Berkeley Journal of Criminal Law

Volume 24
Issue 1 *Spring 2019*

Article 4

2109

No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials

Bess Stiffelman Esq.

Recommended Citation

Bess Stiffelman Esq., *No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials*, 24 Berkeley J. Crim. L. 110 (2019).

Link to publisher version (DOI)
https://doi.org/10.15779/Z384Q7QQ6X

No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials

Bess Stiffelman, Esq.†

DOI: https://doi.org/10.15779/Z384Q7QQ6X

† Public Defender and Staff Attorney at the Legal Aid Society in New York. I want to thank all of my colleagues and friends who provided assistance and support, with special thanks and gratitude to Jessica Goldthwaite and Terri Rosenblatt, Staff Attorneys in the Legal Aid Society of New York's DNA Unit; Chris Flood, of the Federal Defenders of New York; and Sari Kisilevsky, Associate Professor at Queens College CUNY. Thank you for your interest, insight, and guidance.

ABSTRACT

DNA has long been considered the gold standard of forensic evidence, heralded for its ability to exonerate the innocent and convict the guilty. But this new generation of DNA evidence is far from its established predecessor — both in the quality of the evidence collected and the clarity of what is presented in court. With new highly sensitive technology, tiny amounts of DNA, often just a few cells, are now collected at crime scenes. This DNA was left on objects by someone who touched an object, or by someone who touched or was touched by someone who then touched the object, or even further removed. In other words, cells travel and can be easily transferred to an object without the person who was the source of that DNA ever having come into contact with the object. But this is only the beginning of the difference. These crime scene samples often contain multiple people's DNA, and there is degraded and missing information. The samples are incomplete and mixed together. In order to give some evidentiary weight to these complicated and incomplete mixtures, crime labs are turning to something called probabilistic genotyping. These computer programs generate something called a likelihood ratio. These likelihood ratios purport to express the probabilistic relationship between two hypotheses, the hypothesis that the suspect is in the DNA sample compared to the hypothesis that the suspect is not included in the sample. In this paper, I first explain what a likelihood ratio is, how it is generated, and some of the fundamental problems with the evidence. Then I turn to an analysis of the propriety of this evidence in criminal trails. It is my position that because likelihood ratios can only be generated by first presuming guilt (inclusion), they undermine the presumption of innocence, and that, by weighing these hypotheses equally, they water down the burden of proof beyond a reasonable doubt in a criminal trial. This is complicated by the sheer power of the DNA moniker and opacity of the numbers generated.

INTRODUCTION

Something is happening with DNA evidence. Long considered the gold standard in forensic science, it is some of the most powerful evidence a jury can hear. But this new generation of DNA evidence, often from just a few skin cells obtained from objects at crime scenes, bears little similarity to its established predecessor.[1] The science has advanced such

[1] In 2016, The President's Council of Advisors on Science and Technology (PCAST) – an organization of scientific leaders were called upon by President Obama to examine the

that these much smaller amounts of DNA can be viewed, but the samples are often of poor quality, degraded, and hard to interpret. The probative value of the evidence is far from clear. In an effort to quantify the significance of this evidence, the current trend is to estimate a Likelihood Ratio (LR).[2] A Likelihood Ratio is determined by comparing two hypothetical explanations for the evidence. Computer programs using complex algorithms are being used to generate these LRs. The programs are designed to weigh the probability that the sample is either missing information or contains incorrect or misleading information, not truly from the DNA source, and the likelihood of the two hypotheticals given this uncertainty. The hypotheses proposed and compared appear to answer the question the jury is trying to answer, is the defendant the source of the DNA. This is misleading as LRs are not in fact probabilities that the defendant is or is not a contributor to the sample. They merely weigh the relative likelihood of two very specific hypotheses. By purporting to represent the relative likelihood of the lab's proposed defense hypothesis against the proposed prosecutor's hypothesis, LRs usurp the jury's function in a criminal trial, conflict with the presumption of innocence, and undermine the requirement that the prosecutor bears the burden of proving each and every case beyond a reasonable doubt.

In Part I, I explain the history and context of this new way of reporting DNA evidence, what problem it is trying to solve, contextualize it within the history of DNA evidence, and provide a quick snapshot of the state of litigation. In Part II, I analyze this evidence in the context of the history of probabilistic evidence in criminal trials and explain the

---

state of forensic science from the perspective of esteemed scientists to answer the President's question: "[are] there are additional steps on the scientific side . . . that could help ensure the validity of forensic evidence used in the Nation's legal system?" *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY, FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF FEATURE-COMPARISON METHODS x-xi, (2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (hereafter "PCAST"). Probabilistic genotyping systems, the type of DNA testing discussed infra, were reviewed. By separating the two types of DNA analysis, the PCAST scientists could compare them as separate disciplines and contrast their relative reliability and efficacy. The contrast was stark, and the recommendations for a "path forward" was clear: *complex mixture analysis is a unique, stand-alone forensic method, distinct from the "gold standard" of the single source comparisons lauded in the 2009 NAS report. See id.* at 79-85.

[2] LRs are not only or even most often used in reporting forensic DNA results. My criticism here is limited to the use of LRs in reporting the results of forensic DNA testing in criminal cases.

inconsistency of such evidence with the burdens of proof and presumptions in a criminal trial.

I. WHAT ARE LIKELIHOOD RATIOS AND HOW ARE THEY BEING USED IN CRIMINAL CASES

In a criminal trial a defendant is presumed innocent until a jury finds her guilty beyond a reasonable doubt, and the prosecution bears that heavy burden of that proof. These are fundamental principles, albeit, at times, seemingly little more than aspirational. We know that jurors struggle, and often fail to hold the prosecution to that burden, and it is the rare juror that looks upon the defendant at the beginning of a trial, before hearing the evidence, as truly innocent. The mere accusations by the government have such persuasive power that it requires mental gymnastics for someone to truly believe the defendant has done nothing to bring himself to the defense table in that courtroom. But this doesn't undermine the normative import and symbolic power of the principles. In fact, it underscores the need for vigilance in preserving the normative symbolism of a jury trial. For this reason, much of criminal jurisprudence, procedure, and evidence laws are directed at preserving these fundamental burdens and presumptions. As the Supreme Court explained, the presumption of innocence is the "bedrock 'axiomatic and elementary' principle whose 'enforcement lies at the foundation of the administration of our criminal law.'"[3]

As I'll explain in more detail below, LRs are created by assuming the defendant is in fact guilty (his DNA is in the mix), and then weighing that assumption against the assumption that the defendant is not in fact guilty (not in the mixture).[4] The possibility of guilt and innocence (inclusion or exclusion) are treated equally, thus shifting the burden of proof to something more akin to a civil standard. In postulating the hypothesis, the analyst inserts a presumption of guilt into the trial process.

Unfortunately, the sheer opacity of the science and math required to produce these LRs obscures this fact from judges, lawyers, and jurors. Before turning to this argument in depth, I will quickly contextualize this new evidence, and distinguish it from what the general public, and jurors,

---

[3] In re Winship, 397 U.S. 358 (1970).

[4] This is, of course, a somewhat simplistic way to describe LRs. There will always be an inference needed to get from the defendant's DNA being in the mix to the defendant's guilt. However, in many cases, where identity is at issue, for example, as a practical matter, this is how LR evidence is received.

envision when they hear there is DNA evidence.[5]

In order to better understand some of the challenges with degraded crime scene samples, it's important to understand a few things about DNA analysis. A DNA profile is determined by looking at different locations on a genetic chain. The current standard, based on FBI protocols, is to look at 22 specific locations on the genetic chain. A person has, at most, two distinct genetic markers (alleles) at any location — one from her mother and one from her father. A person will often have the same genetic marker from both their mother and their father, so a single location on an individual DNA profile can have either one or two alleles. If there are three alleles at a location, then the sample contains DNA from more than one person.

When one traditionally speaks of a DNA match, what's described is a single source sample, usually from blood, semen, or saliva. A profile is deduced from the sample, a profile is obtained from a suspect, and a comparison is made.[6] The analyst then calculates how rare that profile is, based on databases estimating the frequency of the specific genetic markers in a given population. The analyst then testifies as to both the rarity of the profile and whether the two profiles match. The numbers are usually staggering, such that, if testing was done properly, and the population-based genetic modeling reliable, the matching suspect is almost undeniably the source of that DNA evidence sample.

It is this type of DNA evidence that has been referred to as the gold standard. It is such powerful evidence, that not only has it been used to convict the guilty, it has led to hundreds of exonerations across the country. The first of these exonerations took place in 1989.[7] Since then, the retesting of evidence excluding convicted defendants as the source of the DNA has tremendously improved our understanding of the cause of wrongful convictions, and the frequency of such troubling miscarriages of justice. For example, by studying these cases, we now know how frequently eyewitnesses misidentify the perpetrator of a crime. This has spawned greater research into the cause of such misidentification, and dramatically changed the way cases that depend on stranger

---

[5] I approach this from the perspective of a practicing public defender, in the jurisdiction that has been at the forefront of this expansion of forensic evidence.

[6] For a clear explanation of how these single source samples are analyzed, *see* PCAST, *supra* note 1, at 69-73.

[7] INNOCENCE PROJECT, DNA EXONERATIONS IN THE UNITED STATES (Mar. 6, 2019), https://www.innocenceproject.org/dna-exonerations-in-the-united-states/.

identifications are litigated.[8] These exonerations also demonstrate the surprising frequency in which people make false confessions.[9] But these DNA samples were predominantly from bodily fluids, contained higher quantities of DNA, and individual profiles could be deduced such that they were subject to that much more straightforward analysis. Most forensic evidence is not so clean. As the science progressed, labs developed the ability to view DNA samples obtained from more common forensic evidence, like objects found at crime scenes. But determining the probative value of such evidence has proved daunting.

Before turning to how LRs are being used to try to solve this problem, it is important to understand why the value or evidentiary weight of this evidence is unclear. Because of the advance in sensitivity of the technology, crime labs are testing more and more of this "trace DNA."[10] These trace samples lack the clarity of the more straightforward DNA evidence that can lead to a clear match to a specific individual. An object is found at or near a crime scene. A technician swabs the object to test for that DNA. These trace samples are usually quite small, there is often more than one person's DNA, and the evidence is of a much poorer quality.[11]

When dealing with such small amounts of DNA, there is much greater ambiguity as to how the DNA ended up on the object. For example, the DNA could have been left by someone who touched the object, or even by someone who touched the person who then touched the tested object.[12] And these aren't the only possibilities. In short, small

---

[8] *See, e.g.*, NATIONAL RESEARCH COUNCIL, *IDENTIFYING THE CULPRIT: ASSESSING EYEWITNESS IDENTIFICATION* (2014) (explaining that at least one mistaken identification was present in three quarters of DNA exonerations; summarizing the field of research spawned by this discovery; and making recommendations to law enforcement, lawyers, and courts on how to prevent such injustice).

[9] INNOCENCE PROJECT, *supra* note 7.

[10] *See* PETER GILL, MISLEADING DNA EVIDENCE; REASONS FOR MISCARRIAGES OF JUSTICE 1-3 (2014) (defining trace evidence and outlining the problems with analyzing this type of evidence).

[11] *See, e.g.*, SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS, PUBLIC COMMENTS, www.swgdam.org/#!public-comments/c1t82.

[12] A study focused on determining how frequent secondary transfer could result in DNA detected from an object swabbed. The subjects shook hands with one individual for two minutes, and then touched a knife. "DNA typing results indicated that secondary DNA transfer was detected in 85% of the samples. In five samples, the secondary contributor was either the only contributor or the major contributor identified despite never coming into direct contact with the knife. This study demonstrates the risk of assuming that DNA recovered from an object resulted from direct contact." Cynthia Cale, Madison Earll, Krista Latham & Gay Bush, *Could Secondary DNA Transfer Falsely Place Someone at*

DOI: https://doi.org/10.15779/Z384Q7QQ6X

amounts of DNA can be easily transferred and travels. Because of this, finding someone's DNA on an object is less significant to a determination of guilt or innocence of a suspect.

One interesting study explored the likelihood of DNA transfer among clothes that are washed together. After washing pristine items with one pair of semen stained underwear, fifty percent of the items washed were found to have at least one sperm cell, the source of DNA in semen.[13] The Lukis Anderson case is a terrifying real-life example of what can happen when DNA is unintentionally and unknowingly transferred. Lukis Anderson's DNA was found under the fingernails of a murder victim, and he was charged with murdering a man he had never met. Ultimately, and thankfully, he had an airtight alibi—he had been in a hospital detoxing. He had been taken to the hospital by the same paramedics that later picked up the murder victim. The DNA had been transferred by the paramedics from Lukis Anderson to the murder victim. But for this discovery, he would have been facing the death penalty for a murder he did not commit.[14] Not everyone is so lucky, and without such an airtight alibi, the source of the DNA transfer would have never been discovered. Although the possibility of transfer clearly affects the evidentiary weight of these small DNA samples, this is not accounted for when LRs are reported to a jury.

What LRs try to evaluate is the potentially missing, incomplete, or incorrect information in these small samples, and the difficulty in separating individual profiles when more than one person's DNA is in the sample. The smaller the DNA sample, the higher likelihood of interference, or noise, in the analysis. Something can show up in the sample that comes from the air, or some other source of contamination. The technology also has a certain anticipated amount of random misinformation, referred to as stochastic effects, generated by the testing process itself. These samples also lose information. Alleles drop-out, meaning one of the genetic markers of the contributor just doesn't show up in the testing process. The smaller the sample, the higher the likelihood of incorrect information showing up during the testing process, and the

---

the Scene of a Crime?, 61 J. FORENSIC SCI. 196-203 (2016).

[13] *See generally* E. Kafarowski, A.M. Lyon & M. M. Sloan, *The Retention and Transfer of Spermatozoa in Clothing by Machine Washing*, 29 CAN. SOC. FORENS. SCI. J. 7-11 (1996).

[14] For a good account of the Lukis Anderson case, *see* KATIE WORTH, FRAMED FOR MURDER BY HIS OWN DNA (The Marshall Project, ed. 2018), https://www.themarshallproject.org/2018/04/19/framed-for-murder-by-his-own-dna.

higher likelihood that information is just missing.

If there is more than one person's DNA in the sample, it can be impossible to tell how many people make up that sample.[15] Three alleles seen at a single location could be from two people who happen to have one of the same genetic markers, or one of the individuals had only one allele at the location because she received the same allele from her mother and father. It is also possible that three or four people contributed and either share an allele or had only one allele at the location. Some genetic markers are more common, just as brown and blond hair is more common than red hair. So, if there are three true alleles at a location, it has to be more than one person's DNA, but you don't know for certain how many. This problem of determining the number of contributors obviously becomes more difficult as the number of alleles seen at a location increases.

Accordingly, with this new highly sensitive technology, an analyst is able to view the sample, but she is often unable to distinguish individual profiles, determine how many people's DNA is in the mixture, and there is missing and uncertain information.[16] Until recently, in these

---

[15] *See, e.g.*, David Paoletti et al., *Empirical Analysis of the STR Profiles Resulting from Conceptual Mixtures,* 50 J. FORENSIC SCI. 1 (2005) (concluding that four-person mixtures are not recognized correctly as four-person mixtures 70% of the time when counting the maximum number of alleles at a single locus); Jaheida Perez et al., *Estimating the Number of Contributors to Two-, Three-, and Four-Person Mixtures Containing DNA in High Template and Low Template Amounts*, CROAT. MED J. 315 (2011) ("While locus-by-locus allele counting can provide an estimate of the minimum number of contributors to a mixture, it may not indicate the actual number of contributors to mixtures, particularly those with 3 or more contributors. That is, using this method, a three-person mixture could be classified as a mixture of at least 2 people and a four-person mixture could be designated a mixture of at least 3 people and sometimes as a mixture of at least 2 people.") (citing J.S. Buckleton et al., *Towards Understanding the Effect of Uncertainty in the Number of Contributors to DNA Stains*, FORENSIC SCI. INT'L GENETICS 1:20-8 (2007).

[16] *See* PCAST, *supra* note 1, at 75-76 ("DNA analysis of complex mixtures-defined as mixtures with more than two contributors—is inherently difficult and even more for small amounts of DNA. Such samples result in a DNA profile that superimposes multiple individual DNA profiles. Interpreting a mixed profile is different for multiple reasons: each individual may contribute two, one or zero alleles at each locus; the alleles may overlap with one another; the peak heights may differ considerably, owing to differences in the amount and state of preservation of the DNA from each source; and the 'stutter peaks' that surround alleles (common artifacts of the DNA amplification process) can obscure alleles that are present or suggest alleles that are not present. It is often impossible to tell with certainty which alleles are present in the mixture or how many separate individuals contributed to the mixture, let alone accurately to infer the DNA profile of each individual.").

circumstances, an analyst could only report that an individual suspect could or could not be excluded. The comparison was inconclusive.

Labs have been searching for ways to give some sort of quantitative evidentiary weight to these samples.[17] Crime labs and private software developers have devised different methods, but all to that same end— to find a way to attach numerical evidentiary value to these complicated mixtures of DNA. What is currently being generated by many labs is a likelihood ratio.[18]

A. What Are Likelihood Ratios?

A likelihood ratio (LR) compares the probabilities of two different hypotheses that seek to explain a given piece of evidence. In the context of probabilistic genotyping in forensic science the two hypotheses are most simply and generally, given this piece of evidence, the likelihood that the suspect is the source of some of the DNA in the mixture versus the likelihood that the suspect is not the source of some of the DNA in the mixture. The probabilities used in the formula are always expressly conditional, as in, "given this piece of evidence, it is x times more likely that the suspect is the source of the DNA than it is likely that an unknown unrelated person is the source." Let "Pr" stand for probability, "S" for suspect, and "U"s for unknown individual. Its formulaic expression is, thus: LR = Pr(Evidence/S)/Pr(Evidence/U).[19] Although LRs can be produced for single source samples, the more straightforward RMP method, discussed *infra*, is generally applied. Any two hypotheses can be compared and will generate an LR. No matter how improbable a hypothesis, one will always be reported as more probable than another. Accordingly, nothing about the LR tells you how objectively likely either hypothesis in fact is. This is one of the great dangers of the evidence, and something that jurors will be hard pressed to understand. The number is easily, and most likely taken as the probability of the defendant's guilt.

To be clear, this is a grave mistake. LRs are not, in fact,

---

[17] *Id.*

[18] *Id.* at 78-81.

[19] The Office of the Chief Medical Examiner in New York includes an explanation of an LR on the last page of the report generated when results are given on a case. Their exact language follows: "Likelihood ratio (LR) – A statistical measurement of the strength of support for one hypothesis over another. For example, this would be reported as 'The DNA mixture found on evidence sample is approximately LR times more probable if the sample originated from hypothesis 1 than if it originated from hypothesis 2. Therefore, this supports that reference DNA profile is [included or excluded]as a contributor to this sample.'"

probabilities. They merely express the probabilistic relationship between two hypotheticals.[20] Another step must be taken to convert an LR to a relevant probability, although that is surely something that most judges, lawyers, and jurors will have trouble understanding. As described by the National Research Council, in The Evaluation of Forensic DNA Evidence, there is a great risk that a jury will misconstrue the LR as a "statement of the odds in favor" of the hypothesis that the defendant is the source of the DNA evidence.[21] In order to convert an LR into a probability, a Bayesian analysis has to be applied. Bayes' theorem describes the process by which the probability of an event is combined with new data to produce a posterior probability.[22] In order for an LR to answer anything meaningful to a jury, the juror must first postulate a prior probability of guilt. Again, as described by the National Research Council, "[t]he likelihood ratio is still one step removed from what a judge or jury truly seeks — an estimate of the probability that a suspect was the source of a crime sample, given the observed profile of the DNA extracted from samples."[23] LRs are, thus, fundamentally different from other forms of forensic evidence in that they are only able to provide meaningful information once a Bayesian analysis has been applied. They do not express the probable likelihood of the truth of either hypothesis, or the probability that the defendant is the source of the DNA.

An example that doesn't involve DNA might help clarify this point. One could propose any two hypotheses to answer a question, and theoretically come up with a likelihood ratio. Suppose you returned home to find your dog sitting next to a torn-up pillow, and feathers are everywhere. You could compare the hypothesis that your home was ransacked by burglars with the hypothesis that the pillows you bought at Ikea were designed to explode after six months. After gathering all the relevant data, like that your door was locked when you came home, and nothing else was broken, and the strength of the pillow fibers, you could come up with a likelihood ratio. Let's assume you determine the likelihood ratio that, given the evidence, it was 10,000 times more likely

---

[20] *See* PETER GILL, MISLEADING DNA EVIDENCE; REASONS FOR MISCARRIAGES OF JUSTICE 17, 18 (2014) (defining a likelihood ratio in the context of DNA evidence and citing D.J. BALDING, WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES 22-42 (2005)).

[21] NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE: AN UPDATE 201 (1996).

[22] *Id.* at 31-32, 132-133 (explaining the common logical fallacies made when trying to understand LRs and how Bayes' Theorem is needed to answer the real questions before the court).

[23] *Id.* at 201.

that the pillow just fell apart than it was likely your home was ransacked by burglars. Neither of these hypotheses are in fact the correct explanation for the evidence before you. Of course, your dog played with your pillow like a chew toy and tore it to shreds. So, although it was 10,000 times more likely the pillow just fell apart than it was likely your home was ransacked, this doesn't prove your pillow fell apart on its own. Both of the hypotheses were in fact incorrect.

In the context of forensic DNA evidence, LRs are designed to solve the problem of how to make otherwise inconclusive evidence meaningful. They are primarily used to explain complex mixed samples containing multiple individual's DNA. In these mixed samples, the hypotheses proposed are, for example, "given this piece of evidence, it is x times more likely that the sample came from the suspect and two unknown unrelated individuals than if the sample came from three unknown unrelated individuals." Notice, again, the goal is to find a way to give some prosecutorial value to evidence that would otherwise be reported as inconclusive. And, in order to generate that LR, the analyst tests the assumption that the suspect is in fact a contributor to the mixture. The question being asked is one inference away from the question the jury is trying to answer, "is this the person who committed this crime?"

The following hypothetical should make this a little more concrete. I'll return to this hypothetical later. This is a common sort of scenario, but is not taken from any particular case. A car is pulled over with four African American teenagers, all of whom happen to be of Caribbean descent. The car is searched and a gun is found in the trunk. All four are arrested. The gun is sent to the lab to be swabbed for DNA. The lab reports that a mixture of DNA is found, but that they cannot separate out any individual profiles. They believe there are three or more individuals in the mixture, but they cannot be certain how many. However, the lab reports that the sample is suitable for comparison to the suspects' DNA. A court then orders DNA swabs of all four codefendants. The lab then begins the process of producing a number that purports to weigh the probability, for each of the defendants, that his DNA is in the mixture against the probability that his DNA is not in the mixture. First, the lab develops each of the defendants' DNA profiles. Then, they run each profile through a computer program.

The analyst must propose two specific hypotheses in order to enter the data into the program that produces the LR. In this case, they propose that the DNA found on the gun came from three unknown, unrelated individuals, versus the hypothesis that the DNA sample from

the gun came from the defendant, and two unknown unrelated individuals. They do this for each defendant. Note that the results would be different if the analyst hypothesized that the individuals in the mixture are related or hypothesized a different number of contributors to the mixture. These are things that are uncertain. The hypothesis proposed will have a great impact on the numbers generated, and thus determine how inculpatory or exculpatory the value of the evidence will be reported to the jury.

The program then produces numbers for each of the four codefendants. In this hypothetical, let us assume it is reported that, given the DNA mixture found in the evidence sample, it is 1,000,000 times more probable that the sample originated from Defendant 1 and two unknown unrelated individuals, than three unknown unrelated individuals. The number, with the same hypothesis, for Defendant 2, is reported as 10,000 times more probable that it originated from him and two unknown unrelated individuals than three unknown unrelated individuals. For Defendant 3, it is reported that it is 50 times more probable that the sample originated from the defendant and two unknown unrelated individuals than three unknown unrelated individuals. Finally, as to Defendant 4, it is reported as 10,000 times more probable that three unknown unrelated individuals contributed to the mixture than the defendant and two unknown unrelated individuals. Different labs and programs also have quite varying standards for explaining the significance of these numbers. When the Office of the Chief Medical Examiner in New York used the FST program, the guidelines they provided described an LR of *precisely* the number 1 as no conclusions, a range of 1 to 10 as providing limited support for the hypothesis, 10 to 100 as moderate support, 100 to 1,000 as strong support, and greater than 1,000 as very strong support.[24] Now that the lab has changed to STRmix, they report any LR under 1,000 as statistically insignificant and therefore uninformative.[25] This difference in reporting has nothing to do with the sensitivity of the program itself, STRmix is a more sophisticated program. But, rather, this change reflects a difference in the program designer's position as to the statistical significance of LRs generally.

Not only is there disagreement about the significance of these

---

[24] This description of the statistical significance of the LR reported by the OCME when they used FST was provided at the end of every lab report.

[25] *See* NYC OFFICE OF CHIEF MEDICAL EXAMINER, FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS 28, http://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/protocols-for-forensic-str-analysis/str-results-interpretation-powerplex-fusion-and-strmix.pdf.

DOI: https://doi.org/10.15779/Z384Q7QQ6X

numbers, you can easily see that there are alternative potential hypotheses that could explain the evidence, just as there was an alternative hypothesis with our pillow example. Some of the individuals could in fact be related. More than one of the codefendants could have left DNA on the gun. And the number of potential contributors could be different. Furthermore, there is nothing inherent in the modeling that requires the defense and prosecutor's hypotheses to postulate the same number of contributors, or the same assumptions about relatedness.

Not only are there multiple alternative hypotheses that could explain the evidence, but both hypotheses in the equation could in fact be wrong, and there would still be an LR reported. The number produced is the *relative likelihood* of two specific hypotheses, not the probability that either hypothesis is in fact correct. I again return to the pillow example. Your house was neither ransacked by burglars nor did your pillows spontaneously destruct.

The question of relatedness is far from straightforward. The LRs are produced, in part, based on statistics about the frequency of genetic markers. These frequencies are population based.[26] Family members will see different frequencies of certain genes, just as any discrete population will see different frequencies of certain genes — just as Norwegians will have a higher frequency of blue eyes than, say, Koreans, those of Caribbean descent will see a higher frequency of certain genes than those who are entirely unrelated. Thus, an LR will be different if the technician hypothesizes that the source of the sample was the individual suspect and two unknown individuals of Caribbean descent.[27] We don't know how different.

Additionally, different LRs would be produced if the technician hypothesized that any of the two, three, or four of the individuals in the car contributed DNA to the mixture. There is nothing inherent in the modeling that prevents the technician from supposing more than one of the known suspects contributed DNA to the mixture. In selecting a hypothesis, the analyst proposes a specific defense theory to explain the evidence, without the defendant.

To date, there are at least eight different probabilistic genotyping software programs in use by a small number of labs in the United States.[28]

---

[26] *See* RESEARCH COUNCIL, *supra* note 21, at 89-102 (providing a discussion of the population genetics and databases used to generate these frequencies).

[27] *See generally id.* at 102-27 (examining subpopulation statistics and issues of relatedness).

[28] Hannah Kelly et al., *A Comparison of Statistical Models for the Analysis of Complex*

In New York State, where I practice, as of June of 2018, three different probabilistic software programs have been proffered in different courts, each of which use different methodologies.[29] The statistical, biological, and computer models used by the programs differ greatly.[30] Unsurprisingly, different programs using different models will produce different results.

As defense attorneys, we have no way to know what those LRs would be or how they would change if a different hypothesis were postulated, or how the results would change from program to program. The programs themselves are largely proprietary, and the actual algorithms used to create the LRs are often unknown.

Defense attorneys have been moving for access to the source code for these programs so that our own experts can evaluate the programs, so we can run the data ourselves, and, potentially explore the results generated under the different hypotheses that could explain the evidence. These applications have, until recently, been largely denied. After years of denied applications at the state level, a federal judge finally permitted a defense team to review the source code for FST, the program used for years by the Office of the Chief Medical Examiner in New York (OCME). The review was only permitted under a protective order. "Nathaniel Adams, a computer scientist and an engineer at a private forensics consulting firm in Ohio, reviewed the [FST] code for the defense. He found that the program dropped valuable data from its calculations, in ways that users wouldn't necessarily be aware of, but that could unpredictably affect the likelihood assigned to the defendant's DNA being in the mixture."[31] Shortly after this revelation, the OCME switched

---

Forensic DNA Profiles, 54 J. SCI. & JUST. 66 (2014) (quoting from the abstract that "there is no consensus within the forensic biology community as to how [complex mixtures and small DNA samples] should be interpreted."). There still is no agreement within the scientific community about which, if any, probabilistic software programs or methods to employ when analyzing low template DNA or complex mixture samples.

[29] *See* People v. Wakefield, 47 Misc. 3d 850 (N.Y. Sup. Ct. 2015) (considering the use of the TrueAllele probabilistic software program); *see also* People v. Rodriguez, Ind. No. 5471/2009 (N.Y. Sup. Ct. 2013) (considering the use of the FST probabilistic software program); *see also* People v. Bullard-Daniel, 54 Misc. 3d 177 (Niagara Cty. Ct. 2016) (considering the use of the STRMix probabilistic software program).

[30] *See* Kelly et al., *supra* note 28 at 66-70 (examining current DNA evidence interpretation models); *see also infra* note 34.

[31] Lauren Kirchner, *Traces of Crime: How New York's DNA Evidence Techniques Became Tainted*, N.Y. TIMES, Sept. 4, 2017, https://www.nytimes.com/2017/09/04/nyregion/dna-analysis-evidence-new-york-disputed-techniques.html?_r=1.

DOI: https://doi.org/10.15779/Z384Q7QQ6X

to STRmix.

Although we don't know exactly how the numbers will change depending on variations in the hypotheses, we know they will change. We see this in the Amanda Knox case.[32] Amanda Knox was the American exchange student convicted of murdering her roommate in Italy in 2007, and then exonerated in 2015. Both her conviction and exoneration depended, in part, on contested DNA evidence. When the forensic DNA analyst increased the proposed number of contributors to one of the samples by just one individual, the likelihood ratio that was produced was decreased by ten million.[33] Even postulating the same hypotheses, different programs will generate different numbers. In *People v. Hillary*,[34] for example, the forensic evidence was tested by the two most prominent commercial programs being used by crime labs in the United States, STRmix[35] and TrueAllele. STRmix produced an inclusionary statistic, while TrueAllele[36] did not include Mr. Hillary. Ultimately, the judge did not admit the evidence at trial, but this ruling was based on a lack of internal validation studies, not based on the varied results. The National Institute of Standards and Technology announced in October 2017 that it would be conducting a study comparing the results obtained by different crime labs studying the exact same evidence.[37]

---

[32] Sam Tanenhaus, *Trial and Error*, N.Y. TIMES, May 24, 2013, https://www.nytimes.com/2013/05/26/books/review/trial-and-error.html.

[33] David Balding, *Evaluation of Mixed-Source, Low-Template DNA Profiles in Forensic Science*, 110 PROC. NAT'L ACAD. SCI. 12241, 12241-46 (2013).

[34] The case itself has been sealed, but news reports on the Judge's rulings and findings are available. *See* PCAST, *supra* note 1, at 79, n.212; *see also Newsroom: Cybergenetics Press Information on Forensic DNA Issues and More*, CYBERGENETICS (Feb. 21, 2019), https://www.cybgen.com/information/newsroom/2016/sep/New-York-judge-again-precludes-STRmix-from-Hillary-trial.shtml (collecting up-to-date news reports on the *Hillary* ruling and other forensic DNA issues).

[35] STRMix was first used in a criminal trial in Michigan in 2014. The defendant, Elamin Muhammad, was found guilty in large part based on the testing of a shoe found at the scene using this new approach.

[36] TrueAllele was first used in a criminal courtroom during the 2009 prosecution of Kevin Foley in Pennsylvania, but it was not used again until 2011. In 2011, TrueAllele was used in a Pennsylvania case, three federal cases, and a case in Northern Ireland. *See TrueAllele Admissibility*, CYBERGENETICS (January 21, 2019), https://www.cybgen.com/information/admissibility/page.shtml (collecting rulings of any trial that has made use of the TrueAllele analysis).

[37] NAT'L INST. STANDARDS & TECH., *NIST to Assess Reliability of Forensic Methods for Analyzing DNA Mixtures* (Oct. 3, 2017), https://www.nist.gov/news-events/news/2017/10/nist-assess-reliability-forensic-methods-analyzing-dna-mixtures; *see also* Lauren Kirchner, *Putting Crime Scene DNA Analysis on Trial*, PROPUBLICA,

This variation in results also highlights one of the fundamental problems with this type of evidence; there is no real way to test the results as there's no underlying true value. As Steele and Balding explained in the article, *Statistical Evaluation of Forensic DNA Profile Evidence*, "[l]aboratory procedures to measure a physical quantity such as a concentration can be validated by showing that the measured concentration consistently lies within an acceptable range of error relative to the true concentration. Such validation is infeasible for software aimed at computing an LR because it has no underlying true value (no equivalent to a true concentration exists). The LR expresses our uncertainty about an unknown event and depends on modeling assumptions that cannot be precisely verified in the context of noisy [crime scene profile] data."[38] Thus, labs can test whether the programs will include someone who they know is not in the mix (false positive studies), and studies can test consistency in results across programs, but there's no real way to determine if the LRs reflect the accurate likelihood of one hypothesis over the other as there is no underlying true value.

Although STRmix and TrueAllele are the most prominent, the FST was likely applied in the greatest number of actual cases during the six years it was relied on by the Office of Chief Medical Examiner in New York (OCME). Created by the crime lab itself, FST was not a proprietary of an independent commercial company like STRmix and TrueAllele, but rather the property of the OCME. Even though it was in use for quite a while, OCME discontinued FST and replaced it with STRmix.[39]

According to the lab's own statistics, the FST program was used in 1,350 cases over the six years it was active (2011 until 2017).[40] Most of these cases resulted in pleas, often due to the threat of this seemingly overwhelming DNA evidence. At the state level, defense attorneys moved

---

Oct. 11, 2017, https://www.propublica.org/article/putting-crime-scene-dna-analysis-on-trial.

[38] Christopher Steele & David Balding, *Statistical Evaluation of Forensic DNA Profile Evidence*, ANN. REV. STAT. & APP. 361, 380 (2014).

[39] I am aware of only two cases where FST and STRmix were used to evaluate the same evidence. For one of the mixtures, the FST protocols treated the sample as a mixture of at least two persons and the STRmix protocols treated it as at least three people. Using FST the likelihood ratio was in the millions, but when using STRmix to evaluate the same evidence, the likelihood ratio was reported as either in the billions or trillions. *See* United States v. Jones, 15-CR-153 (VSB), WL 2684101 (S.D.N.Y. Jun. 5, 2018) (including OCME Criminalist Craig O'Connor's testimony that the likelihood ratio reported "went up a couple orders of magnitude").

[40] Kirchner, *supra* note 31.

for admissibility hearings, but these were granted in only a few cases. FST has been the subject of only two New York State Supreme Court admissibility hearings.[41] In *People v. Rodriguez*, Judge Carruthers ruled the evidence admissible, then, in *People v. Collins*, Judge Dwyer ruled the evidence inadmissible.[42] Although these two hearings in New York reached different conclusions as to the admissibility of FST evidence, higher courts in NY have not yet considered the issue — either because the cases were decided on other grounds or the appeals remain pending.[43]

The LRs in many of these cases, both those that have been the subject of hearings and those for which hearings were never been granted, are often quite small, such that minor variation in the hypotheses, or the program's calculations, could easily turn an inculpatory LR into an exculpatory one. The evidence at issue in *United States v. Llamar Lawrence*, et. al., 16 Cr. 76 (JSR), for example, was an LR of 44,[44] while in *Collins*, one of the LRs at issue was 19. But these are just the examples we know from cases in which the use of the evidence was heavily litigated. In the thousands of criminal cases with DNA evidence reported as LRs, we have no idea how many were vulnerable to extreme variation, or so minimally significant.

Not only would these numbers be different if there was variation in the hypotheses or different programs employed to generate the numbers, even the significance of the LRs generated are greatly disputed and evolving. Scientist disagree about what LR values are informative or statistically significant, and how to report the probative value of LRs. As a result, different programs have different guidelines for reporting the significance of numbers.[45]

---

[41] *See People v. Debraux*, 21 N.Y.S.3d 535, 541-542 and *People v. Velez*, 52 N.Y.S.3d 248 (describing the *Frye* hearings in *People v. Rodriguez*, Ind. No. 5471/2009 (Sup. Ct. NY Co., 2013) (Carruthers, J.) (unpublished) (finding FST meets the *Frye* standard)); and *People v. Collins*, 15 N.Y.S.3d 564 (Sup. Ct. Kings Co. 2015) (Dwyer, J.) (finding FST does not meet the *Frye* standard).

[42] *See supra* note 41.

[43] Two intermediate appellate courts in New York State have ruled that it was not an abuse of discretion to deny a hearing on the admissibility of FST evidence, but this is far from a ruling on the merits of the evidence. People v. Gonzalez, 65 N.Y.S.3d 142 (1st Dep't, 2017); People v. Foster Bey, 67 N.Y.S.3d 846 (2nd Dep't, 2018). The highest court in New York State, the Court of Appeals, has not reached the issue.

[44] A *Daubert* hearing was granted as to the admissibility of this evidence in the Southern District of New York. Days before the hearing in front of Judge Jed Rakoff, the prosecutor made an offer to avoid the litigation, potential ruling, and its precedential effect.

[45] *See, e.g.*, Guro Dorum, Oyvind Bleka & Peter Gill, et al., "Exact Computation of the Distribution of Likelihood Ratios with Forensic Applications," Forensic Science

Since the OCME has moved from the FST program to STRmix, they no longer report an LR of anything less than 1,000 as statistically significant.[46] Thus, in addition to the dispute about the significance of these numbers among scientists in the field at large, even this one lab would now deem much of the evidence they previously presented as supporting a conclusion that the defendant was a contributor to the DNA sample as inconclusive. This underscores the fundamental problem with this evidence — as there is no ground truth there is no way to test the true probative value of the evidence. Unfortunately, there is no way to determine how many case outcomes, either pleas or convictions, turned on the stated significance of evidence under prior protocols. Traditional single source DNA evidence is not subject to the same extreme variation in interpretation.

B. The Interpretation of Complex and Degraded DNA Mixtures is Unlike Traditional "Gold Standard" DNA

The weight of a single-source, gold-standard DNA match is most frequently reported as a Random Match Probability (RMP). Generated as to a single DNA moniker (allele) or for a longer genetic profile (x loci), an RMP expresses the rarity of a certain profile, which makes it in essence a frequency statement of certain genes. It's simply based on the frequency of certain genetic markers. When testimony about a simple DNA profile match is presented at trial, two pieces of evidence are presented. First, the frequency of a profile obtained from an evidence sample is reported as an RMP.[47] Then, the suspect's profile is presented and compared to the evidence sample. If all of the same genetic markers are found in the suspect's profile, it is reported as a match. If even one genetic marker is missing or different, it is reported as an exclusion. There is no grey area.

You'll note, neither of these pieces of evidence, nor the conclusions drawn, require making any assumptions,[48] or postulating any

---

International: Genetics, 9 (2014) 93-101 ("The LR is reported as supporting the prosecution hypothesis if it is >1, and if it is <1 then it supports the defence hypothesis. Often likelihood ratios are only reported if the estimate is large and practices vary, but typically this critical 'number' is greater than *1 million.*") (emphasis added).

[46] *See* OCME TECHNICAL MANUAL, FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS 28, http://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/protocols-for-forensic-str-analysis/str-results-interpretation-powerplex-fusion-and-strmix.pdf.

[47] *See* NATIONAL RESEARCH COUNCIL, EVALUATION OF FORENSIC DNA EVIDENCE, *supra* note 21, at 127.

[48] There are assumptions made here about population samples being representative of

DOI: https://doi.org/10.15779/Z384Q7QQ6X

hypotheticals about the suspect's presence in the sample. In other words, the rarity of the profile, is independent of any particular accused. It is then compared to the suspect's profile. The analyst does not attempt to answer a question before the jury, but rather simply presents the finding. For example, the analyst may say, here are two profiles generated from these two samples, these profiles are this rare, and these two samples generated the same profile. This provides support for an issue before the jury, usually one of identity or occurrence of an event, but does so without making any assumptions or postulating any hypotheses about the evidence itself.

Before LRs, labs attempting to solve this question as to how to determine and present the evidentiary weight of samples that contain multiple individual's DNA, and are often degraded or missing information, the most commonly used method was the combined probability of inclusion (CPI). To put it as simply as possible, the CPI takes all of the observed data and considers all possible profiles that could produce that data. Then, it generates a statistic, which expresses the probability that a random person would have any of those generated profiles. Analysts generally believe this method is conservative, meaning that it underestimates the inculpatory value of the evidence. The CPI, however, is often used specifically to avoid what an analyst believes could be incorrect exculpatory results.

An analyst evaluating a degraded mixture generated from a small amount of DNA would see the suspect's genetic markers (alleles) at a number of locations (loci), but one or more of the suspect's markers could be missing from one or more locations. If the analyst hypothesizes that this could be due to degradation or incomplete information, as opposed to concluding that the defendant must be excluded as a contributor, the analyst would ignore that location, assume the sample is incomplete, and generate a CPI based on the other locations. Although analysts argue this is conservative, it is only so if the suspect is in fact a contributor to the DNA mixture. If there is not in fact missing information, the suspect's DNA marker (allele) is missing, such that the suspect should be excluded as a possible contributor. To use an illustrative example, although this is not the portion of the DNA chain being examined in these cases, if the genetic marker for blue eyes is missing, and the suspect has blue eyes, the suspect should be excluded as a contributor to the mixture.

---

an entire population, among other things. But these are assumptions about population genetics in general, not about the case itself.

After learning of this practice in forensic discussion groups, statistician James Curran and forensic scientist John Buckleton became concerned that the approach was producing misleadingly strong evidence against noncontributors. They performed a study to test just this hypothesis that the practice underestimates the exculpatory potential of evidence. The study found, using this CPI method, that 87 percent of the profiles tested that were not in fact in the mixture would generate an inculpatory CPI. In other words, by ignoring potentially exculpatory data, this method would include, as possible contributors to the mixture, 87 percent of the DNA profiles of known noncontributors, or innocent individuals. Thus, on closer analysis, this method is conservative only if the suspect is guilty.[49] The process assumes the suspect is in the mixture in order to generate the probability. If he is not in fact in the mixture, this probability is more than misleading. By proposing another explanation for the missing alleles, besides innocence, the analyst has found a way to make otherwise exculpatory evidence inculpatory.

Most scientist now agree that LRs are an improvement over what was often a troublingly subjective interpretation with CPI. As explained in the PCAST report, this use of CPI was "problematic because subjective choices made by examiners, such as about which alleles to include in the calculation, can dramatically alter the result and lead to inaccurate answers."[50] The report uses the example of a 2003 double-homicide case, *Winston v. Commonwealth*.[51] In that case, the DNA expert for the prosecution testified that the defendant could not be excluded from a mixed DNA sample found on a glove at the crime scene. "The prosecutor told the jury that the chance the match occurred by chance was 1 in 1.1 billion. A 2009 paper, however, makes a reasonable scientific case that

---

[49] *See* Erin Murphy, *The Dark Side of DNA*, 92-94, (citing James M. Curran & John Buckleton, *Inclusino Probabilities and Dropout*, J. FORENSIC SCI. 55 (2010) (describing the motivation for the study, and its results in their abstract: "[r]ecent discussions on a forensic discussion group highlighted the prevalence of a practice in the application of inclusion probabilities when dropout is possible that is of significant concern. In such cases, there appears to be an unpublished practice of calculation of an inclusion probability only for those loci at which the profile of interest (hereafter the suspect) is fully included among the alleles present in the crime scene sample and to omit those loci at which the suspect has alleles that are not fully represented among the alleles in the mixture. The danger is that this approach may produce apparently strong evidence against a surprisingly large fraction of noncontributors. In this paper, the risk associated with the approach of ignoring loci with discordant alleles is assessed by simulation.")

[50] *See* PCAST, *supra* note 1, at 75-78 (discussing the problems with the CPI method).

[51] Winston v. Commonwealth, 604 S.E.2d 21 (Va. 2004).

DOI: https://doi.org/10.15779/Z384Q7QQ6X

that the chance is closer to 1 in 2 — that is, 50 percent of the relevant population could not be excluded."[52] The defendant was sentenced to death.

PCAST cited another striking real-world example out of Texas, that only came to light by chance. In 2015, after the FBI detected an error in its population database, which is used to calculate DNA statistics, the Texas Department of Public Safety issued a letter to the criminal justice community. The errors were not expected to make a significant difference in DNA calculations, but the crime laboratory was willing to, upon request, recalculate the statistical significance of any particular potentially affected evidence. After a number of pieces of evidence were indeed retested, the results were shocking. "The statistics had changed dramatically — *e.g.*, from 1 in 1.4 billion to 1 in 36 in one case, from 1 in 4000 to inconclusive in another."[53] This triggered further investigation. It turned out the dramatic change in the reported statistical significance of the DNA samples was not due to the change in the FBI's population database, but, rather, the way in which analysts were employing the CPI method. It was the significance analyst gave to the potentially missing information at particular locations, and whether that was treated as exculpatory or potentially just degraded or missing information. Ultimately, the PCAST report concluded that the CPI method was so subjective that it is not a valid method for interpreting and reporting the statistical significance of complex mixtures of crime scene DNA.[54]

LRs, in part because they are calculated by computer programs, are thought to be a vast improvement over such subjective interpretation. However, they still seek to determine the possibility of a hypothesis that has no ground truth and, as such, the accuracy cannot be tested. And, as discussed above, there will always be a subjective choice made in what hypotheses to propose to generate the LR, and that decision is being made by crime labs working closely with the prosecution. The defense has no input. The analyst makes these subjective choices before the computer program can calculate the LR. These subjective choices, along with the program that a jurisdiction happens to employ, and the protocols of the local lab will determine how the significance, or probative value, of the DNA evidence is described to the jury.

---

[52] PCAST, *supra* note 1, at 76 (citing W.C. Thompson, *Painting the Target around the Matching Profile: The Texas Sharpshooter Fallacy in Forensic DNA Interpretation*, 8 LAW, PROBABILITY & RISK 257-76 (2009)).

[53] *Id.* at 76.

[54] PCAST, *supra* note 1, at 82.

II. THERE IS A MORE FUNDAMENTAL PROBLEM WITH THIS PROBABILISTIC EVIDENCE

As discussed above, there is much to be critical of in the different methods used to generate LR evidence. Some of these programs are proprietary such that defense attorneys are not given the opportunity to evaluate the programs or test their assumptions.[55] Some of the assumptions behind much of the software are questionable and not well tested.[56] Often the number of individuals in the mixture cannot be determined, which makes the calculations flawed from the start.[57] Some programs have not been developed with proper software development standards. For example, when the source code for the FST program was finally disclosed and examined by an independent software engineer, it turned out not to be running the calculations it purported and supposedly validated.[58] Additionally, these different programs come up with different LRs when evaluating the same evidence and there is no method for determining which, if any, of these numbers is more accurate.[59] Although

---

[55] *See, e.g.*, Lauren Kirchner, *ProPublica Seeks Source Code for New York City's Disputed DNA Software*, https://www.propublica.org/article/propublica-seeks-source-code-for-new-york-city-disputed-dna-software (discussing efforts to obtain access to the source code of proprietary software, with a focus on FST program used by the New York Office of the Chief Medical Examiner, but also discussing efforts in California courts to obtain access to the source code for TrueAllele); *see also*, ERIN E. MURPHY, INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA 97-98 (2015).

[56] Dr. Budowle put it succinctly in the *Collins* hearing on the assumptions behind the FST software: "I think it is in the way it's being used because it makes certain assumptions about DNA typing that no one else would do even in standard DNA typing. The main assumption being made is that all the rates for drop-in, drop-out are based on the amount of DNA." Transcript of Record at 793, *Collins*, 15 N.Y.S.3d 564 (emphasis added).

[57] *See* Perez et al., *supra* note 15.

[58] *See* Kirchner, *supra* note 31.

[59] "Some programs use discrete (semi-continuous) methods, which use only allele information in conjunction with probabilities of allelic dropout and drop-in, while other programs use continuous methods, which also incorporate information about peak height and other information. Within these two classes, the programs differ with respect to how they use the information. Some of the methods involve making assumptions about the number of individuals contributing to the DNA profile, and use this information to clean up noise (such as "stutter" in DNA profiles)." *See, e.g.*, PCAST, *supra* note 1 at 211; Mark W. Perlin, Jennifer M. Hornyak, Garett Sugimoto, & Kevin W.P. Miller, *TrueAllele genotype identification on DNA mixtures containing up to five unknown contributors*, 60

defense attorneys are raising challenges, most courts are admitting the evidence, which is being used to convict individuals of serious crimes.

Challenges to this new category of DNA evidence are usually made under the *Frye* or *Daubert* standards, arguing that the science is not generally accepted in the scientific community, or insufficiently reliable to be admissible.[60] These motions often address the variation in results, lack of agreement among scientists and statisticians as to how to calculate and report these LRs and how to translate the statistical significance of these numbers to a jury, as well as challenging the underlying assumptions employed by the programs. These challenges further question the lack of access to the source code such that the defense is in the dark about what calculations are even being performed to generate the evidence against them. In short, the technology and science is so disputed, that there is insufficient consensus in the scientific community regarding the admissibility of these LRs.

But there is an even larger and quite distinct problem with this evidence. The opacity of the LRs obscures something important that is happening when they are introduced in a criminal trial. The complicated math and science distracts judges, lawyers, and surely jurors from the essential nature of this evidence — that it expresses the relative probability of two hypotheticals. Bayesian reasoning has to be employed to convert the LRs into a meaningful probability, and, as I discuss *supra*, this undermines the presumption of innocence and the prosecutor's burden of proving their case beyond a reasonable doubt. This concern is exacerbated by the sheer power of the DNA moniker, which dwarfs any and all other less purportedly "scientific" evidence.

Although this specific type of evidence is new, it is not the first time a Bayesian approach to the trial process has been proposed. In 1971

JOURNAL OF FORENSIC SCIENCES 857-868 (2015); Susan A. Greenspoon, Lisa Schiermeier-Wood, & Brad C. Jenkins, *Establishing the limits of TrueAllele® Casework: A validation study,* 60 JOURNAL OF FORENSIC SCIENCES 1263–76 (2015); Jo-Anne Bright, Duncan Taylor, Catherine McGovern, Stuart Cooper, Laura Russell, Damien Abarno, & John Buckleton, *Developmental validation of STRmixTM, expert software for the interpretation of forensic DNA profiles*, 23 FORENSIC SCIENCE INTERNATIONAL: GENETICS 226-39 (2016).

[60] The *Frye* test holds that expert scientific evidence is admissible only if it is generally accepted as reliable by the relevant scientific community at the time of the proceedings. Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). The *Daubert* test evaluates the validity and reliability of the underlying scientific methodology and whether the reasoning can be reliably applied to the facts at issue. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 589 (1993).

Laurence Tribe, in *Trial By Mathematics: Precision and Ritual in the Legal Process*, analyzed the use of mathematical methods in the trial process, concluding that this sort of evidence undermines the fundamental normative requirement that each and every case be proven beyond a reasonable doubt. Much of his analysis is equally prescient today. Revisiting Tribe's arguments, and viewing LRs through the lens of the debate that ensued, I hope to clarify why likelihood ratios of the sort employed by these probabilistic genotyping programs have no place in criminal trials.

A. Probabilistic Evidence as to an Ultimate Issue

There are many examples of prosecutors trying to introduce evidence of the probability of a defendant's guilt into a trial, and appellate courts generally reject such evidence.[61] One of the most frequently cited is the 1968 California case, *People v. Collins*.[62] In *Collins*, the prosecution faced identification problems with the two defendants. A woman had been pushed from behind by someone whom she neither saw nor heard approach. When she looked up, she saw a woman running from the scene. Nearby, another witness heard a commotion and saw a woman run from the alley and enter a yellow car driven by a black man with a mustache and beard. That eyewitness described a white woman, slightly over 5 feet tall, with a dark blonde ponytail and dark clothing. The prosecutor charged a married couple: a black man who drove a yellow Lincoln but

---

[61] *See, e.g.*, United States v. Massey, 594 F.2d 676 (8th Cir. 1979) ("Our concern over this evidence is . . . with its potentially exaggerated impact upon the trier of fact. Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established 'beyond a reasonable doubt.' Diligent cross-examination may in some cases minimize statistical manipulation and confine the scope of probability testimony. We are not convinced, however, that such rebuttal would dispel the psychological impact of the suggestion of mathematical precision. . ."); People v. Collins, 68 Cal.2d 319 (Cal. 1968); Chumbler v. Commonwealth, 905 S.W.2d 488 (Ky. 1995) ("The statistical calculations rival a polygraph in their unreliability and propensity to mislead and may have convinced jurors of modest analytical ability that no one but [the defendant] could have committed the crime."); State v. Carlson, 276 N.W.2d 170, 176 (Minn. 1978) ("Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established 'beyond a reasonable doubt.'") (citing Lawrence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971) and *Collins*, 68 Cal.2d 319); State v. Sneed, 76 N.M. 349 (1966).

[62] *See generally Collins*, 68 Cal. 2d 319.

DOI: https://doi.org/10.15779/Z384Q7QQ6X

did not have a beard and a white woman with blonde hair she wore in a ponytail. At trial, the victim couldn't identify the wife as the perpetrator, and had never seen the husband.

In an effort to bolster the identification of the defendants, the prosecutor called an instructor of mathematics to establish that the defendants were the perpetrators. The testimony assumed the robbery was committed by a Caucasian woman with a blonde ponytail who left the scene with a black man with a beard and mustache, and tried to articulate a probability that another couple possessed this combination of characteristics in the Los Angeles area. The prosecutor argued that it was so unlikely that another couple possessed these features that the defendants had to be the guilty parties. Positing probabilities of each of the characteristics, the mathematician applied the product rule, which states that the probability of joint occurrence of a number of mutually independent events is equal to the product of the individual probabilities that each of the events will occur. In other words, for illustration, suppose there is one in ten chance that a lawyer would have red hair, and a one in ten chance that a lawyer would be Catholic, then, applying the product rule, there is a 1 in 100 chance that a random lawyer would possess both these characteristics. The expert then testified that there was a 1 in 12 million chance that any other couple possessed the distinctive characteristics as the defendants. "Accordingly, under this theory, it was to be inferred that there could be but one chance in 12 million that defendants were innocent and that another equally distinctive couple actually committed the robbery."[63]

Here, the prosecutor made a grave and yet logically appealing mistake. I think Tribe expressed it succinctly:

> [The] prosecutor erroneously equated the probability that a randomly chosen couple would possess the incriminating characteristics, with the probability that any given couple possessing those characteristics would be innocent. After all, if the suspect population contained, for example, twenty-four million couples, and if there were a probability of one in twelve million that a couple chosen at random from the suspect population would possess the six characteristics in question, then one could well expect to find two such couples in the suspect population, and there would be a probability of approximately one in two — not one in twelve million — that any given couple possessing the six characteristics would be innocent.[64]

---

[63] *Id.* at 325.

[64] Tribe, *supra* note 61, at 1336.

Thus, if the probability were correct, and the guilty parties possessed the characteristics, there was only a 50% chance they had the right couple.

The reasoning of this testimony was flawed on many other levels. At the outset, there was no evidence that the probabilities of each of the characteristics proffered by the prosecutor were in any way grounded in fact, nor was there any evidence that the characteristics were actually independent, which would be necessary for the application of the product rule. This is not surprising; poorly understood probability statistics invite flawed reasoning. More generally, however, in reversing the conviction, the Supreme Court of California explained that "[q]uite apart from our foregoing objections to the specific technique employed by the prosecution to estimate the probability in question, we think that the entire enterprise upon which the prosecution embarked, and which was directed to the objective of measuring the likelihood of a random couple possessing the characteristics allegedly distinguishing the robbers, was gravely misguided."[65]

Although the court did not go so far as to hold that there is no place for mathematical techniques in the proof of facts, it emphasized particular concern over such evidence in criminal cases. The *Collins* decision inspired academics and commentators to ask the bigger question: does mathematics in fact pose problems of a more pervasive and fundamental character? I will focus primarily on Professor Tribe's analysis. It is my position that he not only got it right, but that history has affirmed this fact.[66]

Tribe starts with the uncontroversial position, and one articulated in the *Collins* case: "no mathematical equation can prove beyond a reasonable doubt (1) that the guilty [party] *in fact* possessed the characteristics described by the People's witnesses, or even (2) that only *one* [party] possessing those characteristics could be found in the [relevant] area."[67] Since the probability evidence could not be sufficient on its own, that evidence must be viewed in light of additional evidence,

---

[65] *Collins*, 68 Cal.2d at 329.

[66] In order to explore this question, Professor Tribe focused on three types of illustrative cases. The categories he identified were: (1) cases where the jury had to determine whether or not an event occurred, (2) cases where the jury was tasked with determining the identity of a party, and (3) cases where the jury sought to determine a party's intention. Mixed samples of DNA are designed to answer the second question, concerning identity. I will, thus, focus on just those cases. Although Professor Tribe studied both civil and criminal cases, the heightened burdens in criminal cases make his work on criminal cases most relevant to the inquiry at hand.

[67] Tribe, *supra* note 61, at 67.

not so easily quantifiable. Herein lies the challenge.

Tribe addressed a proposal for how that might be done. Michael Finkelstein and William B. Fairley, in a *Bayesian Approach to Identification Evidence*,[68] analyzed the *Collins* decision by proposing a method for integrating probability evidence into the criminal trial process by applying Bayesian probability analysis. Bayes' theorem describes the process by which information relevant to the probability of an event is combined with the prior probability of that event to produce a posterior probability. In applying that prior probability, that new information is made meaningful in the form of a posterior probability. What this amounts to, in our context, is that in order to integrate other evidence with the mathematical probability, that other evidence must first be converted into a probability.[69] This is where the problem with probability-based evidence becomes so stark in the context of the presumptions in a criminal case.

Tribe discusses the example analyzed by Finkelstein and Fairley of a palm print found on a murder weapon. Because the print is only a partial print, the expert can only say that this partial print could be produced by no more than one suspect in a thousand. So, how is a jury to integrate this evidence?

> By itself, of course, the "one-in-a-thousand" statistic is not a very meaningful one. It does not, as the California Supreme Court in *Collins* showed, measure the probability of the defendant's innocence—although many jurors would be hard-pressed to understand why not. As Finkelstein and Fairley recognize, even if there were as few as one hundred thousand potential suspects, one would expect approximately one hundred persons to have such prints; if there were a million potential suspects, one would expect to find a thousand or so similar prints. Thus the palm print would hardly pinpoint the defendant in any unique way.[70]

In order to understand the significance of this evidence, the jury must consider the prior probability of guilt. In the context of our example above, the four young men in the car and the gun in the trunk, in order to integrate the new probability evidence, the LR, with other evidence, the

---

[68] 83 HARV. L. REV. 489, 498-517 (1970).

[69] *See generally* Michael Finkelstein and William B. Fairley, *A Comment on "Trial By Mathematics"*, 84 HARV. L. REV. 1801 (1971); Finkelstein and Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); C.R. Kingston & P.L. Kirk, *The Use of Statistics in Criminalistics*, 55 J. CRIM. L. & CRIMINOLOGY 516 (1964); Lawrence Tribe, *A Further Critique of Mathematical Proof*, 84 HARV. L. REV. 1810 (1971); Lawrence Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

[70] Tribe, *supra* note 61, at 1355.

jurors must transform the prior evidence into a probability. They must decide, before hearing about the LR, what they think the probability of guilt is.

One possibility is to assume that, prior to the generated LR, that each hypothesis in question should be treated as equally probable, such that the prior probability is 1— it is equally likely that the prosecutor's hypothesis is true as the posited defense hypothesis. But jurors are not supposed to start from a place where they believe there is an equal chance of guilt as innocence. They are to presume innocence. Another possibility is to start with zero. But the starting place cannot be zero because starting at zero would lead only to verdicts of "not guilty." A zero probability of guilt, combined with any other number, is always going to be zero.

The court in *Skipper* explored this exact issue in the context of the use of paternity tests to prove the element of sexual intercourse in a sexual assault case.[71] In *Skipper*, the issue on appeal was the admissibility of the probability of paternity statistic, calculated from DNA evidence, to prove the element of sexual intercourse. The defendant was accused of sexual intercourse with a high school girl who became pregnant. An expert testified based on the paternity index as to the likelihood that the defendant was the father of the child. The index produces, in essence, an LR that the defendant would produce a child with the phenotypes of the fetus as compared to an unrelated random male.[72] In *Skipper*, the paternity index was 3496, indicating that, given the evidence sample, it was 3496 more probable that the defendant was the father than the probability that a randomly selected unrelated man was the father.

The expert further testified that the index could be converted into a statistic indicating the defendant's probability of paternity by applying Bayes' theorem. He then testified that the probability that the defendant fathered the child was 99.7 percent. In order to produce that percentage, the expert had to posit a prior probability of intercourse. The expert set the prior probability at "50 percent, expressed as odds of one, i.e., fifty-fifty, reasoning that 50 percent is a neutral starting point because it assumes that it is just as likely that the defendant is not the father as it is that he is the father."[73] The court explained that this " 'utilization of Bayes' Theorem by the prosecution [ ] permitted the introduction of evidence predicated on the assumption that there was a fifty-fifty chance

---

[71] *See generally* State v. Skipper, 228 Conn. 610 (1994).
[72] *Id.* at 615.
[73] *Id.* at 617.

that sexual intercourse had occurred in order to prove that sexual intercourse had in fact occurred."[74]

In reversing the conviction, the court concluded that the admission of this evidence undermined the presumption of innocence:

> In fact, if the presumption of innocence were factored into Bayes' Theorem, the probability of paternity statistic would be useless. If we assume that the presumption of innocence standard would require the prior probability of guilt to be zero, the probability of paternity in criminal case would always be zero because Bayes' Theorem requires the paternity index to be multiplied by a positive prior probability in order to have any utility. In other words, Bayes' Theorem can only work if the presumption of innocence disappears from considerations.[75]

The court posited that the presumption of innocence should be represented as zero, which renders the application of Bayes theorem useless — all posterior probabilities would be zero. Setting the odds to one doesn't work because it presumes an equal chance of guilt and innocence, and zero renders the numbers useless. But these, of course, are not the only options.

As discussed *supra,* LRs express the relative probabilities of two hypotheses to explain the evidence, and do not express a true probability. It is only when combined with a prior probability that they become meaningful and express an actual probability about the evidence in the case. Our exploded pillow example above might help clarify this point again. If you assume there is a 50% chance that the pillow exploded because the house was burglarized, and a 50% chance the pillow was designed to explode, the LR becomes meaningful. You have assumed that each of these explanations are equally probable, and account for 100% of the possible explanations — that they are the only two possible explanations. But, in fact, that is an incorrect assumption. There are other possible explanations for the evidence, and, in a criminal trial, 50% is an incorrect starting point. Although it may be a fair starting point if the burden were a preponderance of the evidence, and you had identified the only two possible explanations for a given piece of evidence. That is not the burden of proof in a criminal trial, and, as discussed *infra*, there are more than two potential hypotheses that could explain the evidence. But some prior probability is necessary to give meaning to the LR.

Just as a paternity index of 3496 is not meaningful to a jury

---

[74] *Id.* at 619.

[75] *Id.* at 623 (internal citations omitted).

without an explanation and conversion to a percentage using Bayesian reasoning, an LR is not properly understood without applying Bayes. [B]ecause Bayes' Theorem will be introduced in the State's case . . . *the jury inevitably will be impelled to focus, during the State's case, before all of the evidence is in, on the probability of defendant's guilt.*[76]

This is inconsistent with the jury's sworn obligation to presume the defendant innocence until they have heard all the evidence and found that the state has met their burden of proving guilt beyond a reasonable doubt. Thus, the incorporation of Bayes' Theorem, is in direct conflict with the obligation that a jury presume innocence until the close of the case. This is not true of RMPs, which can be understood without reference to any specific hypothesis or prior assumption about the case before all the evidence is presented.

B. Likelihood Ratios Persuade a Jury that They Should Convict Without Proof Beyond a Reasonable Doubt

Whether Bayes' theorem is required to integrate mathematical evidence with non-mathematical evidence or not, the introduction of probability evidence in criminal trials invites the jury to render a guilty verdict with proof less than beyond a reasonable doubt. The *Collins* court explained that "[i]n essence this argument of the prosecutor was calculated to persuade the jury to convict defendants whether or not they were convinced of their guilt to a moral certainty and beyond a reasonable doubt."[77]

Any time a jury is urged to convict based on an argument that there is a likelihood of guilt to some percentage short of 100%, they are accepting the conviction of a certain percentage of innocent men and women. Of course, we all know that the innocent are convicted. DNA exonerations have had a profound impact on our understanding of the frequency and cause of such wrongful convictions, raising grave concerns about the legitimacy of many other convictions that lack the benefit of DNA evidence. We also know that beyond a reasonable doubt does not mean beyond all doubt. However, in quantifying reasonable doubt, we explicitly accept the conviction of a percentage of innocent men and women.

Shortly after Tribe's famous article, Charles Nesson explored this

---

[76] State v. Spann, 130 N.J. 484, 517 (1993) (citing Tribe, *Trial by Mathematics*, *supra* note 61, at 1368-71).

[77] *Id.* at 332.

issue in the context of the propriety of rebuttable presumptions in criminal cases.[78] Professor Nesson argued that any effort to quantify reasonable doubt would undermine the concept and its role in the legal system:

> Reasonable doubt defies exact definition precisely because it is a concept meant to encompass many different, individual views of how probable guilt must be (or how unlikely innocence must be) to warrant conviction. The closer reasonable doubt comes to explicit quantification, the more any notion of it being a shared concept will break down.[79]

There is something in the very process of trying to quantify the proof beyond a reasonable doubt that undermines its normative power, and the normative symbolism of a jury trial plays a powerful role. To put it in stark terms, there is a huge difference between an innocent man being lynched before being brought to court, and an innocent man being convicted after a fair trial. The outcome may be the same for that individual, but the outcome is not the same for the rule of law and all that it represents. Similarly, "there is a qualitative difference between the outcome of erroneously convicting a man when the trier has been fully convinced of his guilt and the outcome of erroneously convicting a man when the trier has reason to believe he may be innocent."[80]

We have chosen a system in which we demand proof beyond a reasonable doubt. It is not a balancing act. Instead, the risk of erroneous acquittal is not weighed against the risk of erroneous conviction. The standards set forth in the Constitution rightly reject a system that cavalierly accepts any error rate. In each and every case, we want jurors to demand proof beyond a reasonable doubt.

> It is here that the great virtue of mathematical rigor – its demand for precision, completeness, and candor – may become its greatest vice, for it may force jurors to articulate propositions whose truth virtually all might already suspect, but whose explicit and repeated expression may interfere with what seem to me the complex symbolic functions of trial procedure and its associated rhetoric.[81]

---

[78] Charles Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 Harv. L. Rev. 1187 (1979).

[79] *Id.* at 1197.

[80] *See generally* Laurence Tribe, *An Ounce of Detention: Preventative Justice in the World of John Mitchell*, 56 VA. L. REV. 371 (1970).

[81] Tribe, *supra* note 61, at 1371; *see also* R. Jonakait, *When Blood Is Their Argument: Probabilities in Criminal Cases, Genetic Markers, and, Once Again, Bayes' Theorem*, 1983 U. ILL. L. REV. 369, 415-20 (1983) ("Confronted with an equation which purports to yield a numerical index of probable guilt, few juries could resist the temptation to

We don't want a system in which we articulate that it is acceptable if a certain percentage of those convicted are innocent, no matter how small that percentage might be. Of course, we all know that innocent men and women are convicted of crimes. But that cannot be an explicit aim of our criminal justice system. An inescapable consequence of quantification of the burden of proof is that the number will never be absolute, never 100%. Similarly, in any individual case, the probability of guilt will never be 100%.

So, what if we consider 95% sufficient. That would mean a margin of .05% or a chance of one out of 20 that the accused is innocent. "[T]here is something intrinsically immoral about condemning a man as a criminal while telling oneself, 'I believe that there is a one in 20 chance that this defendant is innocent, but a 1/20 risk of sacrificing him erroneously is one I am willing to run. . .'"[82] At the heart of our criminal justice system we have established principles that reject this sort of risk.

C. Likelihood Ratios are Inconsistent with these Burdens

The likelihood ratio pits a theory of the prosecution against a theory of the defense as proposed by the crime lab. "Before computing the LR, one must specify prosecution and defense hypotheses. . ."[83] LRs are created by balancing two hypotheses against each other: that the mixture contains the defendant's DNA and the DNA of x number of unknown unrelated individuals, versus the hypothesis that the mixture contains only the DNA of x number of unknown unrelated individuals. Thus, the technician proposes that it is the defendant's DNA, and weighs that possibility against the possibility that it is not the defendant's DNA in the mix.

As discussed above, in our hypothetical about the four men in the car, the initial problem is that the crime lab proposes the defense hypothesis. Not only does the defendant not need to propose such a hypothesis, but also, there's more than one potential defense hypothesis.

---

accord disproportionate weight to that index."); C. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1225 (1979) ("Any conceptualization of reasonable doubt in probabilistic form is inconsistent with the functional role the concept is designed to play."); *State v. DelVecchio*, 191 Conn. 412, 417-18 (1983) ("Jury instruction using a football field simile and instructing the jury that 'it . . . is up to you to decide' where reasonable doubt lies between the fifty yard line and one hundred yard line diluted the constitutional standard of proof beyond a reasonable doubt.").

[82] Tribe, *supra* note 61, at 1372.

[83] Perez et al., *supra* note 15, at 315.

The analyst, in our hypothetical, proposed that the contributors were unrelated. But what reason does the analyst have to presume this? Surely sometimes people in the same family touch a single object. The analyst also hypothesized that there were three contributors to the DNA mix, but that is not the only possible explanation of the evidence. As discussed *supra*, there is no way to tell for certain, from the evidence itself, how many people contributed to a given mixture of DNA.

More fundamentally, two positions are not supposed to be weighed against each other in a criminal trial. As a judge repeatedly explains to a jury, the defense does not have to put on a case, or present any particular theory of innocence. The defense can sit silently, present no evidence, and the prosecution still must prove their case beyond a reasonable doubt. Jurors are expressly directed not to weigh two positions against each other, but rather to evaluate the prosecution's hypothesis and determine if that hypothesis is proven beyond a reasonable doubt.

To be sure, this is not an easy task. People are used to listening to two sides, and evaluating them against each other — weighing two competing views, and determining which is more persuasive. And this is what happens in a civil trial. But our justice system explicitly rejects this kind of reasoning in the criminal context. Our forefathers, in all their wisdom, decided that when the weight of the government is coming down against an individual, and threatening to take away her life or liberty, we want more proof than that. The government is thus burdened with proving their accusations beyond a reasonable doubt, and the defendant bears no burden of proving her innocence. And much of a criminal trial is aimed at safeguarding the defendant against the natural inclination of juries to weigh the two sides against each other. It is an ideal, for sure, and not always accomplished, but the principle is fundamental and one that our system uncontrovertibly strives to achieve.[84] As expressed in *State v. Hartman,*[85] "it is antithetical to our system of criminal justice to allow the state, through the use of statistical evidence which assumes that the defendant committed the crime, to prove that the defendant committed the crime."

D. Traditional DNA Evidence is not Subject to the Same Criticisms

Traditional DNA evidence, presented as an RMP, is not subject to

---

[84] *See, e.g., Winship*, 397 U.S. at 364.
[85] 145 Wis. 2d 1, 16 (1988).

the same criticisms. RMPs are undeniably probabilistic. In expressing the rarity of a profile, RMPs evaluate the probability that a random person would possess the genetic profile in question. RMPs are often distinguishable by the sheer rarity of genetic profiles expressed as RMPs — the numbers are often so staggering that it is unlikely that another unrelated person has ever possessed such a profile. When the rarity is so staggering, the evidence is certainly easier for a jury to evaluate. They can infer that the defendant was the source of the DNA, and then analyze that evidence with the other less scientific evidence. In essence, this is the opposite of Bayesian— they convert the evidence from a number, to something more qualitative. There is no reasonable doubt that it is his DNA at the scene, for example.

More significantly, however, LRs differ from RMPs in what they express— it's not merely that probabilities are involved. LRs express the relative probability of two hypotheticals significant to the case. More specifically, unlike RMPs, LRs do not actually express a probability about the world, they express the relationship between two specific explanations for a piece of evidence. These two explanations are used as stand-ins for the two positions in the adversarial trial process — the prosecutor's hypothesis versus the defendant's hypothesis. This is exactly the sort of reasoning the prosecutor applied in *Collins*. Inserting a supposed defense hypothesis that there was some other couple with the characteristics in question in the area, the prosecutor weighed that probability against the probability the defendants were the only such couple. These LRs are not meaningful independent of the trial. This is not true of the RMPs. RMPs express a non-evaluative fact, namely the observed rarity of certain genetic markers. They are, in essence, frequency statements.[86]

Additionally, with RMPs, the testimony at trial, that a profile obtained from the defendant matches the suspect profile, is not actually presented in probabilistic terms. The analyst testifies that the same alleles were seen in the suspect profile as the profile obtained in the evidence sample. In order to get from there to the proposition that the suspect is the

[86] Technically RMPs can be easily converted into LRs, and some would say LRs are just the inverse of an RMP. The fact that an RMP can be expressed is an LR doesn't undermine the fundamental point. RMPs describe the frequency of a genetic profile, and the next step is to convert that frequency into the likelihood that a random person would possess those characteristics. But you don't need to know anything about the suspect profile to determine the frequency of the evidence profile. In that sense, what is happening in the context of LRs in probabilistic genotyping is distinguishable from the more traditional RMP.

source of the DNA, an inference must be made. But the probability expressed as an RMP is not the probability that the suspect is the source of the DNA. The Supreme Court of Minnesota expressed this well in *Sate of Minnestoa v. Bloom*.

> There is nothing inherently wrong in a jury using its inference that the match is a true match as the basis for another inference, specifically, that the defendant is the source. What is important is that the jury know that it has to go through the process of making the inference. The probability that a randomly selected person would have the same profile as the sample found at the scene is not the probability that someone other than the defendant is the source. But it is commonly assumed that it is the probability that someone other than defendant is the source. This is what is often referred to as "source probability error." In order to give an opinion as to the probability that someone other than defendant is the source, one would first have to estimate the size of the potential source population.[87]

In *Bloom*, the court established a DNA (specifically RMP) exception to the rule against probabilistic evidence in criminal trials. In so ruling, the court outlined a number of improper inferences. In addition to what was described as the source probability error, the court warned against this evidence being mistakenly equated with the probability that the defendant is the perpetrator of the crime. The Court called this the ultimate issue error, but it is also often referred to as the prosecutor's fallacy. This is the same sort of flawed reasoning exercised by the prosecutor in *Collins*.

Rather than shielding against this flawed reasoning, LRs actually invite this reasoning. LRs are designed to specifically present the relative probabilities that the evidence came from the defendant versus the probability that it came from an unknown unrelated individual. It is this leap that distinguishes RMPs from LRs. LRs try to answer a question before the jury in probabilistic terms, thereby inviting a jury to convict because the defendant is probably guilty.

E. These Presumptions and Burdens in a Criminal Trial Should be Safeguarded

The presumption of innocence is an "axiomatic and elementary" principle, "enforcement [of which] lies at the foundation of the administration of our criminal law."[88] To safeguard that principle, "courts

---

[87] 516 N.W.2d 159, 162-63 (Minn. 1994) (internal citations omitted).

[88] Coffin v. United States, 156 U.S. 432, 453 (1895).

must be alert to factors that may undermine the fairness of the fact-finding process. In the administration of criminal justice, courts must carefully guard against dilution of the principle that guilt is to be established by probative evidence and beyond a reasonable doubt."[89]

At times, in the trial process, jurors are asked to aspire to positions that, realistically, they are unlikely to achieve. It is unlikely that any juror truly believes, if asked to reflect at the beginning of a trial, before hearing any evidence, that there is no reason the defendant has been brought to trial. They are directed not to speculate, but to presume innocence. As Tribe explains, the presumption is more than a rule of evidence, it is an aspiration. The presumption of innocence represents a commitment to the respect, the respect we as a society choose to give a person accused of a crime. Even if, factually speaking, no juror truly believes that defendants stand accused of a crime for no reason at all.

> The presumption retains force not as a factual judgment, but as a normative one —as a judgment that society ought to speak of accused men as innocent, and treat them as innocent, until they have properly convicted after all they have to offer in their defense has been carefully weighed. The suspicion that many are in fact guilty need not undermine either this normative conclusion or its symbolic expression through trial procedure, so long as jurors are not compelled to articulate their suspicion of guilt in any explicit and precise way.[90]

The presentation of DNA evidence in the form of a likelihood ratio can only be understood by articulating a prior probability of guilt. You are not treating a man innocent if you articulate a probability of guilt.

CONCLUSION

Although DNA evidence has had an overwhelmingly positive influence on the criminal justice system— helping to exonerate the innocent, convict the guilty, and shed light on flaws in the system that lead to wrongful convictions— this new generation of DNA evidence shares little with its predecessor. As the tools to examine DNA evidence become more and more sensitive, and smaller and smaller amounts of DNA from a crime scene can be analyzed, that evidence is becoming less and less probative. As this evidence becomes less probative, with often only a few skin cells being examined, forensic science experts search for ways to give that less meaningful evidence some inculpatory weight. This

---

[89] *Winship*, 397 U.S. at 364.

[90] Tribe, *supra* note 61, at 1371.

trend not only waters down the DNA moniker, it is watering down the presumption of innocence. As the science evolves, and the statistical analysis becomes more complicated, judges, lawyers, and jurors are less able to understand the value and significance of this evidence. These statistics, posing as the likelihood of guilt or innocence, create a trial by mathematics — fundamentally inconsistent with the constitutional norms of the criminal justice system.