. A., Goldstein, A., Enriquez, D., Liu, L., & Miller, K. T. (2020). Pre-automation: Insourcing and automating the gig economy. *Sociologica*, 14(3), 167–93. <https://doi.org/10.6092/issn.1971-8853/11657>
- Waardenburg, L., Sergeeva, A., & Huysman, M. (2018). *Hotspots and blind spots: A case of predictive policing in practice* [Conference session]. Working Conference on Information Systems and Organizations (IS&O) (pp. 96–109).
- Walker, C. R. (1958). Life in the automatic factory. *Harvard Business Review*, 36(1), 111–119.
- Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Basic Books.

Author biographies

Hannah Pullen-Blasnik is a Ph.D. Candidate in Sociology at Columbia University. Her work examines the criminal legal system and its relationships to technology and urban development to understand how power gets regulated and contested.

Gil Eyal is a Professor of Sociology and Director of the Trust Collaboratory at INCITE at Columbia University. He is the author, most recently, of *The Crisis of Expertise* (Polity 2019) and co-editor of *The Oxford Handbook of Expertise and Democratic Politics* (OUP 2023).

Amy Weissenbach is a Ph.D. Candidate in Sociology at Columbia University and holds an M.Phil. in History, Philosophy, and Sociology of Science, Technology, and Medicine from the University of Cambridge. She is interested in how people tell stories about themselves in moments of transition and crisis.