PREPARED BY THE COURT

STATE OF NEW JERSEY Plaintiff, v. PAUL J CANEIRO Defendant.	SUPERIOR COURT OF NEW JERSEY LAW DIVISION: CRIMINAL PART MONMOUTH COUNTY INDICTMENT No. 19-02-0283-I CASE No. 18-4915 **ORDER**

FILED
MARCH 6, 2025
Marc C. Lemieux, A.J.S.C.

THIS MATTER having been brought before the Court on application of defendant Paul J Caneiro (Christopher Godin, Esq., Tamar Lerer, Esq., J. Michael Wicke, Esq., and Victoria Howard, Esq., appearing), for an order to exclude evidence, and Raymond S Santiago, Monmouth County Prosecutor, (Christopher J Decker, Assistant Prosecutor, and Nicole Wallace, Assistant Prosecutor, appearing) for the State of New Jersey; and the Court having held a hearing; and having reviewed and duly considered the arguments and papers submitted; and for good cause shown;

IT IS on this 6th day of March, 2025;

ORDERED that the defendant's motion to exclude DNA evidence analyzed utilizing STRmix v. 2.5.11 and v. 2.8.0's is **DENIED**; and it is further

ORDERED that a copy of this order shall be served upon all counsel of record via e-courts.

Hon. Marc C. Lemieux, A.J.S.C.

NOT FOR PUBLICATION WITHOUT THE

APPROVAL OF THE COMMITTEE ON OPINIONS

	SUPERIOR COURT OF NEW JERSEY
STATE OF NEW JERSEY	MONMOUTH COUNTY
	LAW DIVISION-CRIMINAL PART
v.	INDICTMENT No. 19-02-0283
	CASE No. 18-4915
PAUL CANEIRO	CRIMINAL ACTION
Defendant.	

DECIDED: March 6, 2025

FINDINGS AND CONCLUSIONS OF THE COURT

CHRISTOPHER DECKER and NICOLE WALLACE, Assistant Prosecutors, for the State of New Jersey (Raymond Santiago, Monmouth County Prosecutor, Attorney)
CHRISTOPHER GODIN, TAMAR LERER, J. MICHAEL WICKE, and VICTORIA HOWARD Deputy Public Defenders, for defendant (Jeniffer Sellitti, Public Defender, Attorney)

MARC C. LEMIEUX, A.J.S.C.

I. CONTENTS

TABLE OF AUTHORITIES

II. PRELIMINARY STATEMENT

This matter comes before the court by way of defendant's motion for a N.J.R.E. 702/Olenowski[1] hearing on the reliability of STRmix v. 2.5.11, later expanded on remand from the New Jersey Supreme Court to include STRmix v. 2.8.0. State v. Caneiro, 256 N.J. 531 (2024).

The State of New Jersey has charged defendant, Paul Caneiro, with four counts of first-degree murder, two counts of first-degree felony murder, two counts of second-degree aggravated arson, one count of second-degree possession of a weapon for an unlawful purpose, one count of second-degree unlawful possession of a weapon , one count of third-degree possession of a weapon for an unlawful purpose, one count of fourth-degree unlawful possession of a weapon, one count of second-degree theft of movable property, one court of fourth-degree misapplication of entrusted property, and two counts of third-degree hindering the apprehension of oneself. The State intends to introduce at trial DNA analysis of evidence performed by a probabilistic genotyping software called STRmix. "Such programs employ probabilistic genotyping, the 'use of biological modeling, statistical theory, computer algorithms, and probability distributions,' to 'assist,' rather than 're-place,' 'the DNA analyst in the interpretation of forensic DNA typing results.'" State v. Pickett, 466 N.J. Super. 270, 283 (App. Div. 2021) (quoting Science Working Group on DNA Analysis Methods (SWGDAM), Guidelines for the Validation of Probabilistic Genotyping Systems 2 (June 2015) (hereinafter "SWGDAM Guidelines)). "Specifically, the programs use mathematical models and simulations, subject to parameters programmed into the software to account for drop-in or

[1] State v. Olenowski, 253 N.J. 133 (2023) (hereinafter "Olenowski I"). The standard outlined there was later elaborated in State v. Olenowski, 255 N.J. 529 (2023) (hereinafter "Olenowski II").

drop-out effects and other issues, to calculate a likelihood ratio—a statistic measuring the probability that a given individual was a contributor to the sample against the probability that another, unrelated individual was the contributor." Ibid.

The defendant argues that STRmix is not reliable under the standard articulated in Olenowski I/Olenowski II because the underlying methodology is foundationally unreliable, unreliable as implemented by the specific laboratories that performed the analysis, and unreliable as applied to this case. The defendant further argues that, even if the underlying methodology employed by STRmix were sound, the actual software code that applies it is unreliable. The defendant has moved to exclude all DNA evidence as unreliable under N.J.R.E. 702.

The established fundamentals of DNA analysis are not the subject of this hearing. The fundamentals of DNA analysis are understood, and their reliability is well-settled in New Jersey. See, e.g., State v. Washington, 453 N.J. Super. 164, 185 (App. Div. 2018). The Legislature has determined "that DNA testing enhances the State's ability to positively identify an offender, to ascertain whether an individual may be implicated in another offense, and to establish positive identification in the event the offender becomes a fugitive." N.J.S.A. 53:1-20.18. The process of DNA extraction, amplification, quantification, and electrophoresis are common to all forms of DNA analysis, and STRmix does not enter into the equation until after the DNA sample has been reduced to an electropherogram ("EPG"). These steps, therefore, are not subject to a reliability hearing. Likelihood ratios ("LR" or "LRs") are also well-settled in case law. LRs are an established and reliable method of reporting a finding that compares competing hypotheses and are not a feature that is unique to STRmix or to probabilistic genotyping as a whole. Last, the concepts of cognitive bias or general human cognitive limitations are also not relevant to this hearing. If

STRmix works (reliably), then it works. If an individual expert's biases would cause them to process an individual sample improperly, then such a finding of fact should be made by a jury after a vigorous cross-examination.

The experts presented by the defendant fall, broadly, into two groups. The first group is forensic scientists who are wary of STRmix and probabilistic genotyping in general, to varying degrees. They range from staunch critics of STRmix to those who believe it needs to be developed further. The second group is computer scientists who claim that a comprehensive review of STRmix is necessary before it can be deemed reliable. Though they claim this review is essential, none of them have completed one in connection with this hearing. Instead, they have viewed the supporting documentation for STRmix, called it lacking, and gone no further.

Even if the defendant's experts have not proven that STRmix is unreliable, the State must still produce sufficient credible evidence to support a finding that STRmix is reliable under the relevant standards. The State's experts are developers and early proponents of STRmix, as well as forensic scientists who use STRmix in their day-to-day casework.

After hearing twelve days of testimony and argument, and after considering both pre-and post-hearing briefs, supplemental submissions, and over 180 exhibits submitted by both the state and the defendant, this court finds that STRmix is fundamentally reliable, that it is reliable as implemented by both the laboratories in question, and that it has been reliably applied to this case. While some of the samples may be ripe for a vigorous cross-examination based on the theories the State may advance at trial, none are of a kind that is beyond the capabilities of STRmix.

III. FACTUAL AND PROCEDURAL HISTORY

On November 20, 2018, while responding to a house fire in Colts Neck, New Jersey, emergency personnel found four residents deceased: [redacted] his wife [redacted] their

eleven-year-old son [redacted] and eight-year-old daughter [redacted] was found on the front lawn, and the remaining family members inside the home. [redacted] suffered five gunshot wounds. [redacted] suffered one gunshot wound to her head and multiple stab wounds to her torso. [redacted] and [redacted] were found with multiple stab wounds. [redacted] cause of death was later ruled to be smoke inhalation.

Earlier that day, a fire had also been reported at defendant's residence in Ocean Township, New Jersey. Detectives gathered evidence from both the Colts Neck and Ocean Township scenes. The investigation included forensic evidence that was secured, collected, and later tested for the presence of biological material. The crime scene evidence and known biological samples were sent for testing and comparison through traditional DNA analysis, and some were later tested by use of probabilistic genotyping software ("PGS"). Law enforcement collected DNA samples from each victim and from the defendant.

On June 18, 2020, the Monmouth County Prosecutor's Office sent crime scene evidence and DNA samples of the defendant and each victim to the New Jersey State Police for testing. After some samples were deemed unsuitable for testing, the items were sent to Bode Technology, a private laboratory. Bode processed the evidence using STRmix software v2.5.11. STRmix is a PGS that has been jointly developed by the New Zealand Institute of Environmental Science and Research ("ESR") and Forensic Science of South Australia ("FSSA"). This software utilizes probabilistic genotyping analysis to interpret complex DNA profiles, such as those with multiple contributors or with only trace amounts of genetic material. By leveraging statistical methods, the STRmix software generates millions of hypothetical DNA profiles, then evaluates the strength of these profiles against evidential samples and determines the most likely combination of profiles that can account for the observed DNA profile.

Bode examined thirteen items and generated a report on November 20, 2020. Most of the samples had two contributors, but the analysis by Bode Technology indicated that the defendant, [redacted] and [redacted] were, individually, all possible contributors to at least one sample.

On March 4, 2022, the defendant filed a motion to preclude the State from presenting Bode Technology's DNA results using STRmix during trial. On May 12, 2022, the court granted the defendant's motion to hold an evidentiary hearing on the admissibility of STRmix evidence. The parties were ordered to draft a protective order to facilitate the review of STRmix's source code. The terms of that order were discussed at multiple conferences and ultimately agreed upon on August 15, 2022.

At the August 15 conference, defendant's counsel estimated that the review would take one day, and expert reports based on the source code would follow shortly thereafter. The State also informed the court that it would be filing a certification that it was ready for trial and did so on August 25, 2022.

On September 19, 2022, the parties informed the court that they were working out the details of the code review with STRmix representatives and defendant's experts. The code review occurred from November 1 to November 3, 2022. On November 28, 2022, defendant's counsel confirmed that the code review was complete, but additional time was needed to complete expert reports. On December 12, 2022, defendant's counsel stated that defendant's expert reports would be completed by January 10, 2023. The State's responses to defendant's expert report were ordered due by January 24, 2023.

On January 23, 2023, defendant's counsel indicated he was in possession of three of their expert reports. Counsel indicated that the other two reports would be complete by February 3,

2023. The State informed the court they could complete their reports by February 20, 2023, if the defendant's expert reports were turned over by February 3, 2023.

On February 17, 2023, the New Jersey Supreme Court decided Olenowski I, 253 N.J. 133, adopting a Daubert-like[2] standard to assess the admissibility of expert evidence under N.J.R.E. 702 in criminal cases. The Olenowski I decision dramatically changed the focus of the expert reports, and of this hearing.

Numerous status conferences were held between February and May 2023 to discuss defendant's expert reports. In May 2023, the defendant's counsel informed the court they had retained an additional expert. On September 5, 2023, the parties returned to court and the defendant's counsel stated three out of their four expert reports were completed.

On June 7, 2023, the State provided samples previously analyzed by the New Jersey State Police Laboratory in 2020 for testing using STRmix. On September 29, 2023, the New Jersey State Police Laboratory issued a report containing analysis of two samples conducted using STRmix v. 2.8.0. The State provided this report to the defendant on October 3, 2023.

On February 12, 2024, the defendant advised that he objected to the State's intended proffer of any reports generated by STRmix v. 2.8.0. He indicated that the experts were only prepared to discuss the analysis generated by Bode Technologies using STRmix v. 2.5.11.

On February 22, 2024, the defendant filed a motion to adjourn the February 26, 2024, Olenowski hearing, and sought to expand the scope of the hearing to include STRmix v2.80. On February 23, 2024, the court denied that motion.

---

[2] Daubert v. Merrell Dow Pharms. Inc, 509 U.S. 579 (1993).

On February 29, 2024, the Appellate Division affirmed the trial court's decision not to adjourn the evidentiary hearing. The defendant appealed to the New Jersey Supreme Court. On March 22, 2024, the New Jersey Supreme Court reversed the denial of the adjournment request and directed the court to expand the scope of the hearing to include both STRmix v. 2.5.11 and STRmix v. 2.8.0. State v. Caneiro, 256 N.J. 531 (2024).

On March 26, 2024, a new protective order regarding the source code for STRmix v. 2.8.0 was discussed and ultimately signed by the court on April 8, 2024. On April 23, 2024, the defendant, through counsel, confirmed that his experts were scheduled to conduct their source code reviews of STRmix v. 2.8 in May 2024, which did occur.

On June 25, 2024, the defendant advised that the expert reports would be completed "soon." On July 16, 2024, the defendant provided an update on the expert reports. One report was finished, another would be finished by the end of the week, and the last report would be finished by the end of July. The defendant also advised that he would be retaining a statistician as an expert.

On July 30, 2024, the defendant indicated that he was still waiting on expert reports. Counsel stated that they were days away from receiving Dr. Karl Reich's report, and weeks away from Mr. Keith Inman's report. The court set the next status conference for August 9, 2024.

On August 9, 2024, the defendant indicated that Dr. Reich's report would be submitted by the end of the next week. The defendant also stated that reports from Mr. Inman and a Dr. Rabinowitz, who ultimately did not submit a report or testify in this hearing, would be submitted by the end of August, or early September. The court set the next status conference for September 4, 2024.

Due to a lack of progress with the experts' reports, the matter was transferred to this court, which conducted the September 4, 2024, conference. At that time, based upon the representation

of defendant's counsel, Dr. Reich's report had been submitted, but Mr. Inman's and Dr. Rabinowitz's reports still had not. The two reports were ordered to be completed by September 27, 2024, and if not completed by then, the experts were ordered to appear before the court to explain the delay.

On September 23, 2024, defendant requested an adjournment to have more time for expert reports to be completed. The adjournment request was denied. On September 27, 2024, the defendant advised this court that he would be proceeding with Mr. Inman's report and that it would be submitted by September 30, 2024. Dr. Rabinowitz's report would not be submitted. The State was given until October 18, 2024, to respond to Dr. Inman's report. This hearing was scheduled for November 12, 2024.

Both parties provided updated submissions in preparation for this hearing which began on November 12, 2024, and concluded on December 9, 2024. On December 13, 2024, this court heard oral argument, and both parties submitted written summations on January 17, 2025.

IV. SUMMARY OF PRE- AND POST-HEARING BRIEFS

a. State's Brief

On November 3, 2024, the State submitted their brief in response to defendant's motion to exclude DNA evidence analyzed utilizing STRmix software. The State argued that STRmix is not new technology by emphasizing that eighty-four laboratories across the United States have internally validated STRmix and are actively using it. Additionally, more than nine years have passed since SWGDAM issued its Guidelines for the Validation of Probabilistic Genotyping Systems. The Union County Prosecutor's Office, for example, has been using STRmix in their DNA laboratory for seven years.

The State further argues that STRmix has been testified to in New Jersey State and Federal Courts about thirty-three times. The State notes other courts' findings regarding STRmix to be admissible. The State "submits that STRmix evidence clearly meets the [Olenowski] standard and that the challenges put forth by the defense" are proper "for adversarial examination, not grounds for exclusion." For example, many of the specific factual challenges to the use of STRmix in this case, such as individual analyst's decisions or specific details regarding the relationship between the persons suggested to be in the mixture, are questions for a jury and go to the weight, not the admissibility, of the evidence.

The State further notes that the defendant's experts have completed multiple source code reviews. These reviews were to examine the "depths of [the] STRmix code and the depths of the data underlying the validations by the New Jersey State Police and Bode." According to the State, these source code reviews demonstrate that STRmix works. STRmix maintains a list of all discovered miscodes, describing how they were discovered and how they were fixed; no errors outside that list have been noted, and that the defendant's review did not reveal any miscodes or bugs.

The State counters defendant's argument that the STRmix analysis conducted by Bode was outside of its validation by stating that the validation summary supports the analysis done in this case. While the defendant has argued that Bode did not perform any tests of mixtures with minor contributors below 5%, there was testimony at trial that Bode tested some mixtures as low as 3%. Further, the State argues that familial relationships do not decrease the reliability of STRmix's conclusions. The State offered Dr. John Buckleton to show that the reliability is not affected. The State submits that all samples sent to Bode and the New Jersey State Police laboratory were reliable and admissible.

Finally, the State submits that STRmix has been widely accepted as reliable and admissible in U.S. State and Federal courts. Relying on a Daubert-like analysis, the State submits that these courts have found that STRmix, after being subjected to testing and peer-review, is generally accepted by the scientific community. As a result, it has been found reliable and relevant.

b. Defendant's Brief

The defendant contends that STRmix must be shown to meet the standards of reliability in not only the field of forensic biology but also software engineering. Even if the underlying science of STRmix is sound, the software that runs it must be reliable in its operation or else the output of the program cannot be trusted. Defendant argues that STRmix is deficient, and insufficiently tested, in both these regards.

The defendant characterizes STRmix as a "safety critical system," a class of software programs that have the potential to cause catastrophic and irreparable harm if they fail in their intended function. Defendant also argues that safety critical systems must be subjected to a specific procedure in their design, development, and testing, and that systems which do not conform to those standards are *per se* unreliable. Since, in defendant's view, STRmix was not put through this process of independent validation and verification, it can never be found reliable in a New Jersey court.

Even if independent validation and verification were not required, defendant argues STRmix fails to meet the four prongs of the Olenowski analysis. Since the vast majority of peer-reviewed papers on STRmix have been written by its developers, their close colleagues, or scientists affiliated with laboratories that have purchased STRmix, and since these individuals have an interest in seeing the program succeed, any claim by them cannot be considered a "peer-reviewed

publication." Defendant argues that the standards that govern probabilistic genotyping are insufficiently developed to support reliability, and that the error rates and testability of STRmix have not been conclusively established. Consequently, defendant believes that STRmix has not met general acceptance in the relevant scientific fields.

Assuming, arguendo, the court were to find STRmix reliable, the defendant submits that it is not reliably applied to this case. Broadly, the defendant claims the internal validation studies conducted by these laboratories are woefully deficient and these laboratories have no business analyzing any DNA samples at all with STRmix. Since the matter before the court involves DNA samples that are believed to have come from related individuals and involve, in some cases, trace amounts of DNA, defendant also raises concerns that these samples are simply outside the boundaries of what DNA science can ever hope to analyze.

In support of all these arguments, the defendant points to the software validation procedures found in other industries and claims that STRmix falls short. The defendant also argues that the laboratories in question lack sufficient standard operating procedures to properly test samples of this kind and have either failed to set reliable limits for STRmix's use or have failed to adhere to them. The defendant expresses concern that the types of samples analyzed in this case are problematic and prone to mislead a jury, and that the methods of analyzing these types of samples and the preferred method of reporting them to a jury are both misleading and vulnerable to cognitive bias.

The defendant notes that the State's burden to prove STRmix's reliability requires the prosecution to justify gaps in the data and address the shortcomings he has identified; the defendant offers opinions of several experts, who testified to the inherent flaws in STRmix.

V. DNA- GENERALLY

"Deoxyribonucleic acid (DNA) is a molecule of genetic materials shaped like a double-helix or spiral ladder. In every person, each cell with a nucleus contains a copy of that person's DNA. Thus, DNA serves as a blueprint for the human body." State v. Harvey, 151 N.J. 117, 156 (1997). The order of the base pairs along the DNA molecule comprises an individual's genetic code. Human DNA contains approximately three to four billion base pairs, known as the "genome." These base pairs govern the production of bodily proteins. Id. at 156-57. A gene is a sequence of nucleotides on a DNA strand responsible for producing a particular protein. The sequence of the nucleotides can vary. The possible sequences or variations are called "alleles." Thus, an allele is simply a version of a gene. A gene's position on a chromosome is its locus. In different individuals, genes may be "polymorphic," meaning that they may take different forms or contain different sequences of base pairs. The polymorphic genes, which vary from one person to another, provide the basis for DNA identification. Id. at 157. Detecting differences in alleles at corresponding loci is an essential component of human identity testing. John M. Butler, Fundamentals of Forensic DNA Typing 25 (2010) (hereinafter "DNA Typing"). DNA profiling is the process of determining the genotypes present at specific locations within the DNA molecule. Id.

The human genome consists of twenty-two pairs of autosomal chromosomes and two sex-determining chromosomes. Id. at 23. A normal human cell contains forty-six different chromosomes or twenty-three pairs of chromosomes. Ibid. One chromosome in each pair is inherited from each parent. Id. at 24.

DNA evidence was first introduced in the courts in the late 1980s. President's Council of Advisors on Sci. & Tech., Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods 25 (2016) (hereinafter "PCAST Report"). DNA analysis involves comparing DNA profiles taken from evidence to profiles taken directly from a person of interest.

Id. at 69. This comparison is done to see if the known sample may have been the source of the evidentiary sample. A DNA profile is created by chemically extracting biological material from a sample. Then a series of predetermined loci are amplified using polymerase chain reaction (PCR), an enzymatic process that replicates a targeted DNA segment to create millions of copies. Once amplification is complete, the lengths of the DNA fragments are measured using capillary electrophoresis, and this raw data is analyzed to create an EPG. Electrophoresis operates on the principle that longer DNA fragments will move more slowly through a polymer solution than a shorter fragment will, allowing for the alleles present to be plotted along a graph according to size. A person may either inherit the same allele from both parents or a different one from each. Thus, alleles in a pair might register as the same or different heights on an EPG.

Once an EPG is generated from a sample, the DNA profile can be compared against other profiles to determine if they match. DNA Typing at 9. Each person has their own unique genotype. Forensic DNA analysis involves comparing a known person's (or people's) genotypes to a sample from a relevant scene. For a single-source mixture, the analyst must determine if there are any mismatched alleles at a particular locus – if so, they could not have come from the same person. So, for example, if Sample A has a 2,4 at a specific locus and Sample B has a 3,5 from that same locus, the DNA did not come from the same person. If the genotype is the same, then it must be determined how common that genotype is and how common the combination of all the genotypes in a sample are to determine how likely it is that a specific person is the source of the DNA.

The result of this forensic comparison is documented in a laboratory report. DNA Typing at 11. This report represents a summary of the work conducted by a DNA analyst, who performs their work based on standard operating procedures ("SOPs") promulgated by the laboratory where they are employed. DNA Typing at 11-12.

When the questioned sample contains DNA from only one person, it is possible for a trained analyst to compare it directly against a known profile without computer assistance. When a sample contains DNA from two people, it is more difficult to determine which alleles were contributed by which individual, since the EPG does not distinguish between contributors. The first step in accomplishing this is estimating the relative ratio of components in each mixture—how much genetic material from each individual is present. If the difference between their contributions is large enough, it is possible for an analyst to assign the individual alleles to either the major or minor contributor. Only then can the analyst compare these interpreted components to reference profiles for inclusion or exclusion. If the contributors to a two-person mixture are roughly even, it is usually impossible for the analyst to assign a given allele to one contributor or the other, and the sample is deemed not suitable for comparison. Manual DNA interpretation undertaken by a human is largely unable to deal with samples any more complicated than a simple two-person mixture.[3]

a. Probabilistic Genotyping

Samples that contain three or more individuals are generally too difficult for a human being to interpret without assistance, but probabilistic genotyping was invented to try to address the problems of working with these "complex" mixtures that present these challenges.

> There are two methods of probabilistic genotyping—semi-continuous and continuous. Semi-continuous probabilistic genotyping methods use the observed peak heights and incorporate a probability of allele drop-out and drop-in to explain missing or extra alleles. However, such methods do not take into account other variables such as peak height ratios, mixture ratios, and stutter percentages in the calculation of the likelihood ratio. In contrast, continuous methods utilize all peak heights and

---

[3] When discussing mixtures comprised of two or more people, it is common to refer to the person whose DNA makes up the highest portion of that mixture as the "major" contributor, and all other contributors as the "minor" contributor. Proportions of DNA in a mixture can be expressed as either a ratio or a percentage (i.e., a two-person mixture with a 20:1 ratio could also be described as a mixture with a 95% major contributor and a 5% minor). Whenever possible, this opinion will endeavor to express contributor proportions as percentages.

> do not require the analyst to determine whether a given peak is allelic, stutter, or over-stutter but analyzes the observed information. Stated otherwise, the continuous method uses more of the information that is present in a DNA sample than the non-continuous method.
>
> Probabilistic genotyping systems generate what is known as a likelihood ratio. The LR compares how likely the observed DNA evidence is under hypotheses regarding the source of the DNA. Typically, though not always, the LR compares the probability of obtaining the observed DNA profile under the hypothesis that the defendant contributed to the DNA mixture being analyzed (the "inclusion hypothesis") with the probability under the hypothesis that the defendant did not contribute to the DNA mixture (the "exclusion hypothesis"). The likelihood ratio considers how likely the observed findings are under the two hypotheses. In the case of a complex, multi-contributor mixture, as in this case, the likelihood ratio is phrased as follows: This mixture is greater than X times more likely if it originated from the defendant and three unknown unrelated individuals than if it originated from four unknown unrelated individuals.
>
> United States v. Lewis, 442 F. Supp. 3d 1122, 1139–40 (D. Minn. 2020) (internal citations and quotation marks omitted)

Since an EPG is not capable of indicating the source of a given allele, and since it is possible for a human being to not only have two copies of the same allele at a given loci but also to have the same alleles as another person, there are countless combinations of alleles possible across the loci typically examined in DNA analysis, and countless permutations of potential contributors that could produce a particular sample. PGS attempts to determine whether it is more likely than not that a particular individual contributed to a DNA mixture, by using a Markov Chain Monte Carlo (MCMC)[4] process to determine which of those permutations appears to be the best fit. Dr. Buckleton described Markov Chain theory, developed in 1906, as a "memoryless walk" where

[4] Although the implementation of these algorithms and their applicability to DNA analysis has been challenged, the underlying concepts of Markov Chains and statistical analysis are not new. Predictive text messaging functions, for example, often use Markov Chains to determine what words are most likely to come next based on the previous words entered, effectively "remembering" only the most recent information to make its predictions.

each step in a deductive process is made without considering all the steps made before. (7T 37-4).[5] As explained by Dr. Buckleton, the software uses MCMC to run possible genotyping combinations to see what is best used as evidence. (6T 60-61). The programs attempt to determine which combination of DNA profiles from hypothetical contributors is the "best match" to explain the DNA observed in the mixture. MCMC and wandering chains allow for thousands or millions of different explanations for the DNA observed to be tested. (7T 183 to 186).

> The MCMC deconvolution process proceeds in two phases. During the first phase, called "burn-in," the MCMC is run to ensure that in the second phase—called "post burn-in"—the MCMC begins in an area of high probability space. Burn-in begins in a randomly chosen genotype set and fixed mass parameters. Once 100,000 iterations (steps) have been accepted the MCMC is said to have "burned in."
>
> Post burn-in begins with genotype and mass parameter proposals that already start from these higher probability densities. The post burn-in process is undertaken until the iterations also reach a set number of acceptances. The genotype sets accepted during post-burn-in are tallied. At the completion of the MCMC these values are

---

[5] Transcripts from the hearing span multiple days and are cited to using the following legend:

1T 11/12/24 a.m. (Ghannam)
2T 11/12/24 p.m. (Ghannam and Naughton)
3T 11/13/24 a.m. (Naughton and Reed)
4T 11/13/24 p.m. (Reed)
5T 11/14/24 a.m. (Reed)
6T 11/14/24 p.m. (Buckleton)
7T 11/15/24 a.m. (Buckleton)
8T 11/18/24 a.m. (Thayer)
9T 11/18/24 p.m. (Thayer and Schlenker)
10T 11/19/24 a.m. (Coble)
11T 11/19/24 p.m. (Coble)
12T 12/2/24 a.m. (Reich)
13T 12/2/24 p.m. (Reich)
14T 12/3/24 part 1 (Heimdahl)
15T 12/3/24 part 2 (Heimdahl)
16T 12/4/24 a.m. (Adams)
17T 12/4/24 p.m. (Adams)
18T 12/6/24 (Martin)
19T 12/9/24 (Inman)
20T Final Summations

> normalized at each locus so that they range from zero (indicating that the observed data cannot be explained by the proposed genotype set) to one (indicating that this is the only genotype set that explains the data).
>
> [Lewis, 442 F. Supp. 3d at 1142–43.]

Probabilistic genotyping "enables weighting (based on the probability of) specific genotype contributions through biological and statistical models informed by probabilities of missing alleles. These methods incorporate mathematical modeling that can reflect uncertainty in the mixture interpretation." National Institute of Science and Technology (NIST), DNA Mixture Interpretation 34-35 (2024) (hereinafter "Mixture Interpretation").

> Fully continuous probabilistic genotyping software (PGS) . . . unlike binary and semi-continuous models, use quantitative information contained within a profile (e.g. allelic designations, peak heights, molecular weights/fragment length), take into account stochastic effects, model peak height variability, and allow interpretation of low-level and complex DNA mixtures, therein reducing the need to infer using subjective reasoning.
>
> Riman et al, Comparing performance and likelihood ratios for different PG models, at 2.

After the deconvolution process, a PGS (such as STRmix) goes on to interpret the evidence. Its final output is an LR, expressed as a fraction. "The LR involves a ratio of two conditional probabilities: the probability of the evidence given that one proposition (hypothesis or narrative) is true, and the probability of the evidence given an alternative proposition is true." Mixture Interpretation at 48. An LR above 1 provides support for the first hypothesis, while an LR less than 1 provides support for the alternative hypothesis. Ibid. LRs are just comparisons between these two hypotheses and do not suggest that either is likely to actually be true. They are incomplete without reference to what is referred to as a "prior probability," a pre-existing calculation that either hypothesis could have occurred in real life. NIST, Forensic DNA Interpretation and Human Factors 81 (2024) (hereinafter "Human Factors"). Bayes' Theorem is the mathematical framework that

guides the use of LRs, and under this theorem, the LR must be multiplied by a prior probability to arrive at the "posterior probability," the odds that something occurred. Ibid.

In manual DNA analysis, the statistical result that is returned is a "random match probability" (RMP)—the probability that someone else in the population, chosen at random, would have the same genotype as the contributor of the forensic evidence. PCAST Report at 72. By contrast, an LR measures the strength of support for one hypothesis compared to another hypothesis to assess which better explains the observed evidence. In other words, how much more (or less) likely it is that Person A contributed to a mixture, rather than another person.[6] This is not the likelihood that a person *did* contribute to a mixture, merely a way of quantifying how similar their DNA is to that which is reported on the EPG.

"[D]epending on the software being used, the interpretation of the same DNA profile could yield different numeric LR values and, if used, different verbal characterizations. Even if the same PGS is used, the overall LR system could be different and hence will lead to different LRs." Riman at 2. Since LRs are variable across systems, and somewhat variable even within the same system, they are often reported by DNA analysts in terms of a "verbal scale" where different ranges of LRs are assigned uniform language that indicates how much support they provide for a particular hypothesis. The SWGDAM guidelines provide one example of a verbal scale, though some laboratories adopt a more conservative threshold for their reporting. LRs should not be interpreted as dispositive evidence that a defendant's DNA was present in a sample, they are merely statements to illustrate the fact that pieces of DNA *similar to* a defendant's DNA were present, and how unexpected that may or may not be. This is true of a conventional RMP as well: it does not mean

[6] That person may be a specific individual, a related individual, or a random individual.

much to say someone's DNA is a potential match to a sample unless there is also a way to show that person's DNA had a way to turn up in that sample.

"Forensic genetics is experiencing an increase in data volume and complexity, and the interpretation of this data is becoming more and more dependent upon the use of appropriate biostatistical computer programs." Coble et al, Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, 25 Forensic Science International: Genetics 191, 192 (2016). Software for calculating LRs "to evaluate trace evidence or competing kinship scenarios has been in use for many years now," and several groups have "described validation exercises of either in-house, open source, or commercial software packages." Ibid.

VI. RULE 702 HEARING

a. Summary of Expert Testimony and Credibility Determinations

i. State's Expert: Monica Ghannam, Expert in the Fields of Forensic DNA Analysis and Forensic Validations

The first witness before the court was Monica Ghannam. Ms. Ghannam has been employed by the Union County Prosecutor's Office Forensic Laboratory ("UCPO") since 2005 and currently serves as a DNA technological leader. (1T 11-22).

Ms. Ghannam began her testimony by describing the DNA analysis process, which is comprised of four steps. (1T 47-15). Step one is extraction, removing the DNA from the nucleus. (1T 47-16). Step two is to quantitate the sample to see how much DNA is present and ensure it meets the minimum threshold for analysis. (1T 47-25). If there is enough DNA, step three is amplification; millions of copies are made of the twenty-four predetermined loci. (1T 48-13). Step four, DNA is subjected to electrophoresis and peak heights are established. This will create the DNA

profile. (1T 48-21). STRmix has not changed the way this four step laboratory process works. (1T 49-8).

Ms. Ghannam then discussed UCPO's use of STRmix. The UCPO has been using STRmix since 2016 and is accredited with both the FBI Quality Assurance Standard and with ISO 17025. (1T 25-25 to 38-16). ISO 17025 is a general standard that deals with issues of quality and testing for procedures. Ghannam has personally operated STRmix since 2017. LRs are not new to forensic DNA and have previously been used in paternity cases. (1T 63-20). Ms. Ghannam testified that SWGDAM supports using a verbal scale and that her laboratory follows this scale. (1T 66-6). If a LR gives exclusionary support, the same scale will be used. (1T 38-20 to 67-11).

UCPO was interested in STRmix because their previous system was limited to a maximum of three contributors, and then only under limited circumstances. (1T 34-23). Ms. Ghannam's laboratory does not analyze five- or six-person mixtures using STRmix. Union County has internally validated three versions of STRmix: v. 2.4.6, 2.5.11, and 2.7.0—the version UCPO currently uses. (1T 38-16). UCPO uses new versions because they are interested in new features, not due to issues with older versions. (1T 37-24). Ms. Ghannam has testified about STRmix results in New Jersey courts fifteen times, and in Federal court once. (1T 44-20). She has been qualified as an expert in all those proceedings.

Ms. Ghannam gave a preliminary explanation of some common terminology in the Forensic DNA community, including analytical threshold, stutter, stutter filters, drop-in, drop-out, stochastic thresholds, and saturation. (1T 53-4 to 59-19). An analytical threshold is used to separate "noise" out of an EPG by disregarding peaks below a certain RFU—anything below that threshold may not be a "true" DNA peak. Stutter is a byproduct of the amplification process that can make

a "false" peak appear very close to a true peak. Stutter filters are a software-based method of excluding stutters from an EPG. Drop-in generally occurs when an allele that was not part of the evidentiary sample appears on the EPG, from contamination or otherwise. Drop-out occurs when an allele that is present in the sample does not appear on the EPG, usually because there were very few copies of it and they evaded collection. Saturation occurs when there is too much post-amplification DNA and the instruments used to generate the EPG are overwhelmed, like shining too bright a light in someone's eyes. All of these and more, collectively, are referred to as stochastic effects: random or chaotic processes that can alter readings in a laboratory but are difficult to account for. (1T 52-24 to 59-19). None of these phenomena are novel to STRmix. (1T 59-19).

Ms. Ghannam also testified about the validation her laboratory performed on STRmix v. 2.4.6. (2T 28-10). Internal validation allows for the laboratory to determine the limits of the technology. (1T 21-5). Internal validations establish her laboratory's ability to reliably use a technique. (1T 36-16 to 22). For the internal validation, the UCPO's laboratory sent the mixtures and the electropherograms to ESR. (2T 28-20). Ms. Ghannam did review all of the electropherograms prior to sending them to ESR, as it is important to see how STRmix evaluates the data. (2T 31-22 to 32-2). ESR calculated her laboratory's parameters for STRmix, so that they could move on to the second part of the validation study. (2T 28-10). SWGDAM guidelines were applied explicitly during this portion of the validation study. (2T 28-10). No one from her laboratory conducted the deconvolutions themselves. (2T 28-23). All the charts and graphs for the internal validation were created by STRmix and ESR. (2T 29-1-6).

Ms. Ghannam testified about mixture complexity. (2T 35-22). Mixture complexity is a catch all term for determining who contributed to a mixture. (2T 35-36). A mixture having more

contributors in roughly equal parts makes it more complex. (2T 36-4). If there is one major contributor and several minor contributors, it can be hard to distinguish between minor contributors. (2T 37- 14). Degradation can also cause a mixture to be more complex. (2T 36-23). This may occur when one contributor is more degraded than another. (2T 36-37). When contributors to a mixture have the same alleles as a given locus, this can make a mixture more complex and increase the risk of false inclusions. (2T 37-38). Related individuals can share many alleles with each other, which can be a source of further complexity. (2T 38-7 to 13). According to Ms. Ghannam, related contributors, low template profiles, and degradation can cause an analyst to underestimate the number of contributors ("NOC") when performing an analysis. (2T 38-39).

Thus, UCPO limits the NOC it will test in a sample. (2T 39-16). The UCPO laboratory tested two, three, and four person mixtures in its internal validation study for STRmix. (2T 39-19). As a result, Ms. Ghannam, and other analysts, are not permitted to analyze five and six person mixtures. (2T 39-23). Additionally, Ms. Ghannam explained that her laboratory is limited by DNA concentrations. (2T 40-4). It is necessary to obtain a certain amount of picograms, otherwise the sample needs to be amplified again. (2T 40-4). If the electropherograms have at least one allele and five loci, her laboratory can use it for comparison. (2T 41-12). If a minor contributor has two percent or less of the overall mixture proportion, the laboratory will not use it for comparison. (2T 41-22).

Ms. Ghannam was offered, without objection from the defendant, as an expert on forensic DNA analysis generally, not to give an opinion on any evidence analyzed in this case. She did not review the New Jersey State Police laboratory or Bode laboratory analysis. She did not review any of the deconvolutions or LRs generated in this case. (2T 48 to 49). Ms. Ghannam spoke positively

of STRmix and views it as being "open" to analysts: there is no "hidden information" that an analyst would be lacking when evaluating a sample. (1T 109-22 to 24).

The court found Monica Ghannam to be a credible witness. During testimony, Ms. Ghannam was able to sufficiently answer and aid the court in understanding the process of DNA analysis. The court found Ms. Ghannam's testimony to be reliable. She had no interest in the outcome of this hearing, and even though she testified favorably about STRmix, she was objective in her answers on direct and cross examination. Ms. Ghannam explained the methodology of DNA analysis and demonstrated how STRmix's methodology and testing are reliable. Her testimony was given weight, but its relevance was limited to how laboratories validate and use STRmix in a general sense.

ii. State's Expert: Kristen Naughton, Expert in the Fields of Forensic DNA Analysis and Forensic DNA Validations

Kristen Naughton is the Director of Validation Studies at Bode Technologies. (2T 56-1). She has been working there since August of 2023. (2T 56-15). As Director, she supervises a team of forensic DNA analysts (2T 56-57). Ms. Naughton has performed hundreds of validations and continues to perform validations in her role as Director of Validation Studies at Bode. (2T 61-23). An internal validation ensures that the software formulas are being applied appropriately and that results are accurate. (2T 99-3).

According to Ms. Naughton's testimony, Bode limits the use of STRmix to samples with four or fewer contributors, based on the results of their validation study. (2T 113-7). It was Ms. Naughton's opinion STRmix is suitable for complex and low-level mixtures. She indicated that Bode can determine the NOC with a level of certainty by looking at a DNA profile. (3T 90-3).

However, it is possible that in lower template samples, an analyst can underestimate the NOC. (3T 89-23).

Bode has SOPs that address how to proceed if a contributor is known or expected to be in a mixture: If a sample is taken directly from a person's body or from an object of clothing worn by them, then Bode considers their DNA to be "expected" and can analyze samples with their specific alleles already taken into account. (3T 28-3). Bode's laboratory is accredited by ANSI National Accreditation Board, ISO 17025, and the FBI Quality Assurance Standards. (2T 66-23 to 67-1).

During internal validation, Bode conducted experiments to determine the effect of under- or overestimating the NOC in a sample. (3T 31-3). When an analyst put in too few contributors, STRmix often marked the sample as "not plausible" and stopped testing. (3T 31-9). If it was able to run with the wrong NOC, it was usually because there was a major contributor that had a strong template. (3T 31-15). When overestimating contributors, the LRs for the major contributors were unimpacted but minor contributors had lowered LRs. (3T 32-1).

Ms. Naughton testified as to what occurs if there are extra peaks (drop-ins) and how STRmix accounts for them. (3T 38-3). For drop-ins, STRmix granted them no weight until it hit a certain threshold, at which point STRmix stopped and defined the sample as impossible. STRmix doesn't "see" the peak heights directly, it only sees the numerical value as it is entered. (3T 38-17). Lower peak heights reduce what STRmix can interpret due to stochastic effects; high peaks are "tighter and more reproducible" and less subject to chaos. (3T 43-18). An analytical threshold is the setting that can be used to determine if a signal being detected is a true allelic peak or background noise; Bode has set their threshold at 125 RFUs. (2T 87-13).

Ms. Naughton discussed how contributors being related will affect the STRmix results and how the laboratory responds to it. (3T 74-13). Relatedness is a contributing factor to allele sharing. (3T 72-21). It is not the only factor, but it is a source of complexity. (3T 73-2). There are published papers that show that the number of related individuals in a sample have different complexities. (3T 74-13). It is possible to underestimate contributors if you have, for example, a parent and child mixture. (3T 74-75). If relatedness affects interpretation, it may limit Bode's analysis, but it may not serve as an absolute bar. (3T 75-8). For mixtures with first-degree relatives, an "analyst should scrutinize the data to ascertain if the peak height ratios indicate allele sharing to the extent that the deconvolution may not be intuitive." (3T 101-9).

Ms. Naughton did not express any concerns or issues with STRmix's reliability as it operated in her laboratory, or concerns with the findings and conclusions that analysts at Bode generated using STRmix's output. Ms. Naughton's 20 years of experience as a scientist at Bode Technologies was noted by the court.

The court found Kristen Naughton to be a reliable and credible witness. Ms. Naughton has extensive experience in validations and DNA analysis. Ms. Naughton has performed hundreds of validations and serves as the Director of Validation Studies at Bode Technologies. She testified knowingly and with sufficient detail about how STRmix works and how STRmix generates results. Ms. Naughton was able to thoroughly answer questions on both direct and cross examination. Ms. Naughton's professional experience with STRmix allowed her to explain the benefits and drawbacks of the software with objective reasoning. She testified about internal validation studies, different versions of STRmix her laboratory have used, how an analyst puts values into STRmix, and about how relatedness can affect STRmix results. Her opinion was given great weight regarding STRmix's internal validations and Bode's laboratory practices.

iii. State's Expert: Danielle Reed, Expert in the Field of DNA Analysis

Danielle Reed currently works as a forensic DNA analyst at Bode, where she reviews DNA samples, testifies, and peer-reviews other analysts' work. (3T 132-11). She has used STRmix thousands of times and has testified to its results in court sixteen times.

In this case, Bode received a pair of pants, six latex gloves, knife, and swabs from the crime scene. (4T 18-3). Before the evidence was received, Bode received developed DNA profiles for the five persons of interest to be compared to the samples. (4T 18-14). These profiles were of the four victims and the defendant, Paul Caneiro. (4T 18-14). Bode utilized STRmix for thirteen samples. (4T 19-4 to 12). STRmix cannot be used when there is no DNA present, which occurred in sample EO4B (interior of a recovered glove). (4T 82-18).

Before analyzing a sample, Bode tries to determine the NOC based off the details of the investigation. (4T 27-19). Then, once there is a NOC, each profile will be first tested separately. (4T 47-5). If profiles are found to be present, the sample will be then tested for combinations. (4T 49-10).

Ms. Reed knew the five DNA profiles were relatives, due to their last name. (4T 107-15). Reed did not know how they were related to each other, or whether they were first-degree relatives or more distant. (4T 107-18). A first-degree relation refers to either a parent/child or two full siblings. (4T 108-1). Besides identical twins, relatives can share many, but not all, alleles. (4T 109-23). For relatives, Ms. Reed expects peak heights to be higher than normal, which did occur in this

case. (5T 62-23). Ms. Reed testified that the challenges posed by first-degree relatives were mitigated in this case by the presence of a suspected mutation in one of the contributors. (5T 82-20).[7]

On cross examination, defendant's counsel, Mr. Godin, ran simulated tests of different samples while Ms. Reed followed along and confirmed at each step that Godin was performing the test correctly. (5T 19-1). Counsel spent a considerate amount of time working with E06b. (5T 21-8). In this sample, [redacted] was excluded from the sample and Ms. Reed was questioned as to why this occurred. (5T 22-11). [redacted] allele table was compared against the EPG and whether he was visually included or excluded at any locus was confirmed one locus at a time. (5T 22-12 to 25-15). Sample E06b was then run with [redacted] as a proposed contributor, which yielded an inclusionary LR of 9,730—that is to say, STRmix reported that it was 9,730 times more likely that [redacted] had contributed to the mixture than an unknown, unrelated individual. (5T 25-16 to 45-17). This number is below Bode's reporting threshold but is nonetheless inclusionary; it was also noted that the report was inclusionary at locus D12, where Ms. Reed had previously visually excluded [redacted] (5T 45-18 to 48-18).

Initially, Ms. Reed was unsure if there had been an error on her part when she originally ran the sample, or why else this would have happened. (5T 49-9 to 51-7). Ms. Reed was very concerned whether she had made a mistake that had not been caught by her technical reviewer, and characterized her concern as "freaking out" (5T 51-12). After a brief recess where Ms. Reed took a closer look at the reports and her notes, she was able to explain that the demonstrative exhibits had been "go[ing] a little too fast" and she had misspoken slightly. (5T 52-9). She then

---

[7] [redacted] has a mutation, a gene she does not share with any other member of her family. A mutation occurs during embryonic cell division and causes an allele to shift positions on an EPG. (4T 107-3). This allele shift was detected in the samples analyzed by bode, which helped Ms. Reed to differentiate that contributor. (5T 82-20).

clarified that [redacted] should not have been excluded at locus 12, because locus 12 contained the placeholder "Q" allele. At the time of the original STRmix report, Ms. Reed excluded potential contributors at loci with a "Q" allele if the other allele did not match, but her and her colleagues had determined in the intervening years that the more prudent choice would be not to allow for any exclusions based on loci with placeholder alleles. (5T 53-4 to 54-9).

At time of testing, the laboratory had above a 99% confidence that [redacted] was excluded. (5T 52-7). When the samples for this case were processed, Bode's SOP allowed for this level of confidence to be enough to exclude a profile. (5T 52-17). The current policy does not recommend exclusion in these circumstances, so [redacted] would no longer be excluded from the test mentioned above. (5T 52-17).

The court found Danielle Reed to be a credible witness. Ms. Reed is a forensic DNA analyst at Bode Technologies who has used STRmix to test thousands of DNA samples throughout her career and has testified about the program approximately sixteen times. In this case, Ms. Reed utilized STRmix for thirteen samples and did manual interpretation for one sample. Her laboratory uses both STRmix and manual interpretation. Ms. Reed discussed how manual interpretation has a critical role in DNA analysis. Ms. Reed was extremely helpful to the court's understanding on whether a sample is suitable for STRmix or manual interpretation.

The court found Ms. Reed to have extensive knowledge about how Bode Technologies performs visual exclusions of samples, the limits Bode has for STRmix, and how STRmix was used in this case. Her answers on both direct and cross examination were consistent, forthright, detailed, and knowledgeable. On cross-examination, Ms. Reed walked defense counsel through a simulated test of different samples using STRmix. Ms. Reed went step by step through the process of entering data into STRmix with defendant's counsel to show how STRmix generates its results.

On cross-examination, Ms. Reed was asked whether it was appropriate to exclude [redacted] Caniero from a particular locus on a particular sample and she paused while giving her answer. (5T 51-11). Her pause resulted from defendant's counsel performing calculations with STRmix that she had not previously run. While Ms. Reed needed time, her answers were not evasive or deceptive. (5T 51-18). Ms. Reed showed no intent to deceive the court. She was objective with STRmix's conclusions and understood the limitations her laboratory has in using the software.

The court also notes that Ms. Reed further explained how Bode's SOPs have evolved over time. (5T 53-11). While it was originally consistent with Bode's SOP to exclude [redacted] they would now employ a more inclusive approach and not exclude persons of interest unless they are ruled impossible. (5T 54-4). This approach is based on using STRmix thousands of times and recognizing how Bode's conservative approach could be modified. If the tests were performed now [redacted] would not have been excluded from Sample E06B. This further highlights the fact that Ms. Reed's pause on cross-examination did not negatively impact her credibility. Ms. Reed's training in using STRmix and her experience with analyzing the samples in this case are afforded great weight.

iv. State's Expert: Dr. John Buckleton, Expert in the Fields of DNA Analysis, Probabilistic Genotyping, and Software Development

Dr. John Buckleton is a New Zealand resident and forensic scientist. (6T 4-15). He works for the New Zealand Government in the Institute of Environmental Scientific Research (ESR) as a principal scientist. (6T 4-17). At ESR, Dr. Buckleton helped to develop STRmix, but he does not own the software. (6T 46-47). He described himself as a civil servant who helped create the software. (6T 46-21). He has an undergraduate degree in chemistry, a master's degree in chemistry,

and a Ph.D. in chemistry. (6T 12-1). He has a certificate of proficiency in software engineering. (6T 15-1). Dr. Buckleton has been a member of the editorial board for Forensic Science International Genetics since 2007. (6T 151-1). He was published in this journal for the first time in 2007. (6T 15-4).

Dr. Buckleton began working with DNA in 1988. (6T 9-10). From 2014 to 2016, Dr. Buckleton worked at the National Institute of Standards and Technology (NIST). (7T 110-14). NIST is a federal government science organization which employs about 2,800 PhDs and has about 6,000 scientific staff. (7T 110-7).

STRmix is a PGS that assists with the DNA interpretation of complex mixtures. STRmix is not the only PGS on the market. (6T 17-7). There are two others in the United States and two more in Europe. STRmix is the most prevalent PGS in the United States. (6T 17-7 to 18). Human input is still critical before a mixture can run through STRmix, as an analyst needs to make determinations and assess if there is useful data on an EPG. (6T 54-13 to 22). STRmix will then take information from the EPG and generate an LR. (6T 64-2 to 13). STRmix will then take the electronic output, and propose different sets of genotypes, propose proportions, propose degradations, and visualize its results. (6T 64-13). After this visualization, STRmix will start at a random position in the profile, and it will run until it resolves at a suitable combination of genotypes for the mixture at hand. (6T 65).

STRmix is an established, version-controlled software on an annual release cycle, with several mid-year updates. (7T 11-12). Originally, STRmix was coded by Dr. Duncan Taylor, Dr. Buckleton's former colleague and close collaborator. (7T 27-11). After version 2.3, development of STRmix was transferred to an outside company, Orbit systems. (7T 27-23). All testing is conducted by ESR (7T 28-7). Proposed changes are put in a system called U-Track, which is

STRmix's primary disclosure mechanism for feature changes and other patch notes. (7T 11-16). Versions 2.5.11 and 2.8 are considered "mid-life versions," so each year their changes will be marked in U-track. (7T 11-22). Version 2.5.11 is the fortieth version of STRmix. (7T 147-3). Each time ESR prepares a new version of STRmix, they will test it. (7T 147-5). Testing is done by two different teams within ESR—one that tests the core scientific function of the system, and another that tests the graphical interface and other portions. (7T 147-7).

Dr. Buckleton also testified about probabilistic genotyping generally. (6T 15-24). He stated that STRmix was originally developed for the Australian and New Zealand government to use due to TrueAllele licenses being too expensive. (6T 18-12). TrueAllele is another PGS that predates STRmix. (6T 16-17). Dr. Buckleton mentioned that STRmix was not designed with commercial intentions, it was designed as a tool for Australia and New Zealand criminologists as part of Dr. Buckleton's government employment. (6T 49-13). STRmix became a commercial product after United States laboratories expressed an interest in it, circa 2014. (6T 49-50).

Dr. Buckleton discussed licensing sales for STRmix. (7T 124-18). He did not have firsthand knowledge of payments, so he was only able to discuss it in a limited capacity. (7T 124-18). STRmix pays 10 percent of their license sales to FSSA, which is a payment of about one million U.S. dollars annually. (7T 125-4 to 126-2).

Laboratories need to set up parameters before they can use STRmix. (6T 54-13). ESR will assist with the initial tuning of parameters to the laboratory, finalize them, and then check that they are working as intended. (6T 54-14). Once parameters are set, a laboratory needs to perform its own internal validation of the software. (6T 54-19). A laboratory cannot buy STRmix without being trained to use it, and it cannot be used in court without being validated. (6T 54-55). The first

training is a four-day training on the theory and operation of the software. (6T 55-11). Dr. Buckleton has conducted about 230 trainings for laboratory employees to help them become familiar with STRmix. (6T 105-12). Based upon his experience with training analysts, he has not seen any tendency on the part of laboratories to distort or exaggerate STRmix's capabilities; on the contrary, he has seen that laboratories want to know the limitations of STRmix and stay within them. (7T 59-23).

Dr. Buckleton highlighted that if an analyst is going to use STRmix, they need to know what the software does and how to operate it in a rational way without misusing it. (6T 56-3). User errors should not be seen as faults in the program. (6T 56-10). He said that operating STRmix incorrectly, or an analyst making an error when entering data, could produce counterintuitive results. (6T 56-15). One such result is a "wandering chain." (6T 56-17). When STRmix runs the MCMC process it runs 8 Markov Chains simultaneously, all of which should follow similar, but not necessarily identical, paths. (6T 56 to 57). Dr. Buckleton explained that "90 percent of the time" all 8 chains will produce similar results but the rest of time, one of them "just [goes] off" and produces a highly disparate result. (6T 66-4). STRmix will create a diagnostic number, known as the "Gelman-Rubin" diagnostic, which expresses how well the chains converged (6T 66-20). This number is not a hard or fast rule, but a data point to alert an analyst that a result may require closer scrutiny. (6T 67-2). As Dr. Buckleton puts it, the diagnostic is "arbitrary," but if it is "[a]round about 1.2 you, you might tolerate it," as that means there was a good convergence. (6T 67-3 to 67-9).

Dr. Buckleton testified that the software has never had false inclusionary support for samples with high LRs. (7T 24-5 to 74-22).

On direct and cross examination, Dr. Buckleton was asked about standards for software development. (6T 112-12); (7T 143-20). He identified a standard he believed to be relevant, The Institute of Electrical and Electronic Engineers Standards Association, IEEE 1012-2016: Standard for System, Software, and Hardware Verification and Validation (hereinafter "IEEE 1012"). He described IEEE 1012 a relevant but voluntary standard for evaluating the development of PGS such as STRmix. (6T 111 to 112). Dr. Buckleton testified that STRmix is not required to conform to IEEE 1012. (6T 112-12). IEEE 1012 discusses three kinds of independence required for the most stringent Verification and Validation ("V&V") standards: technical, managerial, and financial independence. (7T 144-145). For technical independence, reviewers and debuggers should be separate from the team that programs the software—there must be "fresh eyes" reviewing the code. (7T 144-23). Technical independence can be difficult for smaller groups to maintain. (7T 146-1). Managerial independence requires the team that writes the software and the team that reviews it to report to different managers; it is important because it insulates testing from external pressures on the business. (7T 144-10). Financial independence requires that the control of the testing budget be vested in an organization independent of the developmental organization, so that payment is not dependent on positive or negative results. (7T 145-19). Dr. Buckleton testified that while STRmix is not required to conform with these requirements, there is financial independence with STRmix due to the coders being tenured civil servants; their salaries are set by the government. (7T 29-6). He also testified that STRmix conforms with managerial independence due to the coders and testers being separate teams. (7T 28-18).

Dr. Buckleton further testified that the testing team maintains technical independence by working on their testing plan, attempting to verify STRmix operation and running mathematical calculations in parallel and without consulting the development team, until they reach a result they

cannot verify or fix. At that point, they document the bug and consult the development team. (7T 29-24). Dr. Buckleton will do an additional personal review of the code when necessary, and then another team will go back and check that his solution is sound. (7T 30-5).

Dr. Buckleton testified about the validation process for STRmix. (6T 103-15). For internal validation, first the system needs to be integrated with their other lab equipment to ensure it is running properly. (6T 79-12). Dr. Buckleton admits that ESR often assists with this step, and he does not believe this violates any standards for validations, because the validation itself will still be independent. (6T 80-9). For validations, each laboratory creates the samples used in their tests. (6T 80-21). The validation is being performed, in part, so that clients, courts, and juries can be confident that the necessary work has been done to show the laboratory can operate the equipment correctly. (6T 82-20 to 83-3). Accredited laboratories must go through internal validations before they are able to report their findings in court. (6T 55-5). Validation studies, and the format of their summary reports, generally follow the SWGDAM guidelines. (6T 70-14). Other laboratories might look to the ISFG guidelines. (6T 71-19 to 72-7). Dr. Buckleton testified that both guidelines are not standards, they are generally followed by laboratories but are not strictly required. (6T 71-6). At "most they read them and note the differences and then follow SWGDAM." (6T 72-24 to 73-1). He testified that a "standard is something you should" do, and a "guideline is something you could do," but in the United States "SWGDAM guidelines are the operative guidelines, so people follow them as if they're a standard." (6T 73-8 to 73-19).

Dr. Buckleton acknowledged that there have been criticisms of STRmix. (7T 10-11). In response to criticisms regarding documentation, Dr. Buckleton generated a requirements document for adversarial experts to refer to during code reviews. (7T 12-2). To create this document, Dr. Buckleton himself had to read and understand the source code. (7T 14-7). He mentioned that

STRmix has never changed the functionality of STRmix based on criticism but has made changes to their disclosure or information policies. (7T 15- 18).

Dr. Buckleton has never released a version of STRmix that he has found to be incomplete at the time of release but acknowledges that there have been "miscodes" reported by end-users. (7T 34-12). Dr. Buckleton states that failures from "miscodes," or bugs, can potentially vary from minor to drastic. (7T 50-7). The miscodes are all published on Dr. Buckleton's website. (7T 50-11 to 50-18). Miscodes get reported from members of the scientific community, such as the California Department of Justice. (7T 52- 9). A criminologist named Steven Myers has personally found three miscodes by independently performing, in Microsoft Excel, every calculation performed by STRmix while running individual samples. (7T 53- 6). The miscodes found thus far have been characterized as "minor," and many were "peripheral functionality" that did not impact calculation of the LR. (7T 54-12). Dr. Buckleton has no knowledge of the "miscodes" impacting a criminal proceeding. (7T 54-15).

Dr. Buckleton testified regarding publications on STRmix. (7T 63-6). He has personally written about coding standards and how these standards have been involved with STRmix, as well as authoring or co-authoring several articles on the science and validation of STRmix. (7T 63-10).

On cross-examination, Dr. Buckleton gave a summary of how EPG data is converted into a STRmix report. (7T 160-15). Data on the allele loci and peak heights will be input into STRmix. (7T 160-18). The analyst will also input an estimated NOC for each sample. (7T 160-4). This process will generate a deconvolution report, which gives a list of the potential genotypes at each loci that STRmix has determined. (7T 69-23). Essentially, STRmix is trying to come up with all plausible scenarios at each locus for the contributors. (7T 79-9). From these plausible scenarios, STRmix can then generate a series of LRs. (7T 86-24).

If needed, STRmix can take relatedness into account when generating LRs. (7T 87-20). In the instant case, Dr. Buckleton does not feel that relatives were a catastrophic issue or intractable problem. (7T 88-9). In the DNA samples in this case, only mixtures with two related contributors (dyads) are being proposed, and STRmix does not have issues with dyads. (7T 88- 11). If a laboratory wanted to test a dyad as a triad (three related contributors), they could do so, as an LR could be run with a third proposed relative. (7T 88 to 89).

Dr. Buckleton is of the opinion that most laboratories that use STRmix attempt to get the most conservative value possible. (7T 116-7). A lower LR is unambiguously conservative. (7T 116-23). STRmix's modeling decisions do not favor exclusion but does have three different processes that all favor reporting the lowest proposed LR. (7T 117-22). A statistic that favors exclusion will generally favor a criminal defendant. (7T 117-11).

The regression analysis of STRmix, the analysis of how code has been changed back through previous versions of STRmix, shows that the central element of STRmix has remained stable between versions of 2.3 and 2.9. (7T 91-21). The regression analysis is publicly available and explains the nature of the changes between different versions. (7T 92-4).

Last, Dr. Buckleton addressed the PCAST Report, the 2016 executive summary prepared for the President of the United States on the current state of forensic sciences. (7T 67-15; 95-2). He explained that, after PCAST stated that it was their opinion that PGS was promising but needed to do more to demonstrate its validity for analyzing mixtures with less available DNA at more extreme contribution ratios, he and several colleagues released a peer-reviewed study known as the "31 lab study." (7T 95-1 to 99-12). Their goal, Buckleton explained, was to directly respond to PCAST's requests for more publicly available and peer reviewed data on PGS, though to date PCAST has not indicated whether they are satisfied. (7T 101-1).

It is Dr. Buckleton's opinion that STRmix has addressed PCAST's concerns, is generally accepted in the scientific community, and has been peer reviewed. (7T 106-107). There is no doubt in Dr. Buckleton's mind that STRmix is fit for casework and fit to be used in laboratories across the world. (7T 113-114). These laboratories perform their own testing of the samples, they look for anomalies, look for stutter, drop-in, and drop-out, and will validate the system for their laboratory. (7T 114).

The court found Dr. John Buckleton to be a credible witness. Although Dr. Buckleton's recent experience with software development and engineering is limited to just one program, it just so happens to be the program at issue in this hearing. He may not have written code in 30 years, but his involvement in the continued development of STRmix and its enhancements assists this court with how STRmix works. Dr. Buckleton possesses a certificate in computer science, which is essentially equivalent to a minor in computer science. He was responsible for a large portion of the underlying mathematical model for STRmix. Both of those facts bolster his credibility. He possesses the requisite knowledge to discuss the software engineering principles, and to respond to the criticisms leveled against STRmix. Due to being a developer of STRmix he has special knowledge and experience in the software that assisted the court in understanding the intricacies of the development, use, and limitations of STRmix.

Dr. Buckleton thoroughly explained potential and known "miscodes" of STRmix. He testified that the discovered miscodes have not affected DNA evidence, because the miscodes/bugs that have been detected have had a minor effect on the numerical value of the LR. Per Dr. Buckleton, none of the miscodes have detected through source code review, but all have been identified through users using the program. Dr. Buckleton has never released a version of STRmix that he thought was not adequately tested at the time of release. Dr. Buckleton has published the list of

miscodes on his website, so the public can be informed about identified issues with the source code. Dr. Buckleton's candor about the miscodes assists the court with STRmix's limitations and how potential issues with the software can be handled. Dr. Buckleton's directness in discussing the miscodes showed he was a credible and reliable witness.

Dr. Buckleton's frankness on both direct and cross-examination and his willingness to acknowledge what he believed to be valid criticisms further boosted his credibility. His answers were not vague or qualified, nor did he attempt to change his answers when challenged. Dr. Buckleton knows STRmix can be improved, and he seems more interested in doing that than in deceiving the forensic community into accepting his work. He is willing to hold STRmix to a higher standard of review than is necessary for a software that is not a safety critical system. When asked by the court how STRmix can be improved, he said documentation of more analysis would be an improvement. The court found this honesty to be compelling. Dr. Buckleton is a staunch proponent of his own technology but also showed a commitment to scientific integrity. His opinions are entitled to significant weight.

v. State's Expert: Jennifer Thayer, Expert in the Fields of Forensic DNA Analysis and Forensic DNA Validations

Jennifer Thayer is employed by the New Jersey State Police, Office of Forensic Science, in their DNA laboratory. (8T 4-16). She joined their laboratory in 2003 and has been the laboratory director since 2022. (8T 5-3 to 4-20). Ms. Thayer has a bachelor's degree in biochemistry and a master's degree in forensic science. (8T 7-9).

In her current role as laboratory director, she is responsible for the DNA laboratory, where she oversees analyst casework and entries into the CODIS database. (8T 12-23). She is responsible for all day-to-day activities and the management of the staff. (8T 5-13). She is involved in all internal validations at the New Jersey State Police. (8T 11-24). She described these validations as including multiple scientists since "there definitely has to be more than one person because somebody is going to have to review the work that is done." (8T 12-11).

Throughout her career, she has attended educational trainings and seminars fairly frequently. (8T 13-4). She discussed how the FBI Quality Assurance Standards "requires a minimum of eight hours of DNA education per year," which she has always met. (8T 13-4). At the New Jersey State Police, the scientists will undergo training "in house;" these are conducted regularly. (8T 13-15).

Ms. Thayer explained how forensic DNA laboratories maintain accreditation. (T 18-12). New Jersey State Police is an accredited laboratory and is accredited by the ANSI National Accreditation Board ("ANAB"). (8T 18-22). She explained that the last complete audit by the ANAB was in 2022, but a surprise inspection occurred in October of 2024. (8T 19-1). Every four years the entire office system is audited. (8T 19-10). The New Jersey State Police DNA laboratory is in conformance with FBI Quality Assurance Standards and ISO 17025. (8T 19-16 to 19-21).

Ms. Thayer also described SWGDAM guidelines. (8T 20-10). She explained that a "guideline is something that is recommended that you adhere to," but a "[s]tandard is something that you must meet." (8T 20-3). SWGDAM will issue "various guidelines for different DNA testing and validation and they are also involved with the issuing of the QAS standards." (8T 20-14).

Ms. Thayer was asked to describe the basics of DNA. (8T 34-13). She explained that DNA is the building block of all life. (8T 34-16). For humans, more than 99 percent of our DNA is the

same as everyone else. (8T 34-18). DNA is what makes us human beings. (8T 34-20). Ms. Thayer explained that the <1% that *does* vary from person to person is what makes us unique individuals, and STRmix looks at that 1% to differentiate between contributors to a sample. (8T 34-22).

Ms. Thayer described how saturation affects DNA analysis. (8T 48-13). Saturation is when there is too much DNA to be visualized at the detection step of DNA analysis. (8T 63-5). The fluorescence that is emitted by the DNA can overwhelm the imaging sensor. (8T 63-11). When there is more DNA than the sensor can detect, the electropherogram will be a flat line. (8T 63-18).

In the New Jersey State Police laboratory, analysts will perform manual interpretation for single source samples above their stochastic threshold. (9T 67-68). If seven or more loci resolve they can report that for source attribution, if the RMP is sufficiently high. (9T 26-16 to 26-25). Once the RMP meets or exceeds 1 in 8 trillion, the laboratory's confidence that an individual is the source of a DNA sample is high enough that they will be added to their database for future identification. (9T 67-11). If a sample does not meet their source attribution threshold, it gets run through STRmix. (9T 68-1). There needs to be seven resolved loci for a sample to be run through STRmix. (9T 26-22). Typically, stutter will be eliminated for manual interpretation, but for STRmix, it can be left in. (8T 42-23). The variance between stutter and true peaks is inversely proportional; the higher the peak, the smaller the gap. (8T 56-7).

An analytical threshold is set by every forensic laboratory; any peaks below that threshold will not be labeled on the electropherogram. (8T 38-22). A stochastic threshold is also set during validation. (8T 38-39). Below this threshold, an analyst will see more stochastic fluctuations, including a higher potential for drop-out. (8T 39-1). If there is only one peak at a location below the stochastic threshold, the analyst will not label it as homozygous, because of the risk that a sister allele has dropped out. (8T 39-6).

Ms. Thayer explained the variables that must be put into STRmix. (8T 57-6). A laboratory must set the maximum allowable stutter. (8T 57-7). This means if there is a peak in a stutter position that is more than 20% of the adjoining peak, STRmix will not model it as stutter. (8T 57-8). If it is below that level, then STRmix will consider it as a potential stutter peak and may model it as stutter, allelic, or part stutter/part allelic. (8T 57-13). Ms. Thayer further discussed whether the stutter is going to be modeled per allele or per locus. (8T 57-16). For every locus there will be a regression file and that basically is an equation for a best fit line for the stutter. (8T 57-23). Typically, loci follow an allele regression line meaning you can chart the stutter per allele at a locus and often it will follow a nice, straight line where the larger peaks are giving higher stutter than the smaller peaks. (8T 57-58).

She then testified that after STRmix deconvolutes, an interpretation report is generated. (9T 70-10). This report has case information, diagnostics, weights of genotypes, and if a reference sample was also provided, then STRmix will produce a LR. (9T 70-14). A LR is the ratio of two probabilities to assign weight to evidence and it compares the probability of the evidence under one hypothesis as compared to an alternate hypothesis. (9T 70-14). The New Jersey State Police reports the actual ratio along with a verbal equivalent. (8T 75-5). For LRs below 1, they will calculate the actual ratio and use that as support for the alternate hypothesis as it's easier to understand. (8T 78-13). She further explained that the New Jersey State Police verbal scale puts a LR of .001 to 1000 as "uninformative" and do not assess a weight for inclusion or exclusion. (8T 92-23). In general, if a major contributor adds more to a mixture, that LR is largely unaffected considering minor "trace" contributors are affected more. (8T 102-11).

Ms. Thayer testified about the safeguards in place for New Jersey State Police's laboratory. (8T 16-10). Once all laboratory work is completed an analyst will analyze the raw data and then

interpret the DNA profiles. (8T 16-15). The analyst will compare any evidentiary profiles obtained to any reference samples that have been submitted for the case. (8T 16-17). The analyst drafts a DNA report, which will be technically reviewed. (8T 16-19). The technical reviewer will make sure that they agree with the conclusions and interpretations that are included in the report. (8T 16-23). If there's any discrepancies the analysts will discuss it and come to an agreement. (8T 17-2). Once the analysts agree that everything is accurately reflected in the report, the analysts and technical review will sign off on the report and it will go through an administrative review. (8T 17-3). This review is typically done by supervisors, but it can be done by anyone in the laboratory. (8T 17-6).

When the State Police decided to incorporate probabilistic genotyping, they knew they wanted to use a fully continuous model. (8T 25-7). Ms. Thayer testified that most laboratories use fully continuous models. (8T 25-12). A fully continuous model gathers information from peak heights and mixture ratios, whereas a semi-continuous model does not. (8T 24-25). STRmix was chosen due to its functionality, the support offered by ESR, and the fact that more laboratories within the United States were using STRmix. (8T 25-16).

For the validation of STRmix, there were two stages. (8T 46-11). First, the analyst established the parameters needed to be inputted into STRmix for STRmix to perform its biological modeling. (8T 46-11). Those parameters were based on samples that were run within the New Jersey State Police DNA laboratory. (8T 46-15). The estimated parameters for STRmix in the New Jersey DNA laboratory were listed in exhibit S-161A, which was shown to Ms. Thayer. (8T 47-13). The New Jersey State Police laboratory did not see enough drop-in events to be able determine their own drop-in parameters, so ESR recommended parameters based on other laboratories using the same kit and platform that the New Jersey State Police uses. (9T 38-4).

The second part of the validation was the internal validation study itself. (8T 46-15). STRmix uses biological modeling in its interpretation and to perform that modeling, it needed to know how samples behave in the laboratory. (8T 46-1). Ms. Thayer explained that determining how samples would behave in the New Jersey State Police laboratory was necessary. (8T 48-1). The first step of an internal validation was to run single-source samples through STRmix to determine the inherent variability of the Laboratory's existing equipment. (8T 46-9). Once individual variables were isolated and understood, the New Jersey State Police Laboratory progressed through each portion of the SWGDAM guidelines to conclude their validation study. (8T 71-17) After validating the software, the analysts took a refresher training. (8T 29-30).

Ms. Thayer mentioned that this was the first validation since she started working there that everything wasn't done completely by the analysts at her laboratory. (8T 32-8). Since this was the first time that STRmix had been validated in the New Jersey State Police Laboratory, she felt it important to document what they had done themselves and what ESR assisted them in doing. (8T 32-15). The internal validation only looked at related contributors very minimally. (9T 32-3). She mentioned that ESR told them they could further explore relatedness, but it was not determined to be necessary at that time. (9T 32-10). The New Jersey State Police DNA laboratory has now decided to do relatedness studies, but they are not complete. (9T 32-15).

In this case, the New Jersey State Police were given samples by the Monmouth County Prosecutor's Office. (9T 12-24). Ms. Thayer reported that, while some samples did not yield any informative results, there were conclusions drawn from a pair of jeans submitted. (9T 13-20).

On cross examination, Ms. Thayer was first asked about how has STRmix changed interpretation protocols. (9T 16-8). Before STRmix, interpretation and comparison had two different meanings. (9T 16-12). She explained that they "would interpret the profile first and then compare

it to any references submitted in the case." (9T 16-13). In manual interpretation, if an analyst could determine the NOC, they would then determine whether it was suitable for comparison. (9T 17-17). She agreed with defense counsel that this process of interpretation was a form of manual deconvolution. (9T 17-25).

Ms. Thayer then discussed thresholds and guidelines her laboratory has with different methods before STRmix was implemented. (9T 18-19). If there was a peak height ratio between a major and minor contributor, where it was less than eight to one, there would be additional protocols. (9T 18-6). The New Jersey State Police DNA laboratory had specific guidelines for two person mixtures with a major/minor contributor and different sets of guidelines for three person mixtures. (9T 18-19 to 18-22). To use these interpretation guidelines, there had to be 400 picograms of total input DNA for it to be compared. (9T 19-16 to 19-18). Below 400 picograms, she could not use these guidelines, and this was not discretionary. (9T 19-23). A major contributor also needed an average peak height of at least 400 RFU. (9T 20-6). Having at least 400 RFU would ensure that the major contributor is above the stochastic threshold. (9T 20-6). Her laboratory has since gotten rid of the minimum template threshold due to STRmix now being used. (9T 28-8).

Ms. Thayer summarized that the guidelines for STRmix are objective. (9T 20-21). She also mentioned that SOPs for comparison are objective. (9T 22-5). For a single source profile, all loci for which there is a result would have to match between the suggested contributor and reference sample in order for there to be an inclusion. (9T 22-10). If one locus does not match, there must be an exclusion. (9T 22-13).

She also noted that for mixed DNA profiles the standards are also objective. (9T 23-7). Most of the statistical calculations for these profiles were done in Excel worksheets. (9T 23-10). Determining the thresholds for when to use each method is important and SWGDAM guidelines

require DNA analysts like Ms. Thayer to complete the determination. (9T 23-16). Another hard threshold for this comparison, is there needs to be at least seven loci noted otherwise the sample is not suitable. (9T 27-9). She testified that this is a cutoff for the New Jersey State Police. (9T 27-9). Discretion still exists when an analyst is determining the NOC that can be reliably determined for a sample. (9T 27-18). This discretion is taught in training classes with several different samples. (9T 27-18). She admits that an analyst will potentially underestimate the NOC if the contributors are related. (9T 29-5). Discretion will come down to the training and experience of the analyst. (9T 39-12). She is of the opinion that "if you give the same STRmix report and electropherogram to our analysts, you're going to get the same answers." (9T 51-6).

Ms. Thayer testified that the New Jersey State Police followed SWGDAM guidelines when validating STRmix. (9T 23-23). This validation was performed by seven employees and took "[h]undreds, if not, thousands of hours" of work to complete. (9T 24-6). The data analysis for the validation was all done at the New Jersey State Police laboratory. (9T 24-14). She did acknowledge that ESR calculated the average peak heights for the sensitivity analysis. (9T 25-8).

The court found Jennifer Thayer to be a credible witness. Ms. Thayer has been working at the New Jersey State Police Office of Forensic Science for over twenty years. She has been the laboratory director since 2022. Ms. Thayer's testimony included her vast experience in analysis of DNA profiles, went through each step of DNA analysis, the laboratory analysis of STRmix, and SOPs on how validations are performed. Ms. Thayer was able to adequately and sufficiently explain how stutter filters are used with STRmix and how this affects the way a sample is run through the STRmix 2.8. Ms. Thayer's testimony provided the court a thorough understanding of how samples were interpreted using STRmix at the New Jersey State Police's laboratory, were highly relevant, and are entitled to great weight—particularly as they relate to internal validations.

vi. State's Expert: Christine Schlenker, Expert in the Field of Forensic DNA Analysis

Christine Schlenker is a forensic scientist with the New Jersey State Police Office of Forensic Science DNA laboratory. (9T 57-10). Ms. Schlenker received a bachelor's degree in forensic science from the University of New Haven and a master's degree in pharmaceutical science from the University of Florida. (9T 59-60). She is a member of the American Academy of Forensic Science, a member of the Northeastern Association of Forensic Scientist, and a past member of the New Jersey Association of Forensic Scientists. (9T 62-20).

When Schlenker started working at the New Jersey State Police laboratory, her primary duty was to perform case work. (9T 58-22). This included processing samples, writing reports, and testifying in court when needed. (9T 58-23). She has now testified approximately fifty times as a DNA expert. (9T 65-5). In her current role, she is more responsible for the technical review of case files and assisting with various validations. (9T 58-22 to 59-9).

Ms. Schlenker testified that after the New Jersey State Police decided they would be validating STRmix v. 2.8.0, her laboratory attended a four-day training in May of 2021. (9T 63-12). This training enables her to perform technical reviews of the case files. (9T 64-10). Ms. Schlenker has done write-ups for several dozen cases and, sometimes she examines cold cases using STRmix. (9T 64-10). Primarily, she does technical reviews and administrative reviews to see how STRmix is running. (9T 64-16). She has done over a hundred STRmix reports and altogether has analyzed over 40,000 DNA samples. (9T 64-21 to 65-2).

She described that in the New Jersey State Police DNA laboratory, they perform manual interpretation for single source samples above their stochastic threshold. (9T 66-25 to 67-2). If

seven or more loci resolve, they can analyze that sample for source attribution. (9T 67-7). Once the statistic meets or exceeds 1 in 8 trillion, that New Jersey State Police is confident that somebody is identified as the source. (9T 67-11). She summed this by stating that if a sample contains a major contributor with seven complete loci, then the analyst can manually compare the profile to a reference sample. (9T 67-22). If a sample does not meet these requirements, it gets run through STRmix and an LR is reported. (9T 68-1).

For a sample to be run through STRmix at the NJSP laboratory there are certain criteria that must be met. (9T 68-11). First, the laboratory needs to be able to determine the NOC, because they cannot run a mixture with more than four contributors. (9T 68-11). There must also be information present in at least seven loci; these are hard limits for the NJSP. (9T 68-11).

Before STRmix, Ms. Schlenker was not able to interpret mixtures with more than three contributors. (9T 108-1). She explained that if the laboratory received a mixture with more than three contributors, they would not draw any conclusions from it (9T 108-7). Now with STRmix, she can look to see if the mixture has four contributors or more than four contributors. (9T 108-11).

The court found Christine Schlenker to be a reliable and credible witness. She has been working at the New Jersey State Police, Office of Forensic Scientists since 2002. Ms. Schlenker was able to explain the process of her laboratory using STRmix and the limitations her laboratory has for it. She explained the criteria necessary for a DNA sample to be run through STRmix at the State Police laboratory and how she performs technical reviews of case files for STRmix. Her testimony noted how the State Police ultimately determined STRmix was fit for case work due to the validation study performed. An FBI Quality Assurance Standards ("QAS") audit was performed on STRmix after it was first validated, and she testified that it did not uncover any issues

with the NJSP laboratory using the software. STRmix has allowed Ms. Schlenker to be able interpret more complex mixtures and overall, her laboratory has been satisfied with the work product that has been outputted by STRmix.

On cross examination, Schlenker was questioned about limitations her laboratory has, and the discretion given to an analyst in deciding whether to utilize STRmix. She thoroughly explained the SOPs her laboratory has and how that impacts the interpretation process. Her experience with STRmix was able to provide the court with a greater understanding of how the State Police examined the samples in this case. She was direct with her answers when discussing how STRmix has changed the way her laboratory has approached case work. Ms. Schlenker was able to respond sufficiently to questions, has a strong educational background, and has worked extensively with STRmix in a professional capacity. Her opinions on how STRmix works and her analysis of the samples provided in this case are given great weight.

vii. State's Expert: Dr. Michael Coble, Expert in the Fields of Forensic DNA Analysis, Probabilistic Genotyping, and DNA Mixture Interpretation

Dr. Michael Coble is executive director for the Center for Human Identification at the University of North Texas. As executive director, he oversees the laboratory but does not personally perform any testing. His career has been primarily research focused. He is also a professor at the University of North Texas in the microbiology, immunology, and genetics department. His academic focus is molecular genetics. (10T 4-21 to 6-9). He received his PhD in human genetics in 2004. After completing his PhD, he was employed by the National Institute of Science and Tech-

nology as a forensic biologist. In this role, he conducted research on new standards and technologies. In 2010, there were no probabilistic genotyping systems available, so interpretations were done manually. (10T 15-10 to 17-2).

Dr. Coble testified as an expert witness for STRmix in United States v. Gissantaner, 990 F.3d 457 (6th Cir. 2021), reh'h en banc denied. (10T 34-22). In this case, Dr. Coble was one of two experts appointed by the district court to serve as an independent expert and issued a report and recommendation finding that STRmix satisfied the Daubert factors. (10T 37-14). The district court disagreed with his recommendation, but the 6th circuit ultimately found STRmix reliable and overturned the district court's decision. (10T 38-4).

Dr. Coble discussed the incremental changes to DNA analysis since the late 1980s, and the development of new methods, procedures, and testing. (10T 39-24 to 40-3). Dr. Coble explained became involved with probabilistic genotyping software when he wrote a procurement proposal for NIST in preparation for purchasing TrueAllele. (10T 20-15). NIST performed a limited internal validation for TrueAllele but focused on the confines of a research laboratory—they did not validate it for casework. (10T 23-25 to 24-2). In 2010, SWGDAM promulgated guidelines for interpreting Short Tandem Repeats, the fragments of DNA used in electrophoresis, and recommended that laboratories employ a stochastic threshold for interpreting data. (10T 21-3 to 21-21). Per Dr. Coble, "the real advantage of probabilistic genotyping is that we no longer need that stochastic threshold," it can be done away with and the "software will evaluate the peak probabilistically." (10T 23-1).

Dr. Coble testified about his experience with developmental and internal validations. (10T 24-21). He was personally involved with the internal validation for TrueAllele and believes that validations are very important. (10T 24-25). He explained his experience with other probabilistic

software and how there are two models of probabilistic software, semi-continuous and fully continuous. (10T 25-9). Per Dr. Coble, in a semi-continuous system, "the software doesn't use the peak height information," it "simply asks what peaks are there" and "what alleles are present." (10T 25-10). Fully continuous models, like TrueAllele and STRmix, are much more efficient because they can analyze more data. (10T 25-18). He published a comparative publication about these two approaches. (10T 53-10). The main difference he observed was that "the fully continuous method, STRmix, was able to provide more information about the mixture than the semi-continuous or the manual approaches." (10T 53-19).

Dr. Coble mentioned that he is a member of the International Society for Forensic Genetics and has collaborated on guidelines for the developmental validation of PGS platforms. (10T 32-3 to 32-7). Dr. Coble believed at the time that "probabilistic genotyping [was] really taking off [in Europe] and it would be really helpful if we could write something" to help guide its adoption. (10T 62-21).

In the United States, most laboratories will follow SWGDAM guidelines when validating probabilistic genotyping software, rather than ISFG guidelines. (10T 64-10). Dr. Coble also served on the scientific committee that wrote the standards that would eventually become ANSI/ASB Standard 018. (10T 64-19 to 66-6).

Dr. Coble has published 101 peer-reviewed articles. (10T 28-20). He has written with Dr. Bright, Dr. Buckleton, and Dr. Taylor. (10T 52-25 to 53-6). Dr. Coble explained that part of the process for peer-reviewed publications is subjecting your work to a blind review. (10T 27-3). In the blind review process, the identities of the reviewer and the author are both unknown to the other (10T 27-13). This allows for anonymous and critical evaluation. (10T 27-13).

Dr. Coble testified that STRmix has been subject to peer review. (10T 81-16 to 83-5). There have been publications that discuss the theory, use cases, and software algorithms of STRmix. (10T 83-5 to 83-13). Most publications have a developer of STRmix attached to the paper, but there are papers which were written independent from STRmix developers. (10T 82-83). In 2023, there were about 120 articles in peer reviewed journals that focused on STRmix. (10T 82-2 to 83-2).

Dr. Coble has been familiar with STRmix since he was invited to view an early version, which he believes may have been v. 1.07. (10T 45-23). He testified that he was "[v]ery much" impressed with the software: he brought his own samples to run through STRmix and was impressed at its speed and accuracy. (10T 47-8). After a week of evaluations, Dr. Coble felt confident in STRmix, and NIST and the United States Army Laboratory both purchased the software. (10T 47-23 to 48-15). Laboratories in the United States have been using STRmix since 2014. (10T 49-14). By 2017, there were thirty-seven laboratories using STRmix, and now about eighty-nine laboratory systems are using it. (10T 69-23).

Dr. Coble is familiar with the 2016 PCAST report. (10T 66-13). The PCAST report mentioned probabilistic genotyping and Dr. Coble said it "had a very favorable view of probabilistic genotyping, that it has a lot of promise," and that it allowed for laboratories to interpret mixtures they previously could not. (10T 67-7 to 67-18). The 2016 report made a tentative recommendation to limit testing to three-person samples with a minor contributor of at least twenty percent, until more data could be analyzed. (10T 67-14).

Dr. Coble testified that the report authored by Dr. Bright (Jo-Anne Bright et al, Internal Validation of STRmix – a multi-laboratory response to PCAST, 34 Forensic Science International:

Genetics 11 (2018) (hereinafter "31 Labs")) was, as the title implies, a response to that recommendation. (10T 68-9). The 31 Labs report demonstrated that a three-person mixture at twenty percent is "not the base," laboratories can go lower than that. (10T 68-13). Dr. Buckleton and Dr. Taylor collaborated on this publication. (10T 69-1). Dr. Coble found the 31 Labs report to be accurate. (10T 69-18). Others in his field did criticize the fact that some of STRmix's developers co-authored the study. (10T 70-22). Dr. Coble does not give weight to these critiques because the laboratories who are using STRmix are performing casework for clients; they do not have the same time to commit to publication as researchers. (10T 71-7 to 71-12). If we wait for these laboratories to publish, the topic would no longer be "novel," so it would be the kind of paper that is typically rejected by journals. (10T 71-19). There is also transparency into the fact that STRmix developers participated in the study, and the reader can take this into account. (11T 39-16). There are also papers and publications that have been independent of the STRmix developers. (10T 84-3).

On cross examination, Dr. Coble was questioned further about the publications he has made for STRmix and who he has collaborated with. (11T 56-25). The first article he published with Dr. Buckleton was in 2014 and their most recent publication together was in 2022. (11T 57-15 to 57-21). Both Dr. Coble and Dr. Buckleton are members of the editorial board for the Journal of Forensic Sciences and for Forensic Science International: Genetics. (11T 58-2 to 58-18). Dr. Coble joined the board for FSI Genetics in 2014, and since then he has published around forty articles in FSI Genetics. (11T 58-14). Dr. Buckleton was already a member of the board when Dr. Coble joined. (11T 58-14).

For Dr. Coble's report in this case, he looked at materials from both the New Jersey State Police DNA laboratory and Bode Technologies. (10T 77-6). His review was informed by the Daubert factors, as he understood them. (10T 74-20). He is familiar with these factors as he has

addressed them in the past. (10T 75-4). He found STRmix to have been adequately tested as it has been tested by both researchers and forensic laboratories. (10T 78-10). He does not "have any doubt that STRmix has been tested" due to each laboratory conducting their own independent internal validation. (10T 80-5). He also agrees that STRmix has been subject to peer review. (10T 83-5).

Dr. Coble also testified about error rates, which is another Daubert factor. (10T 85-17). Errors could be due to low information in a mixture or could present as an unintuitive answer that the analyst does not expect. (10T 85-25 to 86-16). Dr. Coble considers it to be error if STRmix gives a LR of less than one for a single-source sample where the analyst expects a highly inclusionary LR. (10T 86-17). He has seen samples where this has happened. (10T 86-24).

A laboratory wants to avoid an inclusion of the wrong person at a very high LR. (10T 88-3-9). False inclusions of this nature are a catastrophic error. (10T 88-2). Dr. Coble has not seen such an error before, but if this error did occur the analyst should be able to figure it out by looking at the diagnostic reports that STRmix generates and determining whether there is something that does not look right. (10T 88-89).

An analyst needs to be able to look at an EPG and determine the NOC. (10T 91-20). This is a matter of analyst discretion and has been since before STRmix. (10T 91-25). An analyst will not know the total picograms in a sample or the ratio of contributors by looking at an EPG, that information comes after the sample has been run through STRmix. (10T 92-18).

Dr. Coble noted that the best way to identify code errors is to test the software. (10T 94-13). Dr. Coble does not know of any source code errors that have been discovered just by looking at the source code for STRmix. (10T 95-11). The ISFG guidelines Dr. Coble helped write address source code review. (10T 95 to 96). The original intent of the ISFG guidelines was to advocate for

transparency with regards to source code. (10T 95-21). The ISFG guidelines also recommend ensuring future availability of source code in the event a developer passes away or the code is needed for later review. (10T 96-14).

A laboratory not including data from a particular locus is not a software error and is not necessarily a user error. (11T 16-23). If a laboratory cannot discern usable DNA from stutter, then the decision should be to avoid the locus. (11T 16-23 to 17-2). Dr. Coble testified that a false inclusion does not necessarily mean an error has occurred. (11T 23-24). Having low levels of DNA means there is a higher risk of false inclusions, but this is the software working as expected. (11T 23-24 to 34-5). Dr. Coble has seen STRmix falsely include people and has seen analysts falsely exclude someone. (11T 37-38). To prevent false exclusions, if an LR of 0 is generated, an analyst should stop and determine why an LR of 0 has been generated before proceeding further. (11T 38-1). ESR provides initial training that reinforces this point and demonstrates the diagnostics available in STRmix should this occur. (11T 38-1). An analyst should determine how and why STRmix created an LR of zero before moving on. (11T 38-9). For this case, Dr. Coble reviewed the reports and diagnostics generated by Bode Technologies (10T 101-24; 11T 23-10). He did not review them as in-depth as he would have if he was performing a technical review. (11T 52-19 to 52-25). He did note that he observed a mutation in some samples consistent with the mutation [redacted] Caniero possesses; this mutation "effectively anchors her." (11T 52-20). No one else in the family can be compared to that reference. (11T 52-20). From the reports, Dr. Coble did not see any single locus exclusions. (11T 53-8). Based on his review of the samples in this case and his experience with PGS, Dr. Coble disagrees with the proposition that anything below five percent is unreliable, and believes the software can be reliable all the way down to a 0% contribution. (11T 55-7).

It is expected for there to be allele sharing between family members. (11T 8-25 to 9-2). Dr. Coble mentioned that it is possible for an internal validation study to test how STRmix acts when there are related contributors, but his assessment is that the LRs STRmix is producing for mixtures where siblings are present is more conservative when it does not take kinship into consideration. (11T 41-21 to 44-4).

Dr. Coble testified regarding the expert report filed by Dr. Karl Reich. In this report, Dr. Reich mentioned a degradation issue with locus D2S441. (11T 13-8). Dr. Reich considered this issue to be a "glitch," discussed infra, but Dr. Coble stated that it was a product of biology. (11T 12-25 to 16-3). Dr. Coble explained that the "glitch" happens when a degraded sample has stutter and loses DNA fragments, making it is difficult to see the separation of alleles, (11T 15-18 to 16-2). This is not, in Dr. Coble's opinion, an error; the decision to ignore that locus and not rely on it for analysis eliminated any issues. (11T 16-23 to 17-2).

On cross examination, Dr. Coble was asked about quantitation. (11T 19-4). At low levels there can be uncertainty about quantitation values for error rates. (11T 19-11). There is a variation around what number will be quantitated, so a quantitation giving a zero does not mean a full profile cannot develop a full profile for a sample. (11T 20-12).

Dr. Coble was also asked about the phrase "ground truth." (11T 32-10). This is a common phrase used in the DNA Analysis field. (11T 32-15). In relation to validation type samples, ground truth means that the analyst knows what is being put into the mixtures because they have the profiles. (11T 32-18). An analyst will never know the ground truth. (11T 32-25).

Dr. Coble would recommend probabilistic genotyping. (10T 82-8). As he frequently says there "are labs that are using probabilistic genotyping and there are labs that will soon be using probabilistic genotyping." (10T 82-10). It is his opinion that "STRmix is generally accepted in the

United States" and that it is "the most prevalent software in use." (10T 99-1). He reviewed Bode's validation summary, and nothing gave him pause, as it looked like every other validation study, he has seen for STRmix. (10T 102-8).

Dr. Michael Coble was a credible witness in the field of forensic DNA analysis, probabilistic genotyping, and DNA mixture interpretation. Dr. Coble thoroughly answered questions on both direct and cross examination. Dr. Coble has performed over a hundred peer reviews in DNA testing generally and has done peer reviews for ISF Genetics. Dr. Coble's first experience with STRmix was in 2014 when he was invited to New Zealand for a weeklong course on it. He has used STRmix to validate samples, has written favorable comparative studies for STRmix and has written a report for the error rates it has.

Dr. Coble thoroughly testified about source code errors and false inclusions with STRmix. He helped draft the IDFG guidelines which evaluated source code review, and Dr. Coble was unaware of any source code errors with STRmix. Dr. Coble testified that false inclusions with STRmix are possible, so an analyst needs to check the LR before excluding an individual.

The court did not find Dr. Coble to have any bias or motive to testify favorably about STRmix. He is a professional in the DNA analysis community who does not have any incentive in the outcome of this hearing. He has never worked for STRmix or been involved in developing the software. He has utilized other probabilistic genotyping software besides STRmix, so he is able to provide objective explanations on the reliability of STRmix in case work. Dr. Coble gave comprehensive answers, was trustworthy in his knowledge of STRmix, and clearly explained all the advantages and disadvantages of STRmix. While Dr. Coble was an early proponent of STRmix, he is not involved in its development and his opinions were therefore found to be objective and worthy of great weight.

viii. Defendant's Expert: Dr. Karl Reich, Expert in the Field of Forensic DNA Analysis

Dr. Karl Reich works in a commercial non-governmental forensic DNA laboratory located in Chicago. (12T 8-4). He has been employed there since 2002 and currently serves as the managing partner and laboratory director. (12T 8-8-13). Dr. Reich received his undergraduate degree in chemistry from Cornell University, his graduate degree from University of California, Los Angeles, and his postdoctoral training at Stanford University. (12T 8-16). He has a background in mathematics. (12T 13-21). Dr. Reich is not a member of any professional organizations because he prefers to be independent. (12T 11-9). Dr. Reich has testified as an expert witness over 170 times in forensic DNA, forensics biology, and forensic statistics. (12T 14-5 to 14-9). He has testified in criminal cases and civil cases for state and federal courts. (12T 14-12).

In his current role as managing partner and laboratory director, Dr. Reich is responsible for administrative tasks and for the science output of the laboratory. (12T 9-15). His laboratory does a variety of work in forensic DNA like testing and manufacturing products for the forensic DNA market. (12T 9-20). Dr. Reich published papers in peer reviewed scientific journals. (12T 13-3). He has published in different journals and most of the papers are on products his laboratory has developed. (12T 13-3).

Dr. Reich testified about his experience with internal validation studies. (12T 10-2). Dr. Reich's laboratory is accredited through three different auditing agencies. (12T 10-2). Each accreditation has its own requirements for validations, but they are similar. (12T 10-3). They all require a series of documentation protocols and SOPs that the laboratory must follow to be accredited. (12T 10-9). Dr. Reich believes that he has completed roughly twenty-five internal validation

studies. (12T 10-19). Each laboratory performs their own validation and does not reference other laboratory's results, validations, or guidelines. (13T 15-3). At Dr. Reich's laboratory, they perform manual interpretation and communicate their results as an RMP. (13T 10 & 22).

Throughout Dr. Reich's career he has analyzed close to a thousand DNA samples. (12T 14-15). At first, Dr. Reich did each step of the DNA analysis himself, but as his team has grown he has delegated much of that responsibility. (12T 10-23 to 11-6). His laboratory deconvolutes mixtures manually, when possible. (13T 9-15).

Dr. Reich testified about DNA generally and gave insight into some of the challenges of the four-step process described supra. One of the challenges in extraction is that an analyst cannot extract the same sample from a piece of evidence every time. (12T 56-57). It could be similar, but never the same. (12T 57-1). Once DNA is extracted, it must be purified before it can be quantified; water, chemicals, and debris from the extraction process must all be removed; this also changes the sample. (12T 30-12).

Dr. Reich testified that the amplification process is not perfect, due to there being noise, or artifacts, that are inherent to the process. (12T 32-4). The most common and concerning noise for Dr. Reich is stutter. (12T 32-11). In addition to back-stutter, where a fragment of an STR chain is lost causing the allele to show at an earlier position on the locus, there is upstream stutter: a chain can pick up a fragment and show seven repeats when only six should be there. (12T 33-7). If you overwhelm the PCR amplification, secondary stutter can occur. (12T 33-21). If another fragment breaks off an already fragmented allele, it can be off by two positions. (12T 33-21). There would need to be a lot of DNA for secondary stutter to be seen. (12T 33-21). These are known stutters that can be identified and quantified but not avoided entirely. Dr. Reich testified that stutter cannot be completely factored out during the amplification process. (12T 34-5). Stutter is a challenge with

DNA mixtures due to it being possible to misinterpret a second contributor's peak as mere stutter to a majority contributor's peak. (12T 45-46). Variation is expected in the amplification step when there are small amounts of DNA. (12T 59-18).

Dr. Reich finalized his testimony about the steps needed for forensic DNA analysis by explaining the electrophoresis portion of the process. (12T 35-6). At this step, an amplified fragment will go into the capillary, and it will be separated. (12T 36-16). There is a detection window in the capillary that allows for data to be recorded. (12T 37-22). An analyst could take the same sample, feed it through the capillary electrophoresis instrument and get different peak heights each time, but they should not get a different set of alleles. (12T 55-10).

Sensitivity, Dr. Reich explained, is how little DNA material a laboratory can analyze and still get results. (12T 40-41). Sensitivity has improved over the years thanks to better equipment and extraction kits. (12T 38-25 to 39-7). Forensic technology is now sensitive enough to gather DNA from the mere touch of a surface. (12T 40-21). Breathing alone can deposit DNA on a surface. (12T 41-3). This increased sensitivity does not mean mixtures have become easier to interpret, however. (12T 41-21).

Mixed samples are difficult to interpret. (12T 46-13). One difficulty is when there are individuals who share chromosomes due to there being less variety in the DNA profiles. (12T 46-17 to 47-3). Children will only have four possible choices of alleles, two each from their mother and father. (12T 46-22). It will be hard to differentiate between sibling profiles due to allele sharing and allele overlap on the EPG. (12T 47-4). But generally, on EPGs, if there are four peaks, there must be a minimum NOC of two, since each contributor can only provide two peaks. (12T 48-24). Even random individuals will have some overlap, as humans DNA is roughly 98% identical. (12T 78-6). If related individuals are a part of a sample, there would be a general increase in inclusionary

LRs. (12T 79-25). Dr. Reich testified that failing to study relatedness could impact the reliability of the software. (12T 79-16).

Dr. Reich also testified that the ratio of contributors can make a mixture more difficult to interpret. (12T 50-12). Extreme ratio proportions are easier for an analyst to interpret, but similar proportions can complicate how an analyst deconvolutes a sample or how they assign different contributors to a sample. (12T 50-23 to 51-1). Also, when there are increasingly disparate peak heights between contributors the analyst can start to lose information from the minor contributor and can mistake it for noise. (12T 52-7).

Severe differences in the amount of DNA material between contributors can influence a hypothesis as to the NOC. (12T 56-57). Not having enough DNA can also lead to there being more variation in LRs. (12T 57-15). There needs to be enough copies of each locus in the original sample to get a consistent result otherwise the analysts will get different outputs. (12T 57-22).

Dr. Reich was asked about his criticisms of STRmix. (12T 52-25). He believes that STRmix will over interpret data and over interpret results. (12T 53-2). He also explained his criticism of the SWGDAM verbal scale. (12T 60-12). Dr. Reich believes verbal scales should be abolished as they do not provide sufficient context for actual numbers. (12T 60-14). In his opinion, the scale is biased towards inclusion and against exclusion. (12T 60-12). They are biased towards inclusion because if numbers increase, the analyst is provided a scale, but if the number is below one, they are not given the same scale. (12T 61-3). An LR is a number which should be interpreted next to other numbers and not next to words. (12T 62-16). He also believes that the thresholds used for the verbal scale are meaningless as each laboratory chooses its own words. (12T 61-18).

Dr. Reich testified about the materials he reviewed for this case. (12T 15-19). He requested all the laboratory reports, the laboratory case file, the electronic data, the SOPs, and any phone and

email records. (12T 15-19 to 16-4). Dr. Reich was also provided the validation documents and the validation summaries for the versions of STRmix used by the respective laboratories. (12T 16-9).

Dr. Reich testified about his review of the internal validation document for Bode Technologies. (12T 66-4). He saw there to be variation between instruments and that individual instruments were not maintained in the same way which led to variability in peak heights for the internal validation. (12T 65-22). So, individuals with the same amount of DNA being inputted, were generating different peak heights. (12T 66-11).

Dr. Reich testified about internal validations broadly before focusing on Bode Technologies and the New Jersey State Police. (12T 66-20). Laboratories do internal validations because they want to be as thorough as possible by knowing more about the method, reagent, or commercial kit they are using. (12T 66-20 to 67-2). In this case, Dr. Reich found many similarities between the two studies in question. (12T 68-18). To some extent, Dr. Reich expected the studies to be similar, but felt the graphs and some of the standard language in both were too close. (12T 68-20 to 69-4). Dr. Reich said this could imply that someone outside the laboratory is doing some of the work for the validation and the laboratory is not using its own tools to validate. (12T 69-6). He was concerned about this, because it speaks to the independence of the validation and strength of the tests used on the method. (12T 69-13).

Bode Technologies created thresholds to determine when a mixture can and cannot be run through STRmix. (12T 91). One of these thresholds, is that the contributor must be at least 5% of the mixture, with the contributor being at least close to 20 picograms. (12T 91). Dr. Reich said there is uncertainty in these measurements because an analyst can't tell if they are at, below, or

just above the thresholds due to there being inherent variation in DNA quantification and in composing mixtures. (12T 91-13). To be reliable, Bode should stay within these thresholds, which is difficult, due to the variable nature of DNA extraction. (12T 92).

Dr. Reich was asked about low template amounts being used by the New Jersey State Police DNA laboratory. (12T 82-20). He is of the opinion that the New Jersey State Police cannot reliably use STRmix at low level template amounts. (12T 82-20). There is not enough data in low amounts of DNA for a mixture to provide enough of a DNA profile for sufficient associations to be drawn. (12T 83-22).

Dr. Reich was concerned about how juries would interpret LR numbers for STRmix, and PGS in general. (12T 94-9). Since very high LRs or RMPs can run into the sextillions range, a jury hearing a LR of 43,000 may overinflate what this number means for DNA. (13T 109-110). Forty-three thousand "[is] a big number for a bank account. It's a big number for the number of books in the library. It is not a big number when it comes to DNA[.]" (12T 94-19). Dr. Reich was more comfortable with a jury hearing LRs if they were contextualized by describing the completeness of the profile they were linked to. (12T 95-20). Dr. Reich does not have any issues with LRs being used in forensic science generally, as they are also used for paternity testing and other non-criminal applications. (13T 3-21).

Regarding the "glitch" observed in one of Bode's samples, discussed supra, Dr. Reich did not agree with Dr. Coble's assessment. He felt that an unresolved allele was indicative of equipment in need of repair and required an analyst to stop, investigate, and refuse to move forward until it was resolved. (12T 105-3 to 106-13).

Dr. Reich has never run STRmix before or validated STRmix but would like to conduct some experiments of his own. (12T 108-14). He is not convinced that he would want to use it or

testify under oath to its results. (12T 108-14). Dr. Reich does not believe STRmix is useful for robust DNA profiles, because laboratories "can interpret those using the existing methods that have been in place for 20 years." (12T 108-21).

On cross examination, Dr. Reich was asked about more of his overall views of STRmix. (13T 7-13). He agreed that STRmix has been peer reviewed, that it has been published extensively, and there has been a lot of instances of testing. (13T 7-15 to 8-3). Dr. Reich felt more testing should be required for a laboratory to use STRmix. (13T 7-25 to 8-3).

Dr. Reich discussed the EPG step of DNA analysis. (13T 19-4). Dr. Reich values EPGs and its worth in DNA analysis. (13T 19-4). Analysts are trained to evaluate EPGs to be able to understand it. (13T 19-9). Probabilistic genotyping does not take this human step out of mixture deconvolution. (13T 19-14). There will always be an analyst who needs to write the code, who needs to design the algorithm, so there is no way for this process to not have any human intervention. (13T 19-21 to 20-1). An analyst must evaluate an EPG, decide if enough data exists for STRmix to be used, and then decide what data from an EPG they are going to input. (13T 20-2 to 20-11).

Regarding the New Jersey State Police validation study, Dr. Reich thought that it was more complete than other validation studies and that their visualizations were clearer than Bode Technologies. (13T 16-2). Dr. Reich found this study to be sufficient regarding major contributors, but had concerns that the study did not fully consider STRmix's effectiveness for minor contributors. (13T 95-10 to 95-18).

For the Bode validation study, Dr. Reich found two issues. (13T 98-20). One, he felt it was sloppy, discussed supra. (13T 98-20; 13T 106-14). Dr. Reich identified it as a proofreading concern. (13T 100-6). Second, Bode's report was close to Bode's limitations for testing. (13T 99-4).

Due to the numerical values of the quantitation, DNA amounts vary, and Reich felt it was better to exercise caution at the edges. (13T 99-4).

The court found Dr. Reich extremely intelligent, but his credibility was undermined by his inconsistencies between his claims that validation studies of STRmix are insufficient and using STRmix, through Bode Technologies, to assist clients. Reich suggested that a client's wish may not be aligned with his thoughts regarding STRmix. This is unpersuasive. If Reich trusts Bode enough to recommend them to clients, then his criticisms here fall flat.

Dr. Reich took great care and time in ensuring that he was understanding the questions being asked for him. The court appreciated that Dr. Reich asked counsel to clarify their question if he was confused and then provided an answer once he knew what was being asked of him. Dr. Reich is credible, but his willingness to refer clients to other laboratories for STRmix analysis prevents the court from giving significant weight to his criticisms.

ix. Defendant's Expert: Dr. Mats Heimdahl, Expert in the Field of Software Engineering

Dr. Mats Heimdahl holds a PhD in Computer Science from University of California, Irvine. (14T 5-24). His PhD dissertation was on software engineering with a focus on safety critical systems. (14T 6-18). He is currently employed as the head of the Computer Science department at the University of Minnesota. At the University of Minnesota, he currently serves in a management role overseeing the research schedule but is still involved in his own research, as well as teaching. His teaching has been primarily in software engineering, so his classes typically relate to requirements, modeling, and software testing and verification. (14T 5-24 to 7-8).

Heimdahl has studied, researched, and taught in the software engineering field for over thirty years, with a special focus on software requirements. Dr. Heimdahl has studied software requirements for air traffic control systems, consulted for Boeing, and also consulted for medical device companies. Much of this work is proprietary and thus is not peer-reviewed, but Heimdahl has been published in some journals. (14T 9-10 to 11-14). Computer science is a "fast moving" field, so published articles are generally brief and presented at conferences rather than more formally reviewed. (14T 11-5).

Most of his work has been published in the Requirements Engineering Journal. Dr. Heimdahl is also a member of IEEE, ACM, Association for Computing Machinery, which are both professional organizations. (14T 14-9).

Software Engineering, according to Heimdahl, is about the "big picture." (14T 19-1). It involves surrounding activities that are not directly related to programming. (14T 19-9). While programming (the act of writing code) and design (the act of deciding what you want code to do) are both encompassed within software engineering, software engineering is a higher-level activity that revolves around generating requirements, documenting activity, and understanding the wants and needs of a customer or stakeholder in one's software. (14T 18-1 to 21-17).

Dr. Heimdahl believes that one such stakeholder in PGS is a person of interest whose DNA is being compared to a sample, and that a PGS cannot properly have its requirements laid out without consulting someone who has their interests in mind. (21-17). He also believes that the courts are a stakeholder, and software engineering requires their input during initial stages in the engineering process. (14T 22-3).

Dr. Heimdahl was critical of STRmix's design and implementation, and he opined that it did not meet the software engineering community's standards for a safety critical system. He based

this opinion off his review of STRmix's requirements documentation, which he found to be insufficient—due, in part, to the fact that it was written after STRmix's development and described the functions of STRmix, rather than STRmix being written to fulfill the requirements contained in the document. (14T 43-23 to 51-23). He characterized STRmix as "not ready for prime time." (14T 51-21)

Much of Dr. Heimdahl's criticisms of STRmix take on a similar tone: his principal objection to the software is that it was not engineered according to what he believed to be the best practices, and steps were taken out of order. However, his analysis of the software ended there. His dismissal of STRmix's reliability was superficial and conclusory and summed up to the assertion that, since STRmix was not built according to what he considered to be the best practices, there *must* be errors within the program and there was no need to look any further. He characterized a piece of software built without a requirements document as the developer telling the end user "trust me, I tried it and it works," (14T 40-25 to 41-1). His rebuttal to that proposition, however, was little more than analogizing STRmix to a car and concluding he did not need to take a test drive. (15T 16-20 to 24).

Dr. Heimdahl was questioned regarding reports he had prepared for other trials involving PGS. Not only did Dr. Heimdahl not examine STRmix in any depth or detail in this case, in prior cases, his reports appear to have included assertions lifted from newspaper articles which he accepted uncritically and refused to investigate for himself. (15T 49-21 to 51-20).

Dr Heimdahl's prior reports included assertions that coding errors have been found in the past, therefore errors must exist in the STRmix software. Dr. Heimdahl based these on newspaper reports of LRs needing to be recalculated, which he referenced in his State v. Pickett amicus brief, and breathalyzer calibrations being called into question, as addressed in State v. Cassidy, 235 N.J.

482 (2018). On cross examination, both these summations were proven inaccurate. (15T 22-19 to 24-24; 27-22 to 31-11).

Dr. Heimdahl's testimony also revealed that he was not familiar with the basic fundamentals of DNA analysis. He testified that he did not believe that "traditional DNA analysis" would generate random match probabilities at or above 10 billion to one. (15T 86-25 to 87-8). Therefore, Dr. Heimdahl's testimony provided limited assistance to this court's determination whether STRmix works and whether it worked in this case. This court gave Dr. Heimdahl's testimony minimal weight.

x. Defendant's Expert: Nathan Adams, Expert in the Fields of Software Engineering and Probabilistic Genotyping

Nathan Adams is a systems engineer at Forensic Bioinformatics Services in Dayton, Ohio. (16T 4-20). In this role, he reviews cases, provides consultation, and occasionally will author expert declarations and reports. (16T 5-1). His company will provide testimony predominantly in criminal cases on the topics of forensic biology and forensic DNA analysis. (16T 5-4). Adams has testified as an expert around twenty times, mostly dealing with probabilistic genotyping software. (16T 19-15-20). He started working for Forensic Bioinformatics in 2012, just as the first PGS programs were being developed. (16T 7-14). Now, the majority of his casework involves probabilistic genotyping—whether it is criticizing STRmix or a competing program. (16T 7-14). Mr. Adams has been involved with over 1,000 cases. (16T 107-13). He received his bachelor's degree from Wright University and his master's degree has been underway for some time. (16T 10-9).

Mr. Adams has reviewed the source code of several PGS versions, including STRmix, prior to this hearing. (16T 12-8 to 17).

Mr. Adams believes that software engineering is a "formal process" that necessarily generates certain types of documentation if it is followed correctly. (16T 31-19 to 32-7). He testified that, in his opinion, the "high level" IEEE procedures are the same for all software, but individual tasks will vary depending on the software's "importance." (16T 35-22). Integrity level is an expression of the software's importance, as determined by the hazards or risks inherent in the software's failure. (16T 36-6 to 14). IEEE 1012 defines four integrity levels, and Mr. Adams classifies STRmix as a level four due to the risk of harm and risk of loss of life. (16T 36-37). The higher the integrity level, the more an analyst must understand the risks of a system failing. (16T 37-16).

As a system's criticality increases, the requirement for independent verification and validation (IV&V) increases as well. (16T 37-25). Independence serves to reduce bias and pressure in verification and validation activities. (16T 38-1). Mr. Adams believes that there is an issue with managerial independence of STRmix, due to ESR's involvement with its development. (16T 41 to 42) Because Mr. Adams considers ESR and Orbit Systems to have the same involvement in development activities, he classifies STRmix as having the lowest level of independence on the IEEE 1012 scale. (16T 46-13).

In reviewing STRmix for this hearing, Mr. Adams had access to both sensitive data and publicly available materials. (16T 49-8). This included risk analysis documents and the source code itself. (16T 49-8). He reviewed the source code for v. 2.5.11. (16T 49-23). For this review, Mr. Adams was reviewing whether a full V&V process had occurred and, if it did occur, what supporting documentation had been created. (16T 51-6). Based off the terms of the protective order put in place, he believed he was going to receive a fully executable version of STRmix, with the ability to view the code as he ran the program. (16T 52-10).

Mr. Adams was unable to build a working version of STRmix during his review. (16T 55-56). He stated that he communicated with local counsel for ESR regarding the roadblocks he was facing. (16T 56-8 to 57-17). Because Mr. Adams was unable to build a working copy of STRmix during his review, he could not confirm whether the source code he viewed was the one used in this case. (16T 59-3).

He was scheduled to complete the source code review from November 1st to November 3rd, of 2022. (17T 123-20). Mr. Adams scheduled these dates with STRmix, after he reached out to them to schedule it. (17T 123-9 to 123-12). He was there all three days, and on the first day he realized he had issues building the source code. (17T 125-10). Mr. Adams requested documents that would have allowed him to continue his review, and he received them a little after lunch time on November 1st. (17T 125-24 to 126-2). He continued to have issues being able to build the code, as he needing a license file. (17T 126-6). Dr. Buckleton was not made aware that Mr. Adams was having issues completing his review until he read Mr. Adams's report. (17T 133-5). Mr. Adams's report was dated July 31, 2023. (17T 131-9).

Although Mr. Adams was unable to build a running copy of STRmix, he did review a specifications document for STRmix v. 2.5.11. (16T 77-18). A specifications document lays out "how or exactly what is going to be used as a demonstration that the system does fulfill its requirements." (16T 77-18). If a requirements document details what a program should do, based on conversations with stakeholders, then a specifications document details how a developer met those requirements. Mr. Adams has previously reviewed the specifications documents for STRmix and believes they do not fulfill their purpose. (16T 77-18).

Mr. Adams concluded that version 2.5.11's development process did not comply with IEEE 1012. (16T 87-4). In his opinion, STRmix came closer to achieving IEEE's requirements than any

other PGS. (16T 107-20). However, he did not believe in "partial credit. You comply or you don't." (17T 26-9).

Mr. Adams testified that STRmix needs to be held to strict standards due to the potential harm of an erroneous result. (16T 97-18). He is concerned about quality assurance with STRmix and believes there is no way to analyze the outputs or LRs. (17T 45-12). There are too many calculations for a human to check, so it is difficult to rely on intuition when evaluating STRmix. (17T 45- 21).

On cross-examination, Mr. Adams conceded that the forensic DNA community appeared to "widely regard STRmix as positive." (17T 23-3). However, he felt that had little bearing on how it was viewed according to software engineering principles. While IEEE has no inherent regulatory authority, Mr. Adams detailed several other regulatory agencies that adopted the IEEE 1012 guidelines either in whole or in part, "at which point compliance is no longer voluntary." (17T 25-8 to 25-25).

Mr. Adams claimed that he is unable to rely upon knowledge he had gained from previous code reviews of STRmix. Mr. Adams did not feel comfortable even detailing whether he expected to see changes in the source code when compared to prior reviews. (16T 135-7).

As to the roadblocks Mr. Adams faced in this review, there is ambiguity as to what those roadblocks were, and to what extent they could have been solved. Mr. Adams claims that he made requests of ESR's local counsel, but he identified documents during redirect examination, purportedly from ESR's counsel, claiming he did not make those requests. (17T 36-20).

Whatever difficulties occurred, and whatever was done or not done to solve them, Mr. Adams consistently maintained on direct, cross, and redirect examination, as well as examination

by this court, that he has yet to see the level of testing and documentation that would satisfy him—those tests and documents "simply don't appear to have occurred for this system." (17T 47-17).

Mr. Nathan Adams is the only expert proffered by the defendant who has been trained to use STRmix.[8] To the extent he has used the program and based his opinions off its limited use, he offers credible testimony on the quality and reliability of the software. His testimony showed he has specialized knowledge and experience with STRmix that could assist the court. The extent that he has refused to engage with the program despite his training and experience, however, concerns the court.

Mr. Adams has suggested that he is aware of serious flaws in the program from prior reviews but was unable to reproduce them during this review. He claims that prior NDAs and protective orders limit his ability to comment on the previous flaws noted, but they do not seem to prevent him from implying they exist. Mr. Adams could not and did not substantiate any of his insinuations in his code review in this case. The court holds Mr. Adams to his word regarding his limitations, but his inability to reproduce his findings in this case gives the court pause in evaluating his credibility.

Whatever technical difficulties Mr. Adams faced that prevented him from completing his source code review in November of 2022, the record is clear that he did not reach out to STRmix for assistance, nor was the court notified of any issues. The court knew that STRmix offered to provide substantial involvement and promised access to defendant's experts. Mr. Adams did not notify Dr. Buckleton or his representatives that he was unable to perform the source code review.

[8] While defense counsel is also trained in the use of STRmix and gave an in-court demonstration during Dr. Reed's cross-examination, counsel is not a witness in this case.

Through Mr. Adam's testimony, the court observed that he had sufficient knowledge and experience to perform the source code review and understood what actions to take if he was unable to complete it. He knew there were representatives at ESR he could have contacted. He chose not to.

It is important to note that defendant's next expert, Dr. Paul Martin, also encountered issues in his code review of STRmix 2.8, contacted ESR, and his problems were solved. (7T 131-24). There is no reason to believe ESR would not have accommodated Mr. Adams, but he did not communicate to find out—not just in November 2022, but for 8 months afterwards. Instead, Mr. Adams simply detailed what he would have liked to do and his best guesses as to what he'd find based on the insufficient information in front of him. It is one thing to encounter a roadblock, but it is another to stop at the first one and declare the whole road ahead to be unreliable. The court finds his lack of effort to review the source code concerning.

Mr. Adams' credibility is further undermined by his mischaracterization of the industry requirements for accuracy studies. In the conclusion of his July 2023 declaration, he suggests that laboratories should go beyond compliance with ANSI/ASB Standard 018's accuracy requirement by showing they can meet the same requirement for samples where the "ground truth is not known." D-17, Declaration of Nathan Adams 15 (July 2023). However, Standard 018 is clear—as Mr. Adams acknowledges—that those types of samples are unsuitable for accuracy studies. When questioned on cross-examination about this inconsistency, Mr. Adams was evasive and did not clarify his position. (16T 118-6). Although he summarized the standard correctly in other portions of his declaration, his final summation attempts to set prerequisites for STRmix's reliability that simply do not exist.

While Mr. Adams' qualifications and experience in the field of software engineering support his credible testimony on the common practices of that industry, and on formal IV&V processes, he was not able to corroborate his testimony regarding STRmix beyond a cursory assessment that its developers did not follow the process he would have chosen. He also could not give a specific example of unreliability in STRmix's source code, only "deficiencies and deviations from software engineering norms." D-17 at 16. His opinions were given minimal weight.

xi. Defendant's Expert: Dr. Paul Martin, Expert in the Field of Software Engineering

Dr. Paul Martin works at Harbor Experts as their chief scientist. (18T 5-12 to 17). In this role, he performs research projects, testifies in court, and conducts casework. (18T 5-17). Dr. Martin is also a professor with the Computer Science Department at John Hopkins University and teaches security and privacy courses. (18T 6-7). He is a member of IEEE and was on the publishing committee for this organization. (18T 14-10). Dr. Martin has a Bachelor of Science and Master of Science in Engineering, and a Ph.D. in Computer Science from John Hopkins University. (18T 9-10). Dr. Martin is familiar with multiple coding languages. (18T 15-4).

Prior to this case, Dr. Martin has performed source code reviews of commercial software. (18T 15-19). He does these reviews many times a year. (18T 15-21). This case is his third or fourth time testifying in court, but he has undergone several depositions. (18T 18-17). He was qualified as an expert in software engineering. (18T 20-12). Dr. Martin testified that in every source code review, the reviewer is trying to figure out how the software works usually. (18T 25-7). This entails reviewing source codes, for many reasons, to include assessing general quality, and looking to see if there are misspellings in the codes. (18T 25-7).

Dr. Martin testified about software failures. (18T 20-22). Failures can happen when there is an update to a system that is not adequately tested. (18T 21-7). So, a bug a can occur which causes the system to crash and not properly work. (18T 21-7). Dr. Martin argued that if your software is failing because it is not meeting your expectations, that can be called a bug. (18T 22-12). A bug is a term for software fault and is a loose term. (18T 22-12). There could be software failures that are not considered bugs. (18T 22-12).

Dr. Martin testified about the validation and verification process. (18T 31-19). V&V is the process of trying to ensure that the software works as intended and meets requirements. (18T 31-19). There are two relevant sets of requirements, according to Martin: customer requirements and engineering requirements. (18T 31-19). Customer requirements are for how the product should function for the customer. (18T 31-19). Engineering requirements are requirements on how the software should function as designed. (18T 31-32). V&V is a formal process that a reviewer should go through to ensure that the software performs as expected and correctly. (18T 3-3). Verification is a process where the tests are done to see that the software performs as expected based on the software requirements. (18T 32-3). STRmix has been published in several peer reviewed publications, which is a part of the process for V&V but not a substitute for it. (18T 33-14). Dr. Martin testified that being published is more in line with the validation step. (18T 33-18). The model itself being published says nothing about the source code or how the model has been implemented. (18T 34-6).

Independence is important for validation. (18T 36-10). Dr. Martin discussed how independence comes into the process when there is concern about conflicts of interest. (18T 36-15). An independent body can confirm code, in order to mitigate bias. (18T 37-5). Dr. Martin testified

that he believes a PGS is a safety critical system due to there being risk to life. (18T 37-19). These systems need independence in their reviews. (18T 37-12).

Dr. Martin had issue with the verification of STRmix. (18T 41-19). His opinion was that there was insufficient testing performed to validate STRmix as software. (18T 42-16). There was no unit testing, integration testing, or component level testing done for version 2.5.11, based on his observations. (18T 42-16). He expected to see these tests as a generally accepted practice. (18T 43-8).

Dr. Martin also testified about his review of the requirements document for this case. (18T 43-25). He called the document "odd compared to other requirements documents." He believed it was prepared after the fact, and not as detailed as Dr. Martin would have liked. (18T 44-23). This did not mean the document was not up to standard, just short on detail. (18T 47-14).

Having reviewed both versions at issue in this case, Dr. Martin testified about some of the differences between versions 2.5.11 and 2.8. (18T 49-50). One improvement with 2.8 was that a function known as SafeMath was included. (18T 49-18). SafeMath checks to see that the mathematical operations are safe due to a computer system only having a fixed amount of space to hold numbers. (18T 40-21). It is designed to prevent errors when numbers are larger than the size allocated for them; it will check and vary the space. (18T 50-11). Dr. Martin explained that SafeMath's absence from v. 2.5.11 gave him concern that those errors were not being anticipated and handled. (18T 50-51). He believed that IV&V for v. 2.5.11 would have caught this. (18T 51-6).

Ultimately, Dr. Martin thought the code was good and complied with engineering standards. (18T 66-6). He thought both versions were "professional" and written well. (18T 66-8). In his view, v. 2.8.0 was better, as he felt it was more rigorously tested and incorporated SafeMath. (18T 66-16).

Dr. Martin's review of both versions was limited by the scope of what he was hired to perform. He did not perform a deep or comprehensive review, but he "looked at" and "surfed across all of it" but only "looked at specific parts" in detail. (18T 78-18) He had "no complaints" about the scope of his hire and would have been "happy" to conduct further reviews, but did not think he was permitted or asked to go further in depth in this matter. He was not "bug hunting" but looking "at the overall quality of the code" in a cursory manner. (18T 97-10 to 99-12).

Both versions, in his opinion, did not comply with IEEE 1012, as he perceived no independence in the validation and verification processes. (18T 67-68). He did acknowledge that IEEE 1012 compliance is not mandatory. (18T 79-15). Dr. Martin only performed a static code review but would have liked to have done a dynamic code review. (18T 98-14).

The court found Dr. Martin credible. While he made the same general critiques about the perceived lack of documentation as some other experts, he was candid in his assessment and consistent on cross-examination. Dr. Martin testified that the requirements document was a little "odd" compared to other ones he has seen because it looked like it was a written after the fact. But later he did testify that it is possible for the software to be developed before one is written, because it is more important to have a requirements document written at some point rather than never. Dr. Martin testimony regarding the requirements documents demonstrated his candor with explaining the advantages and disadvantages of STRmix. He articulated his concerns clearly, while mentioning what he has typically seen occur in other situations. He testified that STRmix's requirements document was detailed in some parts but could have been more descriptive in others. He presented objective testimony regarding the documentation both on direct and cross examination.

On cross examination, Dr. Martin thoroughly answered questions about STRmix and explained his rationale for his opinion in the process. The court found Dr. Martin testimony regarding

general software engineering to be helpful and informative. He testified about the software used to perform DNA analysis, but he has no experience in DNA analysis to help him determine what "requirements" would be needed for that analysis. Dr. Martin only generally understands how STRmix works and how it is used by DNA analysts. (18T 77:20).

Dr. Martin was willing, unlike some other experts, to distinguish between his own preferences for how things should be done, and baseline standards. He believed that STRmix's code was well-written, and he entertained the idea that a professional and well-written code can comply with reasonable engineering standards even if he would have done things differently. He felt his review was constrained and would have preferred a dynamic code review, (18T 98-99), but he was not precluded from doing so by anything other than the scope of what he was hired to do. To the extent that this comment was meant to suggest that Dr. Martin's difficulties were of the same kind that Mr. Adams complained of, the court was not persuaded. Dr. Martin's opinions were afforded significant weight but were found to consist mostly of superficial impressions of STRmix.

xii. Defendant's Expert: Keith Inman, Expert in the Fields of Probabilistic Genotyping and Forensic DNA Analysis

Mr. Keith Inman is an associate professor at California State, East Bay, and he also performs part-time consulting work in forensics. He is currently a visiting professor at the University of Dundee, Scotland. He has a bachelor's degree and master's degree in criminalistics from University of California, Berkeley and is completing his PhD at the University of Dundee. He completed his Master's in 1978. (19T 57-13 to 58-2)

Mr. Inman was previously employed in the Orange County Sheriff's Crime Laboratory, then the Los Angeles County Crime Laboratory. He also worked for the Chief Medical Examiner's

Office in Los Angeles, a private laboratory in Southern California, and then the California Department of Justice Laboratory in Berkeley starting in 1990. (19T 58-3 to 58-17)

At the California Department of Justice, he was one of the founding scientists in their DNA laboratory. After completing the initial validation of the DNA laboratory, Mr. Inman worked as a case worker within that laboratory and was later in charge of training DNA case workers. He was also in charge of implementing CODIS and testified that California was one of the laboratories designated as an experimental developer for the national DNA database. (19T 58-18 to 59-15)

Mr. Inman testified to his experience in designing and implementing internal validation studies. He detailed the process for developing protocols that allow a laboratory to understand "not only the capabilities [of a laboratory] but also the limitations." (19T 60-4). In other words, how one can know when they're wrong. In his career, he participated in many internal validation studies and tested "thousands to tens of thousands" of forensic samples. (19T 61-3).

Mr. Inman testified to his involvement with the National Research Council's 2009 publication, Nat'l Research Council, Strengthening Forensic Science in the United States: A Path Forward (2009), which "assess[ed] the current status of forensic science within the United States, what was good, what needed work, things of that sort." (19T 64-5 to 67-4). He was not a committee member or credited author of this paper but is a credited peer reviewer. Id. at *xiii*. The overall conclusion of this paper was that forensic science was in a "fairly miserable or abysmal state," but that was highly variable by sub-discipline. He was also a peer reviewer for the 2021 NIST Draft Foundation Review for DNA Mixture Interpretation, which has been superseded by the 2024 final publication, cited supra.

Mr. Inman was also questioned about his experience with PGS. He testified that, in 2009, he was first introduced to the concept of Probabilistic Genotyping via a research paper. He implemented the algorithms required for mixture analysis into a simple software program in order to "explore . . . the capabilities" of the new technology. (19T 65-12). While he has not been involved in the development of any software systems since that initial proof-of-concept attempt, he testified to his familiarity with several PGS platforms, noting there were about "10 to 12" programs on the market and that he was particularly familiar with a few platforms prevalent in Europe. (19T 66-6).

Turning to his assessments of STRmix, Mr. Inman summarized his chief concerns with STRmix and PGS platforms in general as "simply one part of a very long process of trying to understand or decide whether someone is a part of a DNA profile. Probabilistic genotyping is really just the end of that, but everything that goes ahead of that is equally important . . . whatever you've done before is going to impact what comes at the very end." (19T 71-3). He testified that it was important "to insure [sic] that all of the processes that go into a DNA profile are well understood," (19T 71-10), because the person making the final analysis must understand how the process impacts the data, and what a result actually means in the context of the case.

Since there are over seven billion people on the planet, Mr. Inman observed, the number of possible combinations of DNA within a mixture is "beyond human comprehension." (19T 72-6). The question, therefore, is how many different combinations of DNA would need to be sampled or simulated in order to know that any given system has been tested against a reasonable cross-section of possible combinations. He detailed four main factors that add to the complexity of a DNA sample: 1) NOC; 2) amount of DNA and the ratio between contributors; 3) degradation; and 4) inhibition. These can occur independently, or all at once and "conspire to make a complex DNA

profile." (19T 74-2). Mr. Inman defined degradation as the DNA breaking down in the environment, and inhibition as "something in the environment that basically gets in the way of your test." (19T 73-18). For a program like STRmix to be properly validated, Inman believes, a laboratory has to make samples "as complex as you expect to find within your casework" (19T 74-12) and even to exceed that complexity as much as possible. This is both to enable you to test samples that exceed the complexity expected in the field but fall within the prepared validation samples, and to allow a laboratory to know when to "back away" from testing. (19T 74-25).

Mr. Inman drew a distinction between the breadth of testing required for internal validation, and the depth of testing required to understand what he referred to as the "factor space." (19T 75-10). In short, he explained that while a laboratory may conduct a study that illustrates their ability to handle a certain range of samples, it is difficult to know if every possible permutation of samples within that range is handled correctly without a huge amount of data; if every laboratory is coming to the same conclusion with limited data, but "may reach some slightly different conclusion" when all the data is taken together, that would show a limited understanding of the factor space. Validation studies that show consistent results with limited varieties of samples show "reproducibility," which is important but not the whole picture. (19T 75-10 to 76-10). Inman mentioned that many labs compiling data into public repositories was something that he "thinks for the most part . . . doesn't happen." (19T-19). This opinion, however, is directly contradicted by several articles and experts.

To illustrate what he meant by factor space and mixture complexity, Mr. Inman described some of the differences between a multi-person mixture made of random contributors, and one made from contributors who are close relatives. He testified that both hypothetical samples wouldn't be much different in terms of the extraction and amplification stages, and it would be

roughly the same process to produce an EPG from both. "Rather, what it complicates is the interpretation," he claims, "essentially, what's the probability of seeing this evidence if it's not the person that you have in hand, but someone at random." (19T 80-8 to 22).

The complication with relatives, he continued, is that the persons present in the mixture are not random. There will not be a broad distribution of possible alleles in the sample and therefore will not be as diverse a section of the general population to compare against for a random match. A mother and father have four potential alleles between them for any given locus. They will each pass one of those alleles onto any progeny. Of the four possible combinations, a child has an equal 25 percent likelihood of inheriting any one combination if each parent has two different alleles—or higher if they happen to have duplicates themselves. Because a LR is calculated based on how common those alleles are in the general population, if a non-contributor is compared to a sample that contains DNA from their father and sister, for example, the calculation may overstate the likelihood that this person's DNA is present. The proposition that an individual is more likely to be a contributor than a random person is "completely separate" from the proposition that an individual is more likely to be a contributor than their brother. Mr. Inman's opinion is that, when it comes to determining which subset of the population we should use as the denominator when calculating an LR, "we don't have a good way of doing that yet." (19T 81-19 to 83-15)

The presence of related contributors, or the presence of contributors that share common alleles for any other reason, also makes it more challenging for an analyst to determine the NOC in a sample. Because the loci examined for DNA analysis are what Mr. Inman describes as "medium heterozygote" loci, there are a limited number of possible alleles to examine. (19T 84-22). When plotted on an EPG, the presence of more alleles at a given location results in greater peak heights. The more contributors present in a sample, the greater the change they share alleles, which

Mr. Inman claims is a proposition supported by several studies. It can be difficult to determine if a raised peak is present because one person contributed more DNA to the mixture, or because multiple people have an allele at that location. The more contributors, the more likely that an analyst will underestimate the number. It is a "perfectly legitimate strategy" to approach this problem by calculating LRs multiple times, with a different NOC, but deciding which result is correct is "a harder problem." (19T 85-1 to 14). In Inman's estimation, underestimating the NOC could result in an artificially deflated LR for an actual contributor, or even a false inclusion or exclusion. (19T 86-2 to 8). He did not view this as unique to STRmix, but as a risk to any PGS that underestimated contributors. (19T 86-17).

One of the "purposes of validation" is to determine what a laboratory should do to solve the problem, Mr. Inman claims, by "determining what a good likelihood ratio should look like." (19T 87-9).

Based on Mr. Inman's observations, he was of the opinion that neither the Bode technologies nor the New Jersey State Police laboratory's internal validations sufficiently tested samples comprised of related contributors. If he were testing a sample under those circumstances, Mr. Inman claimed he would "step away immediately" once he noticed there were related contributors, because he would not "have a solution for what a good likelihood ratio would look like." Later, Mr. Inman expressed his opinion that he did not believe it was possible to achieve a successful testing threshold for samples containing related contributors at all, and that he did not believe it was ever appropriate to calculate an LR based off such a sample. Mr. Inman was also critical of the SWGDAM verbal scale for communicating the support provided by LRs. (19T 86-18 to 91-12).

In another illustration of LRs, Mr. Inman explained that an LR of 1:10,000, which would be characterized as "strong support" under the SWGDAM scale, means that any random individual may have a 1 in 10,000 chance of sharing enough DNA with the sample to be considered a potential contributor. If you compare that to a population of eight million people, it would be expected that eight hundred individuals within that population would be falsely included. (19T 89-25 to 90-8). This holds true whether the sample they are compared to is a multiple-contributor sample, or a partial single-contributor sample where you "don't have all the information." (19T 90-20)

Before going more in-depth regarding the validation studies from Bode and the New Jersey State Police, Mr. Inman also explained the concept of evidential efficiency statistics. In short, he described how the random match probability for a hypothetical person, calculated using a "pristine" DNA sample, would be expressed as a ratio somewhere in the neighborhood of $1:1^{25}$ (a one followed by twenty-five zeros). For another person, it may be only $1:1^{18}$ (a one followed by 18 zeros. These numbers, while both large, are incredibly disparate. The person with fewer zeros has more common DNA and would be expected to generate larger LRs when compared to a partial sample.

If the person with twenty-five zeros in their RMP—a one in ten septillion chance of randomly matching with any given DNA sample—returns an LR of 10,000 for a complex mixture, then their evidential efficiency is 10,000 divided by 1^{25}, or 0.16. In other words, only about 16% of the information that could be generated by a full DNA profile is present. The small LR is only 16% as efficient as a true RMP, for this individual. (19T 94-20 to 96-12).

Turning to Mr. Inman's analysis of the Bode validation summary, he felt there were two shortcomings in relying on the summary documents for his analysis: first, a summary is not the

same as unpacking the raw data. He analogized a summary document to being like a batting average—a compilation of relevant statistics that can give you an overall sense of performance but not the same as watching the game. Put another way, a road map is not a road. His second concern is that there is a necessary level of subjectivity in preparing a summary: someone has made a choice to include the data they felt important, but someone else may come to a different conclusion. (19T 98-2 to 98-17).

Inman did not think that observation was "necessarily terrible," just that examining data with different questions in mind could yield different conclusions. Referring to the Bode summary, he found the information presented to be a "good example" but didn't lend itself to him drawing independent conclusions. Mr. Inman was not tasked with examining the underlying data from the Bode study, and testified later on that doing so would have been a considerable undertaking. The document he examined allowed him to review Bode's validation "generally but not specifically." (19T 98-18 to 99-25).

He had similar concerns about the New Jersey State Police validation summary. Mr. Inman identified some of the graphs within the study that he would have liked to see zoomed in for more detail. He described summaries of the study as the New Jersey State Police trying "to break the system" but could not, in his opinion, determine if they "hit it hard enough… to truly break it." (19T 100-1 to 25).

Unlike the Bode study, Mr. Inman was provided with the underlying data for the New Jersey State Police study. He described it as "an enormous amount of data" that took "easily thousands of hours" to compile. He had made progress on an independent analysis of this data but had not completed it at the time of the hearing. He was confident that, given more time, he would have been able to complete his analysis but wanted six more months to do so. (19T 134-19). He believed

that more edge cases, samples that pushed the boundaries of STRmix's capabilities, could be identified and tested given time. (19T 102-11 to 104-24). Despite receiving the underlying in May or June of 2024 (19T 102-11) and testifying to it in December of 2024, however, Mr. Inman had not been able to identify any flaws or draw any conclusions as to the New Jersey State Police validation study. (19T 141-16).

Moving onto the topic of best practices, Mr. Inman referenced the NIST Human Factors report and gave his opinion on recommendation 3.2: namely, that laboratory analysts should document their work as they do it as well as their judgement calls and decision-making processes. Everyone can forget what they did or why they made one choice over another, so the best practice is therefore to keep a contemporaneous record "transferring what's from your brain onto a piece of paper so people know… why you did what you did." An analyst's determination on how many contributors are present in a sample, for example, should be thoroughly documented as to the judgment calls made. He stated that, while it's not acceptable for an analyst to forget why they made a particular call, everyone should write it down. (19T 105-4 to 108-25)

Inman then observed several validation studies from other jurisdictions but did not opine upon them. (19T 118-9 to 122-14). When asked if he viewed SWGDAM guidelines as mandatory, he observed that they're mandatory "if you want federal funding" or "to put samples into CODIS" and thus were a *de facto* mandatory standard. (19T 118-3).

On Cross-Examination, Mr. Inman clarified that the concept of related individuals as a factor that can impact DNA matches is not a new concept, but understanding its impact was evolving. Ultimately, Inman claims the aim of a validation study is to "determine what way [a PGS platform] is fit for purpose." (19T 136-1). When questioned on his progress in evaluating the New Jersey State Police validation data, he clarified that he was in possession of the data since May or

June of 2024 and, while he had never performed a validation study of STRmix specifically, he had experience validating competing software platforms. (19T 129-16 to 130-16). Given the state of progress in his current review, he was unable to draw any specific conclusions about the underlying validation data for either laboratory. He estimated he probably would not be able to conclude his study of the New Jersey State Police data until "the middle of" 2025. (19T 134-19). He had not received any Bode data at all but thought a review of that might be able to be accomplished "at the same time." (19T 135-12). He was asked about some difficulty he had with sorting through the initial batch of data provided by the New Jersey State Police and conceded that he tried to figure out their file naming conventions on his own rather than calling for clarification, because he was unsure if it was appropriate to communicate with the State Police Laboratory. (19T 150-15 to 151-23).

On examination by the court, Mr. Inman put his concerns with STRmix into the bigger picture. It is the last link in a chain that has its own flaws and propensities for errors along the way means that problems with an earlier step become amplified. STRmix analyzes tiny amounts of data, which means that samples which were not considered trustworthy before are being tested. Inman wants to be sure they should be tested now. (19T 161-1 to 168-10).

The court found Mr. Inman to be credible. He has vast and relevant experience, and has been familiar with probabilistic genotyping since its inception. He has been involved, in an advisory capacity or as a peer reviewer, in the development of many standards and guidelines related to DNA over his career. His answers on direct and cross examination were confident and consistent. He clarified his points and corrected misunderstandings as he answered.

Mr. Inman admitted that he did not review STRmix's data, even though he has had access to this data for months. Mr. Inman never analyzed either laboratory's internal validation studies in

sufficient detail to offer any meaningful opinion. His ability to speak directly to the reliability of STRmix was therefore limited as a result.

His general opinions on best practices and how a laboratory should validate STRmix are relevant, but by his own admission, there is more than one way to conduct a validation study and write and implement a standard operating procedure. Under best practices, a laboratory should allow for variation from the standard operating procedure. Mr. Inman discussed how SOPs can guide an analyst and then allow for them to go a different direction if needed. This part of Mr. Inman's testimony demonstrated his willingness to explore favorable and unfavorable aspects of STRmix. It showed he was examining the information provided to him rationally and objectively.

Mr. Inman's concerns about the variability of DNA testing in general are valid points, but these concerns are not limited to PGS. Conventional DNA testing, like all science, has limitations and must be performed carefully to ensure reliable results. This observation does not render DNA testing itself unreliable by having variability. Mr. Inman explained that STRmix can be right, but the results generated are only as good as the data the analysts are inputting. Mr. Inman's experience and expertise are balanced against his limited review of the actual validation data in this case and are thus, given moderate weight. His testimony is helpful in framing what the court should focus on, but falls short of a comprehensive analysis of STRmix's reliability.

b. Exhibits Presented

The State and the defendant have submitted thousands of pages of exhibits to the court. These include not only the reports generated using STRmix and declarations from their experts, but also various standards and guidelines relevant to forensic DNA analysis, the user manuals for STRmix, numerous scientific articles, college textbooks, Google's testing requirements for their

"App Store," newspaper articles, and internal validation summaries from both the laboratories in question and other laboratories across the country.

Some of these exhibits, such as the laboratory reports generated for this case and the expert declarations, are unquestionably relevant. Others have only limited relevance, due to a lack of proper foundation or explanation from the experts during testimony. Some exhibits, such as chapters of textbooks or summaries of Google's app store requirements, are tangential but useful as background information. The court has reviewed and considered the entirety of the exhibits, but has assigned different weight to each. Several exhibits were of particular relevance to the court:

The expert declarations and reports filed by each testifying expert have been considered extensively. The portions that corroborate their testimony are given the appropriate weight. In instances where the written report and the expert's oral testimony were inconsistent or in conflict, the court evaluated the report and the oral testimony with caution.

For foundational knowledge on forensic DNA analysis, four publications by Dr. John Butler were helpful: John M. Butler, Fundamentals of Forensic DNA Typing (2010), John M. Butler, Advanced Topics in Forensic DNA Typing: Interpretation (2015), John M. Butler, Advanced Topics in Forensic DNA Typing: Methodology (2011), and Suzanne Bell & John M. Butler, Understanding Forensic DNA (2022).

For a general understanding of the principles of software engineering, this court considered Ian Sommerville, Software Engineering (10th ed.) (2016).

The PCAST report, The President's Council of Advisors on Sci. & Tech., Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods (2016), was discussed at length by several experts. It has clearly shaped the conversation on PGS since its 2016 release, despite only briefly addressing the topic.

The SWGDAM guidelines, Scientific Working Group on DNA Analysis Methods, Guidelines for the Validation of Probabilistic Genotyping Systems (June 2015), have become *de facto* standards since their promulgation in 2015. The experts proffered by both the State and the defendant uniformly agree that SWGDAM's expectations are respected and followed by members of the forensic science community.

The standard promulgated by the Institute of Electrical and Electronic Engineers Standards Association, IEEE 1012-2016: Standard for System, Software, and Hardware Verification and Validation, and discussed by the experts at this hearing, has been superseded by a 2024 version. While both versions were considered, the 2016 document is the document that was considered extensively by this court, as it is the version referenced by the witnesses.

There were several other standards, guidelines, and recommendations considered by this court, all of which are addressed infra.

Jo-Anne Bright et al, Internal Validation of STRmix – a multi-laboratory response to PCAST, 34 Forensic Science International: Genetics 11 (2018) is a comprehensive and thoroughly supported article that addresses PCAST's requests for more publicly available multi-laboratory data studies. The court has considered this document extensively and relies upon it. Similarly, the developmental validation study conducted for STRmix, Jo-Anne Bright, et al, Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles, 23 Forensic Science International: Genetics 226 (2016), is relevant to the reliability determination before the court. Other peer reviewed articles and studies which were not discussed by the experts are given lesser weight. In instances where the State provided the titles of articles, but the articles themselves were neither discussed nor provided, the court may acknowledge it but will give it no further

weight in favor of reliability. Articles that were cited in brief, or provided but not directly discussed by experts, were given greater consideration.

Dr. Buckleton's summary of miscodes, provided by this court as exhibit S-137, has been considered but is not taken to be a comprehensive summary of all possible errors. To the extent the identified miscodes within have been explained and contextualized by either the State's or defendant's experts, that information has been considered for that purpose. The court does note that the document is not peer-reviewed as it was taken directly from STRmix's website. The list is available for public viewing and has been updated periodically since 2018.

STRmix's user manuals were of assistance to this court, both to the extent they were discussed by the experts and illuminated decisions made during STRmix's design, and as reference while considering the internal validation studies of the respective laboratories.

The other exhibits that assisted this court are the actual validation studies from the laboratories that have performed casework in the instant matter, and the actual EPGs, lab notes, and STRmix reports generated for the samples analyzed for this case. Validation studies from other laboratories in other jurisdictions are accepted for the limited purpose of illustrating how Bode Technologies and the New Jersey State Police laboratory aligned in their approach, and how they differed. The SOPs and internal manuals from the Bode Technologies and the New Jersey State Police laboratories were also considered.

The absence of an explicit reference to an exhibit above, or the absence of a citation to it within this opinion, should not be interpreted as a statement that an exhibit was not helpful or not considered. As was explained during the hearing, any exhibit not identified or discussed during testimony but still entered into the record was limited in scope, but still considered in its entirety and placed in the appropriate context.

VII. LEGAL STANDARD

a. Summary of the Standard

Under Olenowski II's summation of the non-exclusive Daubert factors, the question before this court is whether STRmix reliably analyzes complex mixtures, and whether the laboratories in question properly validated STRmix to analyze mixtures of the type found in this case. STRmix's general scientific validity is supported by showing:

- that it is governed by adequate standards, primarily the standards of the Forensic DNA community, but mindful of the relevant standards of the software engineering community;
- that it has been published and subject to peer review according to the standards of the scientific community conducting the peer review;
- whether it is testable (and has been tested) sufficiently to determine a general error rate and how said errors can be mitigated;
- the extent to which STRmix is generally accepted by the scientific community that implements it.

[See Olenowski II, 255 N.J 587-604 (2023)]

When considering the adequacy of standards, the standards need not be a "checklist" where those applying the method are "compelled to make a particular finding[.]" Olenowski II, 255 N.J. at 587. A methodology also need not meet every possible standard identified. The inquiry is whether there are adequate standards to support a consistent and reliable implementation generally. Ibid.

The existence of peer reviewed publications discussing the methodology is strong support for a finding of reliability but is not dispositive. Daubert, 509 U.S. at 593 (peer review "is not a sine qua non of admissibility; it does not necessarily correlate with reliability"). Some fields "are not typically studied in peer reviewed academic journals." Olenowski II, 255 N.J. at 590. When considering the reliability in the field of forensic science, where peer review is common and ac-

cepted, the presence of reviewed articles may be strongly probative. In the field of software engineering, peer review takes on a different form and is synonymous with "code review" See Sommerville at 727 ("In a program inspection or peer review, a small team systematically checks the code. They read the code in detail and look for possible errors and omissions. The problems detected are then discussed at a code review meeting.")

Similarly, testability looks different in a laboratory setting and a software setting. Both types of testing support a finding of reliability in different ways and must be considered according to their own standards.

For the general acceptance prong, this court is not seeking unanimous consensus from the experts who testified in the hearing, but a showing that STRmix is accepted by the relevant communities at large. "Scientific acceptability need not be predicated upon a unanimous belief or universal agreement in the total or absolute infallibility of the techniques, methodology or procedures that underlie the scientific evidence." Romano v. Kimmelman, 96 N.J. 66, 80 (1984). This factor, once the sole determinant of reliability, is no longer dispositive on its own:

> It might not be surprising in a particular case, for example, that a claim made by a scientific witness has never been the subject of peer review, for the particular application at issue may never previously have interested any scientist. Nor, on the other hand, does the presence of *Daubert's* general acceptance factor help show that an expert's testimony is reliable where the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology or necromancy.
>
> Kumho Tire Co. v. Carmichael, 526 U.S. 137, 151 (1999).

The reliability determination before this court is constrained to the question of whether STRmix can reliably deconvolute complex mixtures of the type found in this case[9] and whether the laboratories in question sufficiently validated STRmix and used it reliably. Questions about the propriety of LRs or the verbal scales used to report them, the biases or motivations of the experts testifying at trial, or other ancillary matters go to the weight of the evidence.

b. Detailed Standard

As an initial matter, the defendant is correct that the State's burden is to "clearly establish" STRmix's reliability. Cassidy, 235 N.J. at 492. The defendant, however, bears the burden for his own arguments. State v. Rosales, 202 N.J. 549, 562 (2010) ("The party offering the expert testimony has the burden of proof to establish its admissibility."). While "the burden is not on the defendant to sift through the State's evidence and demonstrate the accuracy of the scientific method it is choosing to rely on," (def. Br. at 134), he must support his own specific claims that the method is inaccurate and support his own assertions. See State v. Angeleri, 51 N.J. 382, 385 (1968).

The court is limited in how it may consider evidence that has not been explained by an expert. "Although trial courts are expected to act as gatekeepers to the proper admission of expert testimony, we do not expect courts to investigate *sua sponte* the extent to which the scientific community holds in esteem the particular analytical writings or research that a proponent of testimony advances as foundational to an expert opinion." Hisenaj v. Kuehner, 194 N.J. 6, 16 (2008).

---

[9] The samples in question are all found to be 3- and 4-person mixtures containing probable first- and second-degree relatives. Some mixtures were calculated to contain under 25pg of DNA from the minor contributor, and some were calculated as containing DNA from the minor contributor at 5% or less, but none were reported to be below both those thresholds. See Appendix.

> [I]n determining the soundness of [a] methodology the trial court should [not] directly and independently determine as a matter of law that a controversial and complex scientific methodology is sound. The critical determination is whether comparable experts accept the soundness of the methodology, including the reasonableness of relying on this type of underlying data and information. Great difficulties can arise when judges, assuming the role of scientist, attempt to assess the validity of a complex scientific methodology.
>
> [Rubanick v. Witco Chem. Corp., 125 N.J. 421, 451-52 (1991) (internal citations omitted).]

The underlying processes common to forensic DNA analysis are also not at issue. Nor is the concept of LRs and how they should be communicated in testimony; that question is long settled. State v. Spann, 130 N.J. 484, 519 (1993). DNA evidence has been accepted as admissible in this State starting in the early 1990s, and new and novel methods of interpreting DNA evidence have steadily gained acceptance since. See State v. Harvey, 151 N.J. 117. DNA is a legitimate subject for expert testimony at trial." State v. Marcus, 294 N.J. Super. 267, 288 (App. Div. 1996). In Marcus, the court held that "any dispute regarding the statistical significance of matching DNA profiles does not affect the general acceptance of DNA analysis within the scientific community and consequently should not result in the exclusion of such evidence in criminal trials." Id. at 291.

The issue of DNA admissibility first reached the New Jersey Supreme Court in 1997, in State v. Harvey,. 151 N.J. 117 (1997). The Court, in Harvey, was tasked with evaluating the admissibility of a novel method for analyzing mixed samples: the polymarker test. Ibid. The Supreme Court held that the trial court did not err in admitting testimony on the results of the polymarker test because the "technology is scientifically reliable" and the tests were in "accordance with established procedures." Id. at 176.

"DNA testing is an evolving science[.]" Id. at 155. STRmix is one such evolution. While other jurisdictions have evaluated STRmix's reliability, some under standards consistent with New Jersey jurisprudence, its reliability has not been addressed in New Jersey.

New Jersey Rules of Evidence 702 and 703 control the admissibility of expert testimony. In re Accutane Litig., 234 N.J. 340, 348 (2018). N.J.R.E. 702 governs the admissibility of expert testimony, and states as follows:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify thereto in the form of an opinion or otherwise.
>
> [N.J.R.E. 702.]

New Jersey Rules of Evidence 703 addresses the factual foundation for expert testimony and states:

> The facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to the expert at or before the proceeding. If of a type reasonably relied upon by experts in the particular field in forming opinions or inferences upon the subject, the facts or data need not be admissible in evidence.
>
> [N.J.R.E. 703.]

The party offering the expert testimony has the burden of proving its admissibility by a preponderance of the evidence. Rosales, 202 N.J. at 562. In order to satisfy N.J.R.E. 702, an expert must establish three things: "(1) the subject matter of the testimony must be 'beyond the ken of the average juror'; (2) the field of inquiry 'must be at a state of the art such that an expert's testimony could be sufficiently reliable'; and (3) 'the witness must have sufficient expertise to offer the' testimony." State v. J.L.G., 234 N.J. 265, 280 (2018) (quoting State v. Kelly, 97 N.J. 178, 208 (1984)). The reliability requirement under this rule was formerly governed by the test found in

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923), and "for decades . . . turn[ed] on whether the subject of expert testimony has been 'generally accepted' in the relevant scientific community." Olenowski I, 253 N.J. at 138

Olenowski I summarizes New Jersey's shift away from the Frye test. Until recently, New Jersey courts looked to Frye to determine reliability in criminal matters but began the shift to a methodology-based approach in civil matters in 1991. With the holdings in Rubanick v. Witco Chem. Corp., 125 N.J. 421 (1991), and Landrigan v. Celotex Corp., 127 N.J. 404 (1992), the New Jersey Supreme Court departed from Frye in toxic-tort and asbestos litigation, and allowed scientific evidence that had not yet achieved general acceptance in its field "if, after the trial court engages in rigorous gatekeeping when reviewing for reliability, the proponent persuades the court of the soundness of the expert's reasoning and methodology." Accutane, 234 N.J. at 347. Accutane expanded this standard to all civil cases, summarized New Jersey's evidentiary jurisprudence, and described its parallel evolution to the reliability inquiry provided by the United States Supreme Court in Daubert, 509 U.S. 579, but "stopped short of declaring [New Jersey] a Daubert jurisdiction." Accutane, 234 N.J. at 399. The Daubert factors are helpful determiners of reliability and "should be incorporated," but New Jersey does not "embrace the full body of Daubert case law." Id. at 399-400.

> Distilled, the general factors identified as perhaps pertinent for consideration, but not dispositive or exhaustive, are:
> 1) Whether the scientific theory can be, or at any time has been, tested;
> 2) Whether the scientific theory has been subjected to peer review and publication, noting that publication is one form of peer review but is not a "sine qua non";
> 3) Whether there is any known or potential rate of error and whether there exist any standards for maintaining or controlling the technique's operation; and
> 4) Whether there does exist a general acceptance in the scientific community about the scientific theory.
>
> [Id. at 398.]

To further clarify, the Accutane Court stressed that:

> Our view of proper gatekeeping in a methodology-based approach to reliability for expert scientific testimony requires the proponent to demonstrate that the expert applies his or her *scientifically recognized methodology in the way that others in the field practice the methodology*. When a proponent does not demonstrate the soundness of a methodology, *both in terms of its approach to reasoning and to its use of data*, from the perspective of others within the relevant scientific community, the gatekeeper should exclude the proposed expert testimony on the basis that it is unreliable.
>
> [Id. at 399-400 (emphasis added).]

After years of a separate standard for criminal and civil matters, the New Jersey Supreme Court concluded, in 2023, that "Daubert's focus on methodology and reasoning, which we apply in civil cases, is a superior approach to criminal cases as well." Olenowski I, 253 N.J. at 151. At issue in that case was the admissibility of testimony by Drug Recognition Experts (DREs), trained police officers who use a 12-step method to determine whether a person is under the influence of drugs. Id. at 139. After a special master determined that DRE's methodology was admissible under the Frye general acceptance standard, the Olenowski I Court found that "special justification exists to replace Frye with a Daubert-type standard," in part because the Daubert standard specifically addressed known and potential rates of error, and Frye did not. Olenowski I, 253 N.J. at 143, 153. Once again, the Court stressed that "Daubert's non-exhaustive list of factors does not limit trial judges in their assessment of reliability. The same is true for caselaw from other jurisdictions, which can be persuasive but is not controlling." Olenowski I, 253 N.J. at 154. "The focus . . . belongs on the soundness of the methodology and reasoning used to validate the expert opinion or technique. . . . [and] applies not only to testimony based on scientific knowledge but also to testimony based on technical or other specialized knowledge." Id.

The "opinions in Accutane and Olenowski I both cautioned that the Daubert factors should not be applied rigidly." Olenowski II, 255 N.J. at 584. In fact, the sequence in which a court addresses the Daubert factors here does not reflect their relative importance, all of them bear upon the analysis. Ibid.

While STRmix itself has not been evaluated for reliability in New Jersey courts, other types of probabilistic DNA genotyping software have been considered under the Frye standard. State v. Rochat, 470 N.J. Super. 392 (App. Div. 2022), examined the reliability of a proprietary software used by the Office of the Chief Medical Examiner of the City of New York (OCME) under the Frye standard. The OCME developed a program called the Forensic Statistical Tool (FST) that, much like STRmix, analyzed DNA samples that contained biological material from more than one person and calculated a LR as to whether a particular individual contributed to the sample. Id. at 415. In that case, the trial court found FST to be "generally accepted in the relevant scientific community" and thus admissible under Frye. Rochat, 470 N.J. Super. at 432.

The Appellate Division reviewed the reliability determination *de novo* and reversed, finding that "[a]lthough likelihood ratios may be generally accepted in the relevant scientific community as a method of determining the probability that an individual is a contributor to a DNA sample, the fact that FST calculates likelihood ratios does not establish the reliability or general acceptance of the program *itself*." Id. at 440 (emphasis added). The program making calculations based on reliable scientific methods must reliably implement those methods, just as an expert witness must properly apply the standards of their field when giving testimony. FST's source code was validated internally by the OCME, but "internal validation of proprietary software cannot establish general acceptance" by itself, and approval by the New York State Commission on Forensic Science,

which accredited all laboratories in the state, was not enough to show a general acceptance in the international scientific community. Ibid.

The Appellate Division held that for the State to prove general acceptance it should have provided "testimony regarding six of the articles" that were cited, and the State should have "explained how the articles support the view that LCN DNA testing is generally." Id. at 446. The Appellate Division demanded more evidence supporting general acceptance by the State than just "one published article" and "New York lower court decisions and an appellate decision from Maryland." Id. at 449. The Appellate Division noted the evidence for general acceptance in Harvey was stronger, as "six independent laboratories had verified the accuracy" of the tests and "appellate courts in three other states and the Eight Circuit Court of Appeals found the polymarker test" to be "generally accepted in the scientific community." Ibid.

The interplay between the ultimate product of a PGS (the LR) and the software-based methodology that creates it was further addressed in Pickett, 466 N.J. Super. 270. The purpose of the opinion in Pickett was to determine the scope of the information that should be available to a defendant who wished to challenge the reliability of a proprietary software's code and what would be required for a judge to "reach an informed reliability determination." Id. at 278. The Appellate Division held that "[f]ull independent access [to a program's source code] in an adversarial system is a prerequisite to meaningful cross-examination of the State's expert." Id. at 280. The ability to access the source code, make an independent assessment, and identify potential shortcomings was not only "directly relevant to" the question of whether it has gained general acceptance, it was also an essential part of the fundamental and guaranteed "right to a fair trial . . . pursuant to the Fifth and Sixth Amendments of the United States Constitution, as well as the New Jersey Constitution." Id. at 278, 302.

It also does not grant an unlimited and unrestricted right of access, but rather one curtailed by "an appropriate protective order." Ibid. The aim, as stated supra, is "meaningful cross-examination" at a reliability hearing and not being forced to simply take the proponent's work for it. Ibid. A mere review of a program's outputs without 'looking under the hood,' so to speak, may fail to identify relevant information.

The Pickett court, while not finding any claims of faulty software were necessarily substantiated at that point, agreed that they were "valid concerns" and that "defendant articulated a particularized need" for the code. Id. at 324. A defendant's right to a complete defense required "meaningful opportunity" to prepare that defense and could not be ensured with "anything less" than access to the source code. Ibid.

Those same concerns, articulated by those same experts, have been raised here; the difference is that now the defendant has investigated them and must show more than just a rational basis for the concern. While the complexity of the calculations does not render probabilistic genotyping software of this kind unreliable *per se*, it must be examined closely and with "healthy skepticism." Id. at 317. "Even if the DNA science underpinning probabilistic genotyping analysis has been proven scientifically valid, computer software . . . must also properly implement that analysis in its source code; the source code must do as [its manufacturer] says it does." Ibid.

A defense expert's access to the propriety information of the source code is "directly relevant to" the question of whether it has gained general acceptance, as well as providing an opportunity for an "expert to independently test whether the evidentiary software operates as intended." Id. at 278.

This analysis will proceed in two parts: first, assessing whether STRmix is sufficiently reliable under the Daubert-like standard adopted for use in criminal matters by the New Jersey

Supreme Court in Olenowski I, 253 N.J. 133 (2023). Second, assessing whether the evidence and expert testimony regarding that methodology, as offered by the State, is otherwise admissible under N.J.R.E. 702.

i. Step One- Scientific Reliability

"In a case involving scientific evidence, *evidentiary reliability* will be based on *scientific validity.*" Daubert, 509 U.S. at 590 (emphasis in original). The New Jersey Supreme Court has identified four factors from Daubert and its subsequent caselaw which are roughly equal factors in a reliability determination. Olenowski II, 255 N.J. at 584. These factors are "non-exclusive" and "flexible," and "other considerations may also be pertinent." Id. at 583. While these factors may not be necessary considerations in *all* cases (for example, publication and peer review may not be a helpful factor in assessing a methodology employed by a field that does not customarily engage in peer-reviewed studies) they are relevant to the matter at hand. See Accutane, 234 N.J. at 398-99 (cautioning against a rigid application of irrelevant Daubert factors). The four factors, as reorganized in Olenowski II, are: "(A) adequacy of standards; (B) publication and peer review; (C) testability and error rate; and (D) general acceptance." Olenowski II, 255 N.J. at 585.

(A) Adequacy of Standards

The first prong, adequacy of standards, requires inquiry into "the existence and maintenance of standards controlling [the underlying methodology's] operation." Daubert, 509 U.S. at 594. In Olenowski II, the New Jersey Supreme Court found that the DRE protocols were based on adequate standards because they were "elaborate and standardized," had been used "for decades and . . . periodically modified," "adhere[d] to a standardized manual," and the experts who used the protocol were "extensively trained, and are supervised and recertified." Olenowski II, 255 N.J. at 587. These standards were considered reliable and not mere "idiosyncratic methodology." Id.

This adequate standards test does not apply solely to scientific fields: "there are many different kinds of experts, and many different kinds of expertise." Kumho Tire, 526 U.S. at 150. A standard is inadequate if it is improperly applied, or if its application "[falls] outside the range where experts might reasonably differ, and where the jury must decide among the conflicting views of different experts." Id. at 153. In Kumho Tire, an expert in tire failure was not unreliable because he lacked the necessary education or expertise in the field, but because the standards he employed—a four-part test he had developed himself—were not adequate for the matter at hand. The test was "administered in an undisciplined, standardless fashion" that could not satisfy *any* of the prongs examined today. Olenowski II, 255 N.J. at 586. At issue in Kumho Tire, and at issue now, "was not the reasonableness *in general* of [the approach]. Rather, it was the reasonableness of using such an approach, along with [an expert's] particular method of analyzing the data thereby obtained, to draw a conclusion regarding *the particular matter to which the expert testimony was directly relevant*." Kumho Tire, 526 U.S. at 153–54.

When a methodology is subject to and controlled by adequate standards, "potential differences of opinion do not necessarily make a . . . standard unsound. . . . There can be room for interpretation." Olenowski II, 255 N.J. at 588. The bar for the adequacy of standards prong is simply that: adequacy. Even if improvements to a given methodology are identifiable, or particular aspects are open to interpretation or discretion as was the case for the DRE protocol at issue in Olenowski, the question is whether "the standards are adequate to reasonably support admissibility." Id. at 589. Perceived shortcomings beyond that threshold are "a fair subject of . . .impeachment at trial." Id.

Choosing the appropriate standards is an important part of the Daubert analysis. It is not necessary or appropriate, for example, to establish the scientific validity of a photo retrieval database simply because it operates on a computer; it is only necessary for the officer to establish their understanding of how to operate it. See State v. Joseph, 426 N.J. Super. 204, 218-19 (App. Div. 2012). Similarly, it is most appropriate to consider STRmix according to the standards that are relevant to probabilistic genotyping software, and not necessarily helpful to apply parallel standards or more stringent ones. See, e.g., Lewis, 442 F. Supp. 3d at 1150–51.

(B) Publication and Peer Review

"[F]or scientific experts, 'submission to the scrutiny of the scientific community is a component of good science, in part because it increases the likelihood that substantive flaws in methodology will be detected.'" Olenowski II, 255 N.J. at 589 (quoting Daubert, 526 U.S. at 593). Publication and peer review, however, "is not a *sine qua non* of admissibility; it does not necessarily correlate with reliability, and in some instances well-grounded but innovative theories will not have been published." Daubert, 509 U.S. at 593. In fields where publication and peer review are commonplace, it is appropriate to consider "not only the existence of those studies but also their substantive content and conclusions." Olenowski II, 255 N.J. at 595. Laboratory studies and reviews by government agencies are also relevant; though they may not be prepared from the same perspective as scientific research, and are "not peer reviewed by academics, that does not undermine those studies' relevance or evidential weight." Id.

Where a theory or technology has been published in an appropriate journal, however, it is highly probative. "[P]ublication itself is noteworthy in scientific scholarship—and ultimately why publication in a peer-reviewed journal alone typically satisfies this Daubert inquiry." Gissantaner,

990 F.3d 464–65. The publication and peer review "factor does not demand independent authorship" as "peer review contains its own independence" due to it involving anonymous review of the scientific paper. Id. at 465. After "scientific research" has been "accepted for publication by a reputable journal following" rigorous peer review, that means there is "a significant indication that it is taken seriously by other scientists, i.e., that it meets at least the minimal criteria of good science." Ibid. (quoting Daubert v. Merrell Dow Pharm., 43 F.3d 1311, 1318 (9th Cir. 1995)). The importance of peer review is not that multiple journals have published articles on the scientific research in question, but that it has been accepted by a reputable journal following rigorous peer review. Gissantaner, 990 F.3d at 465

The peer review process is designed to ensure that studies "have passed a rigorous test and are generally considered worthy of consideration by the greater scientific community" before they are published. State v. Henderson, 208 N.J. 208, 242-43 (2011). The "key" to peer review is that the "theory and procedures have been submitted to the scrutiny of the scientific community." United States v. Bonds, 12 F.3d 540, 559 (6th Cir. 1993). When formal peer review is uncommon, as is the case with software code, any alternate method of scrutiny should be considered as well, as it at least "increases the likelihood that substantive flaws in methodology will be detected." Daubert, 509 U.S. at 593

(C) Testability and Error Rates

"Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested." Daubert, 509 U.S. at 593. "An untestable scientific theory is all theory and no science." Gissantaner, 990 F.3d at 463. Testability and error rates, while important to the case at hand, are not "categorically the most important Daubert factors." Olenowski II, 255 N.J. at 602. A "finding[]

of testability and reasonably low error rates from test results [is an] expected—but not always required—element[]of a proponent's reliability showing." Id. at 596. "Admissible evidence must consist of 'knowable fact[s]' relevant to the determination of the question before the trier of fact. If an expert's testimony conveys only a 'conclusion' or an 'assumption,' or if it is mere 'speculation or conjecture,' it is not factual and not helpful to the trier of fact." Id. (internal citations omitted). The Supreme Court of New Jersey has noted that the "fact that a possibility of error exists does not preclude a conclusion that a scientific device is reliable." Romano, 96 N.J. at 80. A failure to "proffer evidence regarding the error rate" of the technology at issue does "not render the expert's testimony unreliable." State v. Stokes, 478 N.J. Super. 392, 422 (Law Div. 2023) (citing State v. Matthews, 479 Md. 278, 1011-18 (2022)). The issue of error rates is "well-suited for 'vigorous cross-examination' and 'presentation of contrary evidence.'" Aviva Sports, Inc. v. Fingerhut Direct Mktg., 829 F. Supp. 2d 802, 830 (D. Minn. 2011) (quoting Daubert, 509 U.S. at 596).

In Olenowski, the DRE protocols examined were a mix of "'soft' sciences such as social sciences" and "'hard' sciences such as physical sciences." Olenowski II, 255 N.J. at 596. "Those and other inherent constraints ma[de] the DRE program less 'testable' and the error rate less 'knowable' than the ideal." Id. at 598. Nevertheless, DRE protocols were found sufficiently testable and admissible. A methodology does not need a 0% error rate to be reliable, and certain methodologies may pose practical barriers to testing. See id. at 602-03. "The absence of a definitive rate of error" will not always be "a dispositive basis to exclude" expert testimony. Id. at 603. It is not possible to prove a "null hypothesis" and guarantee that a process is error-free. Ibid. Rather, a holistic approach should be taken, where "hard" and "soft" science components of a complex methodology are tested, and error rates evaluated, according to the best practices for those relevant fields.

Relevant considerations in an analysis of error rates for a given methodology are its sensitivity, specificity, and accuracy. Olenowski II described how each of those terms applied to the DRE protocols before the Court.

> Sensitivity refers to the detection of true positives. In this context, it calculates the percentage of times a DRE correctly opined the presence of specific drug categories (under the certification match criteria described above) out of the total number of instances where the drivers had drugs in their systems. Mathematically, that entails dividing the number of true positives by the sum of true positives and false negatives.
>
> [Olenowski II, 255 N.J. at 599 (internal footnotes and citations omitted).]
>
> Specificity refers to the detection of true negatives. In this context, it means how often the DRE will opine that persons have no drugs in their system, if they indeed have no drugs in their systems. The specificity is calculated by dividing the number of true negatives by the sum of true negatives and false positives. The false positive rate shows in this case how often DREs opine that drivers have a drug or drugs in their system when, according to a toxicology report, they do not. It can be calculated by dividing the number of false positives by the sum of false positives and true negatives, or by subtracting the specificity rate from 100%.
>
> [Ibid.]
>
> Accuracy "summarizes the ability of the test being able to truly discriminate between true positives and true negatives." It considers when the subject condition is present and when it is not. One of the State's statistical experts, Dr. Brian D. Martin, testified that it is "the most commonly valued statistic associated with a test." As expressed in a mathematical formula, it means taking the sum of true negatives and true positives and dividing that figure by the sum of all four potential outcomes -- true positive, false positive, true negative, and false negative.
>
> [Id. at 600.]

(D) General Acceptance

While general acceptance is no longer the single dispositive factor that it was under the Frye standard, it is still relevant. See Olenowski II, 255 N.J. at 604 (reliability under Daubert permits, but does not require, a showing of general acceptance). Under the former standard, there were three main ways of demonstrating general acceptance:

> (1) by expert testimony as to the general acceptance, among those in the profession, of the premises on which the proffered expert witness based his or her analysis; (2) by authoritative scientific and legal writings indicating that the scientific community accepts the premises underlying the proffered testimony; and (3) by judicial opinions that indicate the expert's premises have gained general acceptance.
>
> [Kelly, 97 N.J. at 210.]

In Kelly, the proffered experts testified about battered-woman's syndrome, not in the field of DNA analysis. Ibid. At the time, battered-woman's syndrome was a relatively new field of research, but "numerous books, articles and papers" indicated "the presence of a growing field of study and research about the battered woman's syndrome and recognition of the syndrome in the scientific field." Id. at 211. Experts testifying to either "widespread acceptance" or "minimal support" can be an important consideration, Daubert, 509 U.S. at 594, but contradictory experts on both sides can quickly turn a hearing into one of credibility, not reliability. "The focus . . . must be solely on principles and methodology, not on the conclusions that [experts] generate." Id. at 595.

Authoritative writings, to the extent they differ from peer-reviewed publications already discussed, are similarly helpful but not dispositive. Judicial opinions are particularly useful when the methodology at issue is new or cutting-edge and of the type commonly used in forensic investigations—such as probabilistic genome sequencing. They are not, however, a substitute for actual testimony and analysis. Judges "are not scientists" and judicial opinions alone do not provide "the level of certainty necessary to approve" a particular methodology. State v. Doriguzzi, 334 N.J. Super. 530, 540 (App. Div. 2000). This does not mean that out-of-state opinions are of no value at all: Olenowski II was replete with analysis of out-of-state cases. Many of the experts who testified in the instant hearing are also featured in some of the major cases discussed infra, and it simply

does not follow that opinions on the same subject matter before the court, guided by testimony from the same experts offered in this hearing, have no ability to persuade.[10]

Regardless of which way general acceptance is shown, "[o]ur case law has instructed that there need not be complete agreement within the scientific community to satisfy the general acceptance test." Olenowski II, 255 N.J. at 605. "Proof of general acceptance does not mean that there must be complete agreement in the scientific community about the techniques, methodology, or procedures that underlie the scientific evidence." State v. Chun, 194 N.J. 54, 91–92 (2008). The Supreme Court has previously noted that "every new scientific discovery has its detractors and unbelievers, but neither unanimity of opinion nor universal infallibility is required for judicial acceptance of generally recognized matters." State v. Johnson, 42 N.J. 146, 171 (1964).

The proponent of the scientific evidence bears the burden to establish "that the device or the test meets the standard of general acceptance as we have defined it." Chun, 194 N.J. at 92. In Chun, the Court considered this test as it related to the Alcotest breathalyzing device's features and corresponding code, and found that "evaluation of the exhaustive record relating to the source code leaves us confident that its errors have been revealed," and the court did not agree with the concern that the source code is "likely to generate inaccurate results simply because, from a source code writer's viewpoint, it is complex or prolix." Id. at 131-32. Furthermore, in Chun, the Supreme Court declined to follow the "defendants' invitation to require that the firmware comply with any specific programming standards" as there was "no evidence in the record that these asserted shortcomings are anything more than stylistic, theoretical challenges." Ibid. The Supreme Court noted

[10] The defense has suggested that all the out-of-state opinions on STRmix lack the scrutiny required to be persuasive. The expert testimony, special master's reports, and multiple layers of appellate review that went into many of those opinions, however, demonstrate that those cases were "thoroughly and thoughtfully litigated." Doriguzzi, 334 N.J. Super at 546.

that defendant's expert opined that "the source code revealed thousands of programming errors," and after reviewing the record, the Court felt confident the errors found in the source code were not of the type likely to generate inaccurate results. Ibid.

ii. Step Two- The Requirements of Rule 702

Ultimately, it is not the factors found in Daubert or Olenowski II that determine whether the testimony offered here is admissible, it is New Jersey's rules of evidence. See Olenowski I, 253 N.J. at 154. In addition to analyzing the reliability of the methods underlying STRmix using the Olenowski factors, this court must also determine that the State's experts have reliably applied that methodology. Whether expert testimony is admissible is governed by N.J.R.E. 702, which states "[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify thereto in the form of an opinion or otherwise." "[U]nder our Rule 702 jurisprudence, trial judges still have a gatekeeping responsibility in ensuring that . . . expert witnesses demonstrate that they have reliably applied the[ir] methodology. . . ." Olenowski II, 255 N.J. at 616.

> The Rule imposes three basic requirements: (1) the intended testimony must concern a subject matter that is beyond the ken of the average juror; (2) the field testified to must be at a state of the art such that an expert's testimony could be sufficiently reliable; and (3) the witness must have sufficient expertise to offer the intended testimony.
>
> Landrigan, 127 N.J. at 413 (citing Kelly, 97 N.J. at 208).

These requirements "are construed liberally in light of Rule 702's tilt in favor of the admissibility of expert testimony." State v. Jenewicz, 193 N.J. 440, 454 (2008). "The standard for the admissibility of expert testimony is not whether it is unassailable and totally reliable, but whether it has a substantial degree of reliability and would be 'an aid to the court or jury in determining the question

in issue.'" State v. Wanczyk, 196 N.J. Super. 397, 402 (Law. Div. 1984) (quoting State v. Cavallo, 88 N.J. 508, 517 (1982)). In other words, is the testimony helpful? State v. Berry, 140 N.J. 280, 290 (1995).

1. Whether the evidence is beyond the ken of the average juror.

"Under [Rule 702], expert testimony is not appropriate to explain what a jury can understand by itself." J.L.G., 234 N.J. at 305. Under the Federal Rules of Evidence, which similarly permit evidence that will help a jury understand a fact, there is an "undeniable preference" for admitting expert testimony that may prove helpful. Planned Parenthood of Cent. New Jersey v. Verniero, 22 F. Supp. 2d 331, 338 (D.N.J. 1998). Here, too, "the 'helpfulness' standard of [F.R.E. 702] is imbedded in our own jurisprudence concerning admissibility of expert testimony." Berry, 140 N.J. at 290.

Expert testimony may be helpful if the expert is explaining a necessary concept which a juror may not be familiar with, such as how illicit drug transactions are usually conducted, as in Berry; it may not be helpful if the expert is simply providing an alternative theory of causation or offering a conclusion to the jury. See, e.g., State v. Free, 351 N.J. Super. 203 (2002) (testimony on "false confession syndrome" not helpful because a jury could understand the concept of a coerced confession without expert assistance). "Expert testimony is not necessary to tell the jury the 'obvious' or to resolve issues that the jury can figure out on its own." State v. Simms, 224 N.J. 393, 403 (2016). The test is not whether the issue the expert opines on is commonplace or rare, but whether they have "peculiar knowledge or experience not common to the world which renders their opinions founded on such knowledge or experience any aid to the court or jury." Berry, 140 N.J. at 291 (quoting Rempfer v. Deerfield Packing Corp., 4 N.J. 135, 142 (1950)).

2. Whether the evidence is sufficiently reliable and can be applied to the facts at issue.

The question of reliability when applied to the underlying methodology an expert relies upon is addressed by the Daubert factors, discussed supra. An expert's testimony must be based on a reliable and adequately founded methodology, and the expert must identify the basis for their conclusion and demonstrate it is supported by both reliable facts, and reliable methodology. See Bahrle v. Exxon Corp., 279 N.J. Super 5, 32-34 (App. Div. 1995). Each step in the analysis must be reliable and link the facts to the conclusion. Castro v. Sanofi Pasteur Inc., 134 F. Supp. 3d 820 (D.N.J. 2015). The question, therefore, is not only whether STRmix is reliable, but whether it has been used *reliably*—that is to say, for its intended purpose.

3. Whether the expert witnesses have sufficient expertise

Finally, an expert must be sufficiently qualified to give their opinion. "The court's function is to distinguish scientifically sound reasoning from that of the self-validating expert, who uses scientific terminology to present unsubstantiated personal beliefs." Landrigan, 127 N.J. at 414. "[T]rial courts take a liberal approach when assessing a person's qualifications." Jenewicz, 193 N.J. at 454. While "the strength of an individual's qualifications may be undermined through cross-examination" and the weight a jury gives their opinion may be affected, sparsity of experience "is not a sound basis for precluding an expert from testifying". See Id. at 455. This liberal approach in examining an expert's qualifications is consistent with the policy of allowing testimony that will assist the jury. See Berry, 140 N.J. at 290.

VIII. LEGAL ANALYSIS: OLENOWSKI FACTORS

a. Adequacy of Standards

More than one scientific community can be relevant when considering the reliability of a given methodology. Pickett, 466 N.J. Super. at 277. Here, as in Pickett, the scientific disciplines of forensic DNA analysis and computer science are both relevant to the inquiry.

Although the defendant has framed the issue to suggest that "software engineering" is the proper sub-field of computer science to draw standards from, this is too broad. "Software engineering is an engineering discipline that is concerned with all aspects of software production from the early stages of system specification through to maintaining the system after it has gone into use." Ian Sommerville, Software Engineering 21 (2016, 10th Ed.). As the defendant points out, this includes both "technical process[es]" and "product management." Ibid. Product management best practices, such as interviewing stakeholders to develop lists of desirable features or coordinating launch schedules, are not directly relevant. A functional and reliable website or piece of software can be coded in a college dormitory without a project management team or consultation with stakeholders. It may not be viable as a business or a product until it undergoes the more commercial or administrative aspects of software engineering, but history is full of examples of products that birthed a business—rather than the other way around.

Only the aspects of software engineering and development that provide standards for STRmix's operation and output are relevant to a determination of reliability, which is to say the principles of developing, testing, and evaluating the code itself.

"Software engineering is an engineering discipline that is concerned with all aspects of software production from the early stages of system specification through to maintaining the system after it has gone into use." Sommerville at 21. It is not software design, and it is not programming. It is a discipline that is focused on the design, testing, and implementation of software prod-

ucts in a cost-effective and timely manner. Id. at 19. "Software engineers are concerned with developing software products, that is, software that can be sold to a customer." Id. at 20. An individual may reliably program a piece of software without adhering to any of the principles of software engineering. That software could be testable and found reliable without a requirements document. It could subsequently be sold to a larger company and undergo the software engineering process after the fact, to ensure scalability, marketability, and profitability. During that process, it is possible that they would not change the source code at all, because it was found to be both reliable and suitable for its new purpose.

> There are no universal notations, methods, or techniques for software engineering because different types of software require different approaches. Developing an organizational information system is completely different from developing a controller for a scientific instrument. Neither of these systems has much in common with a graphics-intensive computer game. All of these applications need software engineering; they do not all need the same software engineering methods and techniques.
>
> [Id. at 18]

In short, software engineering spans disciplines, and the tools it employs varies according to the areas where it is implemented. In some cases, a "formal approach simplifies the production of a safety or security case. This demonstrates to customers or regulators that the system actually meets its safety or security requirements. However, because of the high costs of developing a formal specification, this development model is rarely used except for critical systems engineering." Id. at 49. In other cases, software may be developed incrementally, which is less formal, less costly, and more agile. It is more common than formal development and "the case for most business systems and software products." Id. at 50.

> Incremental development reflects the way that we solve problems. We rarely work out a complete problem solution in advance but move toward a solution in a series of steps, backtracking when we realize that we have made a mistake. By developing

> the software incrementally, it is cheaper and easier to make changes in the software as it is being developed.
>
> [Ibid.]

This description of the incremental process fits better with the way STRmix was actually developed: a problem was identified, a rough solution was designed, and then that solution was reviewed and refined until the end product was ready for production. STRmix should be judged by this software standard unless it is required to conform to the higher bar of the formal development process, which would require evidence that it is, or must be treated as, a critical system.

Several experts have testified to their opinion that STRmix should be a safety critical system and should be subject to formal requirements both pre- and post-development. (16T 36-37). STRmix's creator, Dr. Buckleton, has also testified that he thinks STRmix could be treated as a safety-critical system. (6T 117-12). There is no consensus between the defendant and the prosecution as to which level of safety critical system STRmix would be if it were ever considered such: the defendant believes it is an integrity level 4 system, while Dr. Buckleton believes it should be somewhere between integrity levels 3 and 4. (16T 47-8); (6T 117-12).

To be clear, Dr. Buckleton does not believe STRmix *is* a level three or a level four critical system, he only says he is willing to put in the work to meet those requirements. The defendant has argued that "coming close" to these requirements is insufficient, and that "demonstrated conformance to verification and validation standards for safety-critical systems is the only way to demonstrate the reliability of that system." Def. br. at 49. There is a vast difference between believing a piece of software "should be considered" safety critical and a piece of software actually being a safety critical system. No expert has shown—nor is there any evidence to support the claim—that STRmix has ever been subject to regulation as a safety-critical system or any standards that require its treatment as such.

The court therefore finds that it is better to analyze STRmix using the software standards and practices that focus on testing software and assessing its actual reliability, rather than focusing on the formality of its construction. Dr. Coble described such practices. (10T 94-13). He also articulated these standards in publication. Coble et al, Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, 25 Forensic Science International: Genetics 191 (2016). This is not a binding document promulgated by a regulatory agency. It is, however, a statement made by qualified members of the Forensic DNA community articulating their understanding of the software community's accepted practices. They distinguish between the responsibilities of software developers and incorporate, by reference, relevant portions of IEEE 1012. Id. at 192.

In summation, the court finds that the software community's standards vary depending on which community the users of that software are associated. The relevant software standards for PGS, as distilled and implemented in the DNA community, include ensuring the underlying method implemented by the software is valid, disclosed, and reproducible outside the software; verification and validation data is made available; testing and documentation is provided for each version of the software; and bugs, error fixes, and patches are described and disclosed. See generally Coble et al, 192-94 (proposing these as standard practices as well as others related to training, user manual documentation, and recommendations for laboratory internal validations). "Validation is the essence behind the reliable analytical data required by the Frye and Daubert rulings that impact court admissibility of scientific results." Butler, Debunking Some Urban Legends Surrounding Validation Within the Forensic DNA community, *Profiles in DNA*, Sept. 2006, at 3. STRmix, as described before, is a probabilistic genotyping software. PGS analyzes DNA samples

with genetic material from multiple contributors, calculates possible combinations that would result in the sample given, and generates a corresponding LR as to whether any particular individual's DNA is present in the mixture. "A higher LR is typically obtained when evaluating a person of interest (POI) who is a true contributor to the evidence profile, and a lower LR is typically obtained when the POI is not a true contributor." SWGDAM Guidelines at 2. It "is not intended to replace the human evaluation of the forensic DNA typing results or the human review of the output prior to reporting." Ibid. Rather, it is an interpretive tool which requires human supervision.

Probabilistic genotyping software requires validation to be reliable. As Ms. Naughton testified, internal validation allows for the software to be implemented in the laboratory, so these validations are a requirement. (2T 99-11). Validation, in the context of Forensic DNA analysis, is "the acquisition of test data to verify the functionality of the system, the accuracy of statistical calculations and other results, the appropriateness of analytical and statistical parameters, and the determination of limitations" Id. at 5. At issue in this case is the validation of STRmix as a whole, as well as the sufficiency of the validation processes used by the Bode laboratory and the New Jersey State Police laboratory, and the standards that govern both validation processes. In addition to formalized standards, this court will look to "recognized professional societies [which] may have positions that help determine the soundness of the [methodology]." Landrigan, 127 N.J. at 417.

The testifying experts identified several sources that provide recommendations, guidelines, and standards for implementing PGS in a laboratory setting, and the distinction between those terms. Standards are the practices that "must" be followed. (6T 73-8). Guidelines are generally not mandatory practices, though some of them are "rightly [considered] an expectation of the justice

system" and treated as such. (6T 55-6). Several guidelines and standards were identified over the course of the hearing:

Whether they are described as standards, guidelines, general suggestions, or best practices, some of these documents have "teeth." NIST used the same term when summarizing the forensic DNA community's desire for standards with "impact or real influence." Mixture Interpretation at 28. The court finds the general community's desire for influential standards to be highly relevant when determining which of the above documents are the relevant standards of the community, regardless of what they are titled. The ANSI/ASB Standard 18, SWGDAM guidelines, and FBI QA standards are all relevant. The ISFG recommendations do not have "teeth," but are relevant to the extent they reflect the PGS community's willingness to engage with their peers in the Computer Sciences. Other standards have limited relevance, but are still instructive:

1. ANSI/ASB 018.

While there are several guidelines that apply to PGS validation and its use in casework, there are relatively few binding standards. One such standard is the ANSI/ASB Standard 018: Standard for the Validation of Probabilistic Genotyping Systems (2020). Also referred to as ANAB, or the ANSI National Accreditation Board. Standard 018, as its name suggests, directly addresses the validation process for PGS in laboratories. It contains both mandatory and recommended provisions. Standard 018 at iii. It requires both developmental and internal validation, and although the defendant claims it is not a reliable standard because it does not specify a minimum threshold of results, Def. Br. at 105, the responsibility for determining the sufficiency of test results lies with a laboratory's technical leader. This has been found to be a standard practice in the scientific community. Both Bode and the New Jersey State Police are accredited to Standard 018. (2T 66-21; 8T 18-18).

ANSI/ASB's internal validation standards govern the laboratory-specific implementation of STRmix and require laboratories to use "case-type" profiles that represent the plausible range of casework samples the laboratory intends to analyze. The only hard limitation placed on internal validation studies is that they cannot exceed the scope of the developmental validation. If a laboratory wishes to start analyzing samples beyond those contemplated by the developer, additional developmental validation studies must be conducted. Id. at 3. This standard requires testing with ground truth samples.

2. SWGDAM Guidelines

Another standard applicable to both STRmix as a PGS and to the laboratories that implement it are the SWGDAM guidelines. The Scientific Working Group on DNA Analysis Methods, Guidelines for the Validation of Probabilistic Genotyping Systems (June 2015). The SWGDAM guidelines directly address PGS. While they are published as guidelines, they are "de facto standards" and adhering to them is required for federal funding and access to CODIS. (19T 118-5).

Compliance with these guidelines is also a prerequisite to testifying to results in court. (6T 55-2). The SWGDAM guidelines are thus the most common guidelines that laboratories will use when going through the accreditation process. (6T 70-14). SWGDAM is a consortium of approximately 50 scientists that represent state and local forensic DNA laboratories in the United States and Canada. SWGDAM's guidelines provide that developmental validation may be conducted by either the manufacturer of the application or the testing laboratory, and that developmental validation demonstrates any "known or potential limitations" of the system and further provides a number of steps in the validation process. A laboratory's internal validation, meanwhile, determines whether a PGS is suitable for use in their system. This is similar to the distinctions drawn in the ANSI/ASB standard, but the SWGDAM guidelines acknowledge that some of the work done for

a developmental validation can apply to a laboratory's internal validation, if conducted in the same facility.

While the SWGDAM guidelines are not strictly binding on every DNA laboratory in the country, the FBI does require compliance with them as a precondition to accessing CODIS. As such, they can be considered as guidelines "with teeth." Bode and the New Jersey State Police laboratory both comply with the SWGDAM guidelines and are subject to regular audits confirming that fact. Both Kristin Naughton and Jennifer Thayer testified to that effect. Defense expert, Mr. Inman, testified that SWGDAM guidelines are mandatory "if you want federal funding" or "to put samples into CODIS," so his opinion is that they should be treated as mandatory. (19T 118-3). But he did include that looking at the SWGDAM guidelines as mandatory is a "[t]ough question." (19T 118-3).

3. FBI QAS Standards

FBI, Quality Assurance Standards for Forensic DNA Testing Laboratories (2011). The FBI Quality Assurance Standards are not specific to PGS, but a laboratory's use of PGS as a part of their overall DNA analysis workflow is governed by these standards. FBI at 18.

4. IEEE 1012

The IEEE 1012-2016 standard for software engineering has been suggested to be a relevant standard for STRmix. Both the creators of STRmix and the experts proffered by the defendant are of the opinion that STRmix should be held to IEEE 1012's standards. IEEE 1012, however, is not a binding standard and does not purport to be. Defense expert, Dr. Martin, testified that IEEE 1012 is not a mandatory standard for a program like STRmix. (18T 79-15). While IEEE 1012 may be instructive and may represent a "gold standard" for software engineering that important programs like STRmix should aspire to, there is no credible evidence in the record that STRmix is a safety

critical system that is required to adhere to IEEE 1012. Dr. Buckleton testified that he would be willing to treat STRmix as a Safety Critical System, but did not agree that this is a level 4 integrity system. (6T 117-12).

Some systems are referred to as Safety Critical Systems. Sommerville at 341. These systems are when "it is essential that system operation is always safe." Ibid. These systems should "never damage people or the system's environment." Ibid. Examples of these systems are "control and monitoring systems in aircraft, process control systems in chemical and pharmaceutical plants, and automobile control systems." Ibid. It has been noted that "[r]elatively few people worldwide have been killed or injured because of fault software." Id. at 343. Even for these systems where it is essential to always be safe, it is "impossible to make a system 100% safe." Id. at 344. This level of systems requirement is not the standard for all software and only applies for safety critical systems.

While the evidence generated by STRmix potentially can impact the life and liberty of a criminal defendant, the difference between STRmix and other safety-critical systems is the immediacy of harm. If a safety-critical system fails, human intervention may not be able to correct the failure, whether the harm suffered is immediate or delayed. A brake failure will cause an immediate crash. A radiation overdose will take time to make itself known, but the harm became inevitable at the point of failure. But STRmix does not convict a defendant. Mr. Adams testified that a wrongful conviction, and a loss of liberty, could be characterized as equivalent to a loss of life. (16T 47-48). A conviction and a radiation overdose, however, are not the same. STRmix is an investigative tool, and it interprets evidence. That interpretation is then presented to a jury through the testimony of an expert who explains its context and is subjected to cross examination, and the jury then deliberates upon it. STRmix must be reliable to be trusted, yes, but that reliability is determined

through the judicial process, not by specifically prescribed software and project management workflows.

Olenowski II addressed a similar argument. There, the defense counsel and amici argued that DRE experts were not medical professionals and could not reliably issue a diagnosis, and therefore could not administer the DRE protocol. But there, as here, that was a higher standard than appropriate. Olenowski II, 255 N.J. at 588. A DRE expert can reliably conduct certain medical-related tasks without a medical license, and STRmix can be reliably engineered without complying with every part of IEEE 1012. There is sufficient credible testimony to find that STRmix complies with all the relevant standards and guidelines for PGS, and makes an admirable effort to go further and address the standards of IEEE 1012 and other related industries. Dr. Martin testified that he thought multiple versions of the source code complied with engineering rules, were written reasonably well, and were professional. (18T 66-8).

There is no reason to declare STRmix a safety-critical system today. The risk to life and liberty inherent in an inaccurate DNA result is not of the same nature as the risk of death, injury, or economic loss attributable to safety-critical systems: if an autopilot or a nuclear reactor fails, the risk is immediate and may be beyond the reach of human intervention to correct. In short, the harm occurs at the time of failure. Our system of justice, however, has safeguards built into it; no machine, software, or piece of evidence can directly deprive a defendant of their liberty, the adversarial system and a jury of one's peers stand guard to evaluate evidence. Furthermore, limiting instructions to the jury could be used to assist jurors with determining how to analyze these results.

ESR may consider it a best practice to hold their software to the higher standards of a safety-critical system to ensure confidence in their results, but setting a higher target for oneself

does not change the threshold of reliability in a court proceeding. The court does not find that STRmix is a safety-critical system that requires a heightened standard of reliability.

5. Independent Verification and Validation requirements

The verification requirements that the defendant asks this court to require are not the test-ability requirements contemplated by Olenowski II and Daubert. The concern is not how STRmix was invented, the concern is how it works. The Olenowski court, in discussing the "tire-failure expert" of Kumho Tire, 526 U.S. 137, made this distinction clear. The problem with the Kumho tire test was not its origin or its development, or even the conclusions the expert drew from the test, but that its implementation was "undisciplined" and "standardless" Olenowski II, 255 N.J. at 586.

There is also nothing in the record to indicate that STRmix is required to undergo Independent Verification and Validation. Verification is not a standard of reliability. It is a way of determining that the product built is the one that the customer asked for. An insufficiently detailed requirements document might make it more difficult to formally verify a program, but verification has nothing to do with the reliable functionality of a program. As discussed supra, STRmix was developed via an ad-hoc process; it addressed a pressing need in the Australasian forensic community, and ESR did not contemplate developing a commercial product at first. (6T 18-3). Asking ESR to go back in time and employ a different development process just to check off a box on an inapplicable standard is uninformative. This argument, put forth by all the defendant's software engineers, is where they fail to understand STRmix.

Software verification is "the process of checking that the software meets its stated functional and non-functional requirements." Sommerville at 228. Validation is a different process. Sommerville at 228. The purpose of validation is to "ensure that the software meets the customer's

expectations" and that the developers built the right product. IEEE at 228. Validation should be done to show that "all stakeholders are satisfied by the set of systems requirements." IEEE at 65. Defendant argues the stakeholders include criminal defendants, prosecutors, the criminal justice system, and the public. But as Dr. Buckleton testified, STRmix was originally developed just to be used in Australia and New Zealand, and the original intent behind developing this product was not as a commercial venture. (6T 50-10).

The Defendant's experts envision a typical scenario where a company has an idea, identifies a market for that idea, designs a product, and hires a software developer to write the code. This, obviously, requires the creators of the code to understand what it is that the company wants the software to do, and the company to understand what their customers want in turn. When Drs. Buckleton and Taylor created STRmix, they knew what the problem was, and they knew how to create the software. While they were the developers, they were also the customers and stakeholders. Their experience included statistics, software engineering, and forensic DNA analysis. Dr. Heimdahl acknowledged this on cross-examination: when he was asked if it was "plausible to mandate someone . . . that wants to solve a problem in Australia . . . come to the United States and consult all relevant stakeholders?" His answer was "No." (T15-21-3 to 21-8). Sitting down with all possible stakeholders to discuss their needs and concerns is desirable when creating any kind of software, but it is not always a viable option. It makes good business sense to do so, but it has no bearing on scientific reliability. It is similarly unrealistic to criticize STRmix's origins because it did not follow a "waterfall" method of development, where each step follows another in a rigid process. Sommerville at 47, fig. 2.1. A perceived failure to consult stakeholders prior to creating this software is not significant to this court, because doing so would have been impossible.

There is more than one way to verify and validate a system, and verification is not a required step for all types of software. Since verification requires certain steps to be implemented at the outset of a project, there are times when strict adherence to verification procedures is impossible, but this does not render software inherently unreliable. Verification is also not necessary for reliability, though the process helps demonstrate validity. Verification "provides objective evidence" of development practices and whether a software moved through the phases of the development lifecycle "correctly" IEEE 1012-2016 at 16.

A product built on an assembly line must move through all the assembly stations in the correct order to satisfy validation and generate a document that it was built "correctly." Following the steps in order, however, does not tell anyone how well the steps were followed. Validation is the stage where the actual quality of a product is assessed. "The Verification Process and the Validation Process are interrelated and complementary" and verification "may be useful in performing the developer's tests and evaluations," but it is not the tests themselves. Ibid. IEEE 1012's verification and validation recommendations are also "wholly voluntary" for the forensic community. Id. at 3. Although defendant's experts described verification as a binary process, one that does not give "partial credit" (17T 26-9), the IEEE standard and other evidence all suggest that the requirements for V&V exists on a spectrum.

The inquiry is also not whether STRmix has been validated by a sufficiently independent organization, it is whether it has been *validated*, and whether it has been validated *reliably.* Independent validation is a practice employed at the highest level of safety-critical systems. Even if STRmix were found to be safety-critical, there is no consensus that they would require this level. Dr. Buckleton testified that even though STRmix is not a safety-critical system, it does maintain technical, managerial, and financial independence. (7T 29-30).

6. Forensic Science Regulator (FSR)

The Forensic Science Regulator of the United Kingdom has published their own guidance, but it is not a relevant standard in the United States. Forensic Science Regulator, Forensic Science Regulator Guidance, FSR-G-223 (2020). (16T 161-18). To the extent that the FSR's views align with standards in the United States, it provides additional support for the finding that those common standards are relevant. The United Kingdom views validation, and validity more generally as

> the process of providing objective evidence that a method, process or device is fit for the specific purpose intended, i.e. can be relied upon. The Criminal Practice Directions suggest that the court takes into account when determining the reliability of expert opinion . . . the extent and quality of the data on which the expert's opinion is based, and the validity of the methods by which they were obtained.
>
> FSR-G-223, at 1.1.1

Though promulgated by a United Kingdom government official and not binding on United States laboratories, the FSR guidelines contain useful information for laboratories implementing STRmix. "Software Validation for DNA Mixture Interpretation" has more detailed requirements then SWGDAM for validation. S-132. Dr. Buckleton said the same during this hearing, discussing both S-132 and S-159, two FSR guidance documents; the first being for interpretation of DNA evidence and the latter being a guideline for software validation.

7. ISFG Recommendations

The ISFG Recommendations, discussed supra, are not standards. Coble et al, Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, 25 Forensic Science International: Genetics 191 (2016). They are, however, consistent with the guidance found within other standards, and provide reassurance that the forensic community follows accepted practices in the software community.

The International Society for Forensic Genetics (ISFG) published these guidelines in 2016 for the validation of PGS. ISFG Guidance proposes minimum requirements for validation and addresses developmental and internal validation. Dr. Coble, who chaired the DNA Commission that authored this document, discussed how this guideline addresses software engineering and source code review. (T10 95-22 to 96-3). He testified that, in addition to recommendations addressing testability, documentation, and version control for PGS, the commission included recommendations for testing and code review. Recommendation 7 states that "the DNA Commission does not consider examination of the source code to be a useful fact-finding measure in a legal setting." See Coble et al., at 194; (T10-96-25 to 97-3). That Recommendation goes on to state that "[a] rigorous validation study (both developmental and internal) should be sufficient to reveal shortcomings or errors in coding… however, if requested by the legal system, the code should be made available subject to the software provider's legitimate copyright or commercial interests being safeguarded." Coble et al., at 194.

8. ISO 17025

ISO 17025 is a general standard mentioned by both Ms. Ghannam and Ms. Thayer as being relevant in their laboratories. ISO/IEC International Standard 17025, General requirements for the competence of testing and calibration laboratories (2017) (1T 26-3; 8T 19-3). As its full title suggests, it is not focused on PGS. Rather, it details best practices for accumulating and storing relevant data in preparation for later audits. Finding that a laboratory complies with this standard may offer reassurance that they are generally organized and scientific in their operations but does not go to STRmix's reliability.

9. The PCAST Report

Published in 2016, and relying on data from 2014, the PCAST report is neither a standard nor a guideline. It is a report to the President of the United States laying out the observations and recommendations of a committee of forensic scientists. The document, which spans 174 pages, devotes 5 pages to PGS. It mentions that the FBI laboratory began using STRmix in 2015 but had not yet completed its internal validation. PCAST Report at 79. The document as a whole is a survey of the current state of the art across the forensic sciences, and its findings and recommendations are not binding—they are meant to address "whether there are additional steps on the scientific side . . . that could help ensure the validity of forensic evidence used in the Nation's legal system." Id. at *x*.

The PCAST report is a useful document. It expresses the opinions of six academic and industry leaders, including two processors of computer science, as to the state of the art and the general acceptance of PG in 2016. Id. at *vii*. These experts were advised on the interplay between science and the law by six Federal Judges, two law professors, and two statisticians. Id. at viii. While not an academic journal and not an in-depth discussion, it is best thought of as the summary findings of an independent review conducted by a group of peers.[11]

b. Publication and Peer Review

Publication, as stated supra, is not the "sine qua non" of scientific validity. Daubert, 509 U.S. at 593. For scientific fields such as statistics and computer science, however, it is extremely

[11] The report's discussion amounts to a preliminary conclusion that PG is foundationally valid but requires further analysis. It does not propose standards or outline steps that must be taken to determine validity. It finds PG to be "new and promising" and that "evidence supports the foundational validity" of PG in a limited capacity that is "likely to grow." PCAST Report at 82. The 31 Labs report, discussed infra, was written as a direct response to PCAST's desire for more information.

important. The purpose of engaging in peer review publication, per Dr. Coble, is that it allows reviewers to give an anonymous "critical evaluation of the paper." (11T 27-3). These reviewers will not know "any information about the authors, their affiliations and so forth, so the reviewer will get just a paper that has no information about who it's from." (11T 27-3). Peer review allows for anonymous feedback and criticism without fear of interpersonal conflict or offense, so that when the paper is published there is some "confidence that other people have looked at this first" and approved of the findings. (11T. 27-13).

In PCAST's 2016 report, it states that future studies should be performed by or should include independent researchers not connected with the developers, but this is not the standard. PCAST would just like to see independent authors be at least coauthored with the developers. 31 Labs at 23. Dr. Coble testified that he does not give great weight to the critique that the developer should not be a part of a publication for a software. (10T 71-1). Developers, and researchers generally, have the time and ability to take weeks to review the software and write publications. (10T 71-9). Independent authors might be involved with casework in which they do not have the same opportunity to perform studies. (10T 71-12).

The peer review process is blind. It consists of an anonymous critical evaluation of a submission. The "key" for peer review is whether "the theory and procedures have been submitted to the scrutiny of the scientific community." Gissantaner, 990 F.3d at 464. The independence and integrity of the reviewers is the crucial aspect of the peer review process, because they decide whether to approve or reject a paper for publication. Published papers authored by the developers of a method are subject to the blind peer review process, so the theory and procedures within the paper will be scrutinized and accepted.

This court takes notice that there have been over one hundred articles[12] published in peer reviewed journals that relate to some aspect of STRmix. Ninety-five of those have been authored by Drs. Buckleton, Bright, or Taylor, either individually or in some combination with each other. (S-186 at 3). Seventeen articles include none of the principal architects of STRmix. Twelve more articles address other topics in forensic science but include data generated with STRmix.

Articles written without crediting STRmix's developers include:

- Alladio, et al., DNA mixtures interpretation; A proof-of-concept multi-software comparison highlighting different probabilistic methods; performances on challenging samples, 37 Forensic Science International: Genetics 143 (2018).
- Noël, et al., 41 STRmix put to the test: 300,000 non-contributor profiles compared to four-contributor DNA mixtures and the impact of replicates, Forensic Science International: Genetics 24 (2019).
- Myers and Duke, Systematic evaluation of STRmix; performance on degraded DNA profile data, 44 Forensic Science International: Genetics 102174 (2020)
- Riman, et al., Exploring DNA interpretation software using the PROVEDIt dataset, 7(1) Forensic Science International: Genetics Supplement Series. 724 (2019).
- Huffman, et al., Probabilistic genotyping of single cell replicates from complex DNA mixtures recovers higher contributor LRs than standard analysis. 62(2) Science & Justice 156 (2022).

[12] Per State's exhibit S-186, as of October 25, 2023, there were approximately 120 articles in peer-reviewed journals that focused on STRmix's underlying model, its application, the use of LRs, software validations, or otherwise.

- Mallinder, et al., Interpretation and reporting of mixed DNA profiles by seven forensic laboratories in the UK and Ireland, 58 Forensic Science International: Genetics 102674 (2022)
- Riman, et al., Examining performance and likelihood ratios for two likelihood ratio systems using the PROVEDIt dataset, 16 PLoS ONE 0256714 (2021).
- Costa et al., Quantification of forensic genetic evidence: Comparison of results obtained by qualitative and quantitative software for real casework samples. 59 Forensic Science International: Genetics 102715 (2022).
- Costa, et al., Statistical analysis tools of mixture DNA samples: When the same software provides different results. 8 Forensic Science International: Genetics Supplement Series 37 (2022).
- Duke, et al., Compound and Conditioned Likelihood Ratio Behavior within a Probabilistic Genotyping Context, 13(11) Genes 2031 (2022).
- Orozco, TrueAllele and STRmix: A Comparison of Two Probabilistic Genotyping Software Programs in Forensic DNA Profile Analysis, *University of California, Davis* (2023).
- Greenspoon, et al., A tale of two PG systems: A comparison of the two most widely used continuous probabilistic genotyping systems in the United States, 69(5) Journal of Forensic Sciences 1840 (2024).
- Boodoosingh, et al., An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories, 71 Forensic Science International: Genetics 103046 (2024).

- McCarthy-Allen et al., 'Low' LRs obtained from DNA mixtures: On calibration and discrimination performance of probabilistic genotyping software, 73 Forensic Science International: Genetics 103099 (2024)

In Gissantaner, the court found STRmix had been the subject of "more than 50 published peer-reviewed articles" and that the two published validation studies addressing its reliability as-applied were, taken together, sufficient to clear the peer review hurdle. 990 F.3d at 465, 467. Dr. Buckleton confirmed in testimony that these publications are independent from any STRmix developer. (7T 60:11). Many of these articles were not discussed by the testifying experts or in either party's brief, except to make note of their existence. Judges are "not scientists" and an independent literature review of every article would be inappropriate and runs the risk of misleading the court. Doriguzzi, 334 N.J. Super. at 540. The court takes notice of the number of articles listed, and that the majority of them were written with the input of at least one of STRmix's creators. That said, the court has reviewed the publications, and the testimony discussing the same, and finds them relevant and instructive.

The 31 Labs article is one such study. Written as a response to the PCAST report's specific request for more information on certain tests, the 31 Labs report is a thorough and focused exploration of STRmix's capabilities and limitations. It not only addresses the effects of allele sharing on estimating NOC, but it also specifically uses related contributors to do so. See, e.g. 31 Labs at 12, 16. It corroborates Dr. Buckleton's point about transportability of data. (7T 94-1). The article explains, while addressing PCAST's call for more validation data, that "[e]ach different combination of genotypes is a unique contributor combination." 31 Labs at 13. This study represents a compiled testing of "approximately 20 million false contributors." 31 Labs at 15.

The Boodoosingh, et al, article, An inter-laboratory comparison of probabilistic genotyping parameters and evaluation of performance on DNA mixtures from different laboratories, also illustrates how the repeated validation of similar samples across laboratories creates an opportunity for peer-reviewed articles to test aspects of STRmix: by taking the interpreted data from one laboratory and the parameters from another, a third-party can independently test the outputs of STRmix and plot their consistency. (7T 93-9). This is further support for the argument that STRmix has been horizontally validated across hundreds of laboratories worldwide. While one laboratory's internal validation study cannot stand in for another's, the aggregated data from all the internal validation studies conducted allows for more thorough evaluation of STRmix. (7T 94-1).

Throughout this hearing, there has been an argument that many articles are not independent due to the developers of STRmix being listed as authors for many of the publications. This challenge has been raised and reviewed by other courts. Lewis, 442 F. Supp.3d at 1155. This court takes notice that in Lewis, the Special Master report explains that although some "authors are well known to have an interest in the outcome, the process of peer review acts as a check and balance on that interest." Id. The authors being developers does not undermine STRmix's "general acceptance by the scientific community or its reliability as a principle and method." Id.[13] The purpose of the peer review factor is to demonstrate that the theory and procedures have been submitted to the scrutiny of the scientific community, and every article satisfies that factor. Gissantaner, 990 F.3d at 464.

[13] In Lewis, Dr. John Buckleton testified regarding STRmix and supplied a list of 7 peer reviewed articles regarding probabilistic genotyping and STRmix. Most of these articles described STRmix as reliable. Lewis, 442 F. Supp. 3d at 1155.

Formal, published peer review for software is uncommon. But the Olenowski factors are not only non-exhaustive, but they also "do not all necessarily apply even in every instance in which the reliability of scientific testimony is challenged. . . . And case law has been mindful that several non-scientific fields of expertise are not typically studied in peer reviewed academic journals." Olenowski II, 255 N.J. 589–90.

Though peer review is not commonplace in the computer science community, independent testing and adversarial source code review serve analogous purposes. As part of the validation process, software team members will perform inspections which are "peer reviews where team members collaborate to find bugs in the program that is being developed." Sommerville at 713. Software like STRmix is constantly being reviewed and checked as it is being developed. Dr. Buckleton himself will test code, which he did not write, to ensure it meets quality standards. (6T 47-4).

Dr. Martin's review of STRmix's source code was also a form of peer-review. He has performed many such reviews, and he describes them as a quality assessment of the source code. (18T 16-5). Through Dr. Martin's analysis, source code is scrutinized by the software community. For STRmix, he reviewed the source code and found it to be "quite good," in compliance with "most reasonable engineering standards," and to be "professionally developed." (18T 66-6).

No New Jersey court has held that adversarial source code review was meant to be a substitute for IV&V, nor has the defendant requested IV&V and been told to settle for source code review as a compromise. However, source code review—even adversarial code review—is a type of peer review. It is a fair opportunity for a peer (who is not tainted by a motivation to blindly agree with the creator) to apply another source of independent rigor. The purpose of source code review was established in Pickett, and the purpose was to give the defendant a fair opportunity to

investigate their claim and "present a complete defense, by meaningful cross-examination at the appropriate juncture." Pickett, 466 N.J. Super. at 303.

The court finds that standard debugging and validation by groups other than the programmers, and even adversarial source code review, can provide the crucible for testing software that publication in peer-reviewed journals provides for other disciplines.

Surviving a source code review will not conclusively demonstrate reliability, it only shows that no affirmative proof of unreliability was found. Source code review in this case was consistent with the principles in Pickett and was an opportunity for the defendant to investigate and gather facts to support their claim that STRmix's code was faulty. The defendant has conceded that he found no bugs in v. 2.5.11 or v. 2.8.0 of STRmix's code. Def. Br. at 50. Having found none, there is only speculation on the part of the defendant that bugs exist, but the only consequence is that he needs to support his arguments another way.

The defendant not only failed to find flaws in their code review, but Dr. Martin also found the code to be "quite good" and in compliance "with most reasonable engineering standards.". (18T 66-6). He does claim that he was not able to review the code to his full satisfaction; the record is clear, however, that the roadblocks encountered by Mr. Adams and Dr. Martin were partly of their own making. Dr. Martin was not prevented from dynamically reviewing the code, it was simply outside the scope of his employment. Mr. Adams, meanwhile, chose not to reach out and request a solution to any problems he encountered. Rather, he waited for months and now claims that the review was incomplete. Complete or incomplete, they did not find what they were looking for.

Whether or not it was likely to find bugs during a source code review has no bearing on the reliability of STRmix, and the confident assessment of Dr. Heimdahl that bugs exist and were

simply difficult to find is not an opinion supported by any discernable fact or evidence. The defendant has suggested that "[n]either Mr. Adams nor Dr. Martin was able to debug the code, 'step through the code,' or perform a dynamic inspection of the code, tasks that would have been very helpful in assessing its reliability." Their inability to do so, if any, rests entirely on them. Mr. Adams was given three days for his review but "didn't follow the instructions and didn't appeal for help to the people who were standing by to give him help." (7T 133-7). He performed his review from November 1-3, 2022, and submitted his findings on July 31, 2023. D-17, Declaration of Nathan Adams 1 (July 2023). He did not attempt, in the intervening 8 months, to resolve these issues or try again. Instead, his report primarily focused on the quality of supporting documentation. See ibid.

Dr. Martin's limitations in reviewing the code, on the other hand, were partly by design. By his admission, he was not trying to perform an in-depth review, but rather was "surf[ing] around" and gathering an overall impression. (18T 78-18). He was not "bug hunting" but looking "at the overall quality of the code" in a cursory manner. (18T 97-10 to 99-12).

Dr. Martin responded to any issues he encountered by communicating with ESR, and when he did, he got what he asked for. (18T 74-13). The defendant's claim that Dr. Martin "faced the same limitations Mr. Adams faced," Def. br. at 51, may be true in the sense that they were both under a protective order, but is untrue in how each expert conducted their review.

c. Testability and Error Rates

It is important that a methodology be testable, or it remains only theory. Gissantaner, 990 F.3d at 463. Without testability, there can be no way to show that the challenged methodology works. This is consistent with Olenowski I's focus on the "soundness of the methodology and

reasoning used to validate the expert opinion or technique." Olenowski I, 253 N.J. at 154. Importantly, at the center of this inquiry is whether the methodology "can be 'assessed for reliability,' not whether it always gets it right." Gissantaner 990 F.3d at 464 (quoting F.R.E. 702 advisory committee's note to 2000 amendment). When the dispute focuses on the "'adequacy of the [theory's] testing' or about the 'accuracy of [a theory's] results,' generally speaking, [the arguments] provide grist for adversarial examination, not grounds for exclusion." Id. (quoting Bonds, 12 F.3d at 558-59. Thus, the testing of the software and the validation of the laboratory method are both relevant to the analysis under this prong.

Software can be developed and tested according to formal processes, or as part of an incremental process. Incremental, or iterative, processes are more agile and more common. Sommerville at 50. Testing in an iterative process often involves "backtracking when we realize that we have made a mistake." Ibid. Dr. Buckleton testified that he believes this "back and forth" is "a good thing." (7T 146-20).

The defendant suggests that it is especially important that the software is tested since STRmix's outputs can't be independently checked. While testing is undoubtably important, the assertion that STRmix's outputs can't be checked is untrue. While STRmix performs complex calculations and does so incredibly quickly, its standard output "writes out every step in every calculation." (7T 52-25). It may be time-consuming to check, but this court finds that every output by STRmix can be independently error-tested. In fact, three coding errors were found by a California DOJ criminologist, and he found them through his work manually tabulating every calculation from STRmix outputs. (7T 52-22). Consistency is still important, because not every output will necessarily be checked, but every output *can be* checked.

The Software experts had different opinions about the design and professionalism of the code, but no expert found a critical flaw. The defendant claims "[t]he failure to find a bug is never proof that a bug doesn't exist; it is axiomatic that '[t]esting can only show the presence of errors, not their absence.'" Def. Br. at 51 (quoting Sommerville, at 227). While it is true that the absence of proof is not the same as proof of absence, the defendant cannot simply point to a phrase and ask this court to take their word for it that bugs are lurking around every corner. The defendant's experts claim it is "likely" that there are undetected errors, and the defendant points to their three experts' consensus as sufficient proof. Def. Br at 51-52.

This is nothing more than speculation. The fact that more than one person made an unsupported guess does not increase its credibility. The existence of prior errors does not guarantee future errors, and the failure to find errors has, at best, no bearing on their existence. The absence of previous errors does not mean that errors must exist. A null hypothesis cannot be proven, only disproven. See Olenowski II, 255 N.J. at 603.

The defendant has argued that specification and requirements documents are required to confirm that the software is working correctly, and that having these documents is the only way to ensure the software does not fail. A requirements document does not necessarily need to be written before the software can be used, according to, Dr. Martin's testimony. Dr. Martin testified, "typically you come up with the requirements document first and maybe you revise it," but that is not to say you can't develop it without a requirements document and then create one later. Making a requirements document after the fact is "better than not having one." (18T 44-3 to 18). Usually, these documents are written before source code is developed, but Dr. Martin testified that this process is not mandatory. Dr. Martin felt the requirements document ultimately made was less

detailed than he would have hoped, but he did not find it to be incomplete or inadequate. (18T 44-23).

Defendant argues that the testing done was lacking in scope and in traceability, but every DNA laboratory who runs STRmix enlarges the scope of testing. Dr. Buckleton has asserted that 85% of STRmix's code has been tested; Dr. Martin challenges this estimate without any concrete data to say otherwise. (18T 63 to 64).

Developmental validation, in the sense that the forensic science community uses the phrase, is intended to demonstrate that a PGS is foundationally sound and suitable for purpose, but it is not required to set the outer limits of testing scenarios. ANSI/ASB Standard 018 at 1 (stating that "[d]evelopmental validation should also demonstrate any known limitations of the system" but also noting that statements prefaced by "should" are *not* mandatory components of the standard).

The Defendant claims that STRmix has not been validated in the forensic science sense of the word—that is to say, tested against sufficient samples to ensure reliable output. The Defendant concedes many tests have been done, but "not independent, they do not address the versions used in this case, and they do not test STRmix across the full range of samples STRmix is used on." Def. Br. at 89. The court finds none of these statements to be accurate.

First, to the defendant's point that there has been only one published developmental validation, and that this study was not independent: the requirement for independent developmental validation has not been found to exist. The PCAST report's comment that validation "should be performed by or should include independent research groups not connected with the developers of the methods and with no stake in the outcome," PCAST Report at 78-81, was a single non-binding comment by an advisory committee in 2016 that has not been universally adopted.

Second, as to their claim that the versions used in this case were not tested in the referenced validation study, this is true. The court takes notice of the fact that the study in question, Developmental Validation of STRmix, Expert Software for the Interpretation of Forensic DNA Profiles, 23 Forensic Science International: Genetics 226 (2016), was conducted in 2016. Any versions created after 2016 would not have been part of that study. This does not mean that these versions were not tested or that the validity of STRmix's underlying mathematical concept must be proved each time. After the first published validation study, there is little incentive to publish each subsequent study as "the results would be considered 'no longer novel' once the first paper was published." S-186, Declaration of Dr. Michael Coble 3 (October 2023). Each version has been developmentally validated, even if those results have not been sent to a journal for publication. (7T 33-7).

Third, as to the range of samples, the defendant has artificially narrowed their gaze and suggested that only 31 samples have been tested. This is false. Samples discussed in an article meant for publication will necessarily be abridged for the purpose of publication; the published data and the raw data for any article are different.

Dr. Coble testified that there are 240 crime laboratories in the United States and around 212 of those perform DNA testing. (11T 84-5). At least 120 of those laboratories use STRmix. (10T 99-1); (11T 84-8). Based upon an examination of over 100 laboratories, every internal validation study that validates STRmix performs tests and establishes horizontal validity across the sample space. (7T 93-5); see also Boodoosingh, et al. The 31 Labs study also combines and summarizes the data from thirty-one internal validation studies conducted with STRmix 2.5.02, using simulated non-donors and Caucasian allele frequencies. 31 Labs at 12. At 14 pages, it is but a summary of the underlying data, but it is based off analysis of 2,825 separate mixtures ranging

from three to six contributors. Ibid. The study specifically addressed allele sharing, including related contributors. Id. at 13. As the NOC increased, the amount of allele sharing increased as well. Id. at 15. "Large inclusionary LRs (LR > 1) for false contributors and exclusionary LRs (LR < 1) for true contributors where the [average peak height] was relatively high were investigated. For any given mixture, there is a chance that a given false contributor will have sufficient matching alleles, by chance, to give an LR > 1." Id.

The data gathered from these laboratories (including those used in the 31-laboratory study) is independent, even though these laboratories pay for STRmix; DNA analysts are not clients, they are not prosecutors or defense attorneys, and they are not defendants with a personal stake in the results of their tests. They are professional scientists, and they are also employees. The installation of STRmix in a laboratory is analogous to a company or agency installing a software suite on their employee's computers, or a new piece of equipment in their office. Individual employees have no stake in the success or failure of a new software program or customer retention platform, regardless of how much their employer paid for it. If it does not work satisfactorily, they do not have any intrinsic motivation to say otherwise.

The foundational validity of STRmix has been also independently demonstrated. One such demonstration was by using ground truth samples developed at one laboratory and running them through the parameters set by another laboratory. Boodoosingh et al, An Inter-Laboratory Comparison of Probabilistic Genotyping Parameters and Evaluation of Performance on DNA Mixtures from Different Laboratories, Forensic Science International: Genetics 71 (2024) Boodoosingh's study suggests that STRmix is relatively unaffected by differences in parameter settings. (7T 93-9 to 93-25) This study was conducted independent of the developers of STRmix.

Validations across different versions can still demonstrate the foundational reliability of STRmix as it relates to its performance in a laboratory setting. This court does find, similar to the court in Lewis, 442 F.Supp.3d 1122, that although code review is best thought of as version-specific, the validity of one version can still support the validity of other versions to the extent that the operation of the program remains the same.

Last, and as discussed above, to the extent that "no limits" were found in the initial developmental validation study, that does not defeat a finding of foundational validity. A developmental validation should discuss known or potential limits when possible, but the true limits for casework are the province of the individual laboratories.

STRmix has been validated repeatedly across 89 different laboratory systems in the US alone, which shows it has been well-engineered. See S-140.While these internal validations may take different approaches from a software validation, both have the same end goal in mind: ensuring that the output of the system is consistent, reliable, and conforms to expectations. While no laboratory's validation can stand in for another's, each study adds to the growing sample size of tests already run and demonstrates STRmix's testability. Even if every laboratory in the United States, and every test those laboratories have performed, were removed from the conversation, Dr. Buckleton testified that there have been *9 billion* false donor tests conducted on STRmix since its inception. (6T 98-20).

While the amount of raw data available is significant, tests need to be analyzed in order to be valuable. Internal validation studies, which has been performed and published for 32 U.S. laboratories at a minimum, provide that analysis. The FBI's internal validation study, Moretti et al., Internal Validation of STRmix for the Interpretation of Single Source and Mixed DNA Profiles,

29 Forensic Science International: Genetics 126 (2017) is one such example. It was concluded shortly after publication of the PCAST Report.

That study's findings were discussed at length in United States v. Lewis, 442 F. Supp. 3d 1122. In Lewis, the court appointed a neutral Special Master, who provided a 50-page report, entitled "Special Master's Report on the Scientific Foundations of STRmix."[14] United States Magistrate Judge David T. Schulz adopted these findings in his Report and Recommendation, and both the special master's report and the memorandum opinion in Lewis illustrate the lengths that even a single validation study goes to testing STRmix. The Special Master's report indicated that the aforementioned Bright and Moretti studies showed "persuasively that STRmix is capable of producing accurate results with extremely low error rates: that STRmix not only works, it seems to work extremely well, at least when used in the manner it was used in these studies." Id. at 1129.

Both Dr. Buckleton and Dr. Coble testified that repeated testing with known (ground-truth) samples is the best way to not only determine the error rates of STRmix's output, but to reveal software errors. There is no other way to validate or test STRmix but with ground truth samples. (6T 96-20). Mr. Inman makes a well-reasoned point, worthy of consideration, that many different laboratories testing the same sample repeatedly would not show anything other than the reproducibility of results. (19T 76-9). That is not, however, what has been done. Laboratories are not working off a standardized set of samples containing the same contributors, at the same ratio, with the same peaks at the same loci. If that were the case, it would indeed just test reproducibility. Defense counsel gave the court a very helpful demonstration of this fact: by setting a laptop running

[14] available at: https://bpb-us-e2.wpmucdn.com/faculty.sites.uci.edu/dist/0/594/files/2021/08/Special-Master-Report-10-31-19.pdf (last visited February 26, 2025).

STRmix to the same parameters as a previous test, and entering the same EPG data, defense attorney Godin was able to produce identical results. (5T 18-4 to 42-6).

However, Mr. Inman's point about whether a new laboratory's internal validation study is creating new tests misses the mark. When different laboratories are preparing their own ground truth samples for their own validation studies, they are making new mixtures. They use the DNA from different human beings. The EPG peaks will be different. The ratios will vary. The amount of DNA amplified will change. These are not identical reproductions because no humans, other than identical twins, have identical DNA.

Testing with ground truth samples is also how every coding error has been discovered in STRmix to date. While source-code review can reveal many things about a software program, experts have testified that there is no guarantee that a code review will spot an error, and most are found "through serendipity." D-11, Declaration of Dr. Mats Heimdahl 24 (July 2023). There was also repeated testimony that, in a non-continuous program,[15] any input has a chance of triggering a previously undetected error. Source code review might not reveal that, but brute-force testing could. This aligns with the philosophy of IEEE 1012, which acknowledges "[t]he dynamics of complex systems and the multitude of different logic paths available within the system in response to varying stimuli and conditions" means that any and all variations in input can produce an anomalous result. "The unlimited combination of system conditions presents the V&V effort with the challenge of using a finite set of analytical, test, simulation, and demonstration techniques to establish a reasonable body of evidence that the system is correct." IEEE 1012-2016 at 17.

[15] STRmix's non-continuous code is not to be confused with its fully-continuous analysis of samples, just as software validation should not be confused with laboratory validation.

The ever-growing body of ground truth testing has shown a minimal error rate. The developers of STRmix continue to promote more testing and more documentation—while their documentation may still weather some criticism, the potential error rate remains clear. STRmix, through Dr. Buckleton, pushes its bound with more testing and more variations in samples to show its willingness to comply with IEEE and respond, proactively, to criticisms from its peers. While that is an admirable goal, it is not the court's test for reliability.

The court finds that there may be limits on testing, and that not every possible variable can be reproduced in a laboratory sample, but that does not render the method unreliable. Olenowski II identified seven limitations that prevented a perfect replication of the 12-step DRE protocol in a laboratory setting. Olenowski II, 255 N.J. at 597-98. Nonetheless, despite the DRE protocol not being perfectly "testable" or "knowable," there was still sufficient data available to support an understanding of the general testability and error rates at hand. Ibid.

Dr. Buckleton has been characterized as stating that he does not believe STRmix has any limits. It appears to this court that is a mischaracterization. When Dr. Buckleton stated he believes STRmix "[doesn't] have a bound [at the low end]," he is referring to charting how STRmix behaves with a high-template sample vs. a low-template sample. (6T 84-1). As the template amounts reduce, Dr. Buckleton's observation is that the LRs generated also reduce, and that this is plottable on a graph down to 0. (6T 84-1). Dr. Buckleton states that he has not identified a point on the graph where the calculated LRs behave inconsistently or break from that curve. As a result, Dr. Buckleton has focused his developmental validation testing on other variables where STRmix *does* encounter challenges, such as "[c]ertain mixtures of relatives[, n]on-resolution of peaks[, and t]ri-allelic genetic patterns. (6T 82-6 to 84-1).

This means that for Dr. Buckleton, the total template amount is not a lower bound of STRmix, but other variables may be. This is very different from refusing to set limits or "explicitly refusing to give [error rates]." Def. Br. at 99. This is Dr. Buckleton explaining what, in his opinion, is important to address when establishing STRmix's limits and what is not. This is consistent with the SWGDAM guidelines' understanding of developmental validation as the "acquisition of test data to *verify* . . . the determination of limitations. SWGDAM Guidelines at 5 (emphasis added). This statement is consistent with the further explanation that developmental validation is concerned with demonstrating "known or potential limitations" but not necessarily actual ones. Ibid.

There is nothing to suggest that STRmix is more prone to bugs or coding errors at lower-template mixtures than it is at higher levels. If the software works, and can be shown to work, then it can be reliable. A high-template sample has the same rough odds of triggering an unexpected and unintuitive bug as a low-template sample. Nothing changes within the coding or algorithm in STRmix when you input data from an electropherogram. When a smaller number of alleles appear in a sample, the math, the methods, and the algorithm remain the same, it just leads to less probative results. If STRmix produces a LR of 450, it is not substantially different from a random match probability of 1 in 450. Such a result would fall within the uninformative range of some laboratories reporting scales, but that does not mean it is inaccurate and unreliable – it just means it is less probative. While the risk of a miscode or a bug randomly affecting a calculation will always exist—and has been found in the past—testimony from both sides agrees that the chances of a bug occurring exists independently of the inputs into the system.

There are some types of casework samples interpreted by STRmix that, despite the ground truth not being known, can be empirically tested. If a casework sample is robust enough to be

analyzed through traditional methods and generate a result sufficient for source attribution[16] and STRmix produces a corroborating LR for that sample, or if STRmix and another PGS provide similar results, then this shows the output of STRmix is testable under at least some circumstances.

The case before the court gives an example of this type of testability. Some of the samples in question were tested by two different Laboratories, Bode Technologies and the New Jersey State Police laboratory. Of those samples, some were tested with STRmix by one laboratory and manually by another—for example, a pair of jeans from the defendant's home. The results from Bode's use of STRmix indicates the presence of a profile consistent with [redacted] DNA on these jeans in two locations (E01a – interior thigh above the knee with a LR of 470 sextillion & E01c – exterior front right thigh with a LR of 2.1 septillion). The traditional, manual analysis originally done by the New Jersey State Police DNA laboratory in 2018 also detected this DNA, indicating that [redacted] was consistent with the "source" of DNA found on the jeans in five separate areas (samples 6-1-2-1; 6-1-3-1; 6-1-5-1; 6-1-7-1 and 6-1-8-1). Ex. S-184A. The threshold for source attribution at the New Jersey State Police DNA laboratory was defined by Christine Schlenker as exceeding 1 in 8 trillion. (9T 85-9). It is hard to imagine a better way to demonstrate that the STRmix results are reliable – here, the STRmix results and the traditional results support consistent conclusions for the same item.

Disputes about the adequacy of testing or the accuracy of results are the crux of cross-examination, not grounds for exclusion. Bonds, 12 F.3d. at 558-559; United States v. Baines, 573 F.3d 979, 989-90 (10th Cir. 2009). Even where independent experts disagree on the adequacy of

[16] The New Jersey State Police Laboratory, for example, considers an RMP of 1 in 8 trillion to be high enough that they will report, with a high degree of scientific confidence, that a specific individual is the source of DNA. (9T 67-11).

testing, it does not mean the theory is untestable. In Gissantaner, the Court of Appeals stated that the District Court identified "shortcomings" in STRmix, but even "serious deficiencies" in testing do not render a method untestable. Gissantaner, 990 F.3d at 468; see also Bonds, 12 F.3d at 559. At stake is "scientific validity," not "scientific precision." Bonds, 12 F.3d at 558. From both a software and a scientific perspective, STRmix "can" and "has been" tested. Daubert, 509 U.S. at 593.

Testability and error rates are, as discussed above, intertwined. The latter part focuses on "whether the scientific community has established standards that forensic scientists can use to mitigate the risk of error." Gissantaner, 990 F.3d at 465; see also Daubert, 509 U.S. at 594. So, for example, if the identified methodology has a high error rate, and lacks standards and guidelines to minimize these risks, this would be of concern. Gissantaner, 990 F.3d at 465. If the government wishes to use STRmix to match a defendant to DNA on a piece of evidence, it can do so only if the results are the "product of reliable principles and methods" under R. 702. If STRmix has a high error rate, if it has trouble "avoid[ing] . . . false positives," and if there are no standards or guidelines to avoid or lessen these risks, then it should not be used. Bonds, 12 F.3d at 559; United States v. Mitchell, 365 F.3d 215, 241 (3d Cir. 2004).

"The fact that a possibility of error exists does not preclude a conclusion that a scientific device is reliable." Romano, 96 N.J. at 80. The defendant has argued that knowing the error rate "exists and is indeterminate . . . is insufficient," Def. Br. at 159, but misstates the issue. "The absence of a definitive rate of error" is not always "a dispositive basis to exclude" a methodology. Olenowski II, 255 N.J. at 603. The factor, relevant to this case law, for assessing reliability is "the known or potential rate of error," so error rates need only be potential. Id.

The true, definitive error rate of STRmix may never be known. DNA itself does not deal in conclusory statements, but in probabilities. STRmix creates a LR. Traditional DNA analysis generates an RMP. Neither of these figures, standing alone, are statements that lend themselves to a clean categorization of "accurate" or "erroneous." There is information available that suggests that STRmix may exclude an actual contributor less than 7% of the time and includes a non-contributor between 1% and 2% of the time. Whether these are "errors," functions of the inescapable reality of biological analysis, or expressions of STRmix's intentional conservatism[17] is less apparent. A definitive error rate is not always possible, nor is it always required. Asking STRmix to prove it does not have errors requires it to prove the null hypothesis, and "no set of statistical results is capable of establishing that a null hypothesis is actually true or false." Olenowski II, 255 N.J. at 603. "[A]bsolute scientific certainty is not the standard for the admissibility of expert testimony." Id. (quoting Paolino v. Ferreira, 153 A.3d 505, 523 (R.I. 2017)).

LRs having variation does not mean that STRmix is unreliable. Other courts have considered the accuracy of LRs generated by STRmix and found that variation in the precise LR numbers generated by STRmix from one test to another is very small. Lewis, 442 F.Supp. at 1130. Altogether, the LRs produced by STRmix "are properly calibrated and do not overstate the value of incriminating evidence," so the "absence of a specified error rate does not render STRmix unreliable." Id. at 1131.

Dr. Coble discussed these "fortuitous" situations that some like to call errors. He explained how these are not errors, just "an example of genetics, biology… when you're only looking at two

[17] Dr. Buckleton has testified that STRmix is designed to favor "conservative" results and tend towards exclusion when information is limited, which may partly explain the higher false exclusion rate.

or three markers, you may find by random chance people who could give a profile that "matches" with the person of interest, but they're not that contributor. So, that to me, is not a true error." (10T-86-9 to 16). He described what would be an error that we want to avoid – a high LR supporting inclusion for a non-contributor. He said he has seen "nothing that would be at that level, at that high quadrillions type statistics." See (10T 88-1 to 88-13). This is consistent with what he explained in his report, that "STRmix will produce an LR greater than x from about 1 in x false donors… this is the most concise expression of error rate available." Ex. S-152 at 25.

While there is no precise rate for inaccurate inclusion/exclusion, Dr. Buckleton has previously suggested that it is immeasurably small, putting it at "somewhat less than one over the LR". See Special Master's Report at 33-34, Lewis, 442 F.Supp.3d at 1154. Further, the Special Master found that false inclusions were rare and when occurred, occurred only as often as would be expected due to similarity amongst the different profiles involved. Id. at 33. Dr. Coble supports this very conclusion in his discussion of "known or potential error rate," describing this "false inclusion" scenario as being better described as a "fortuitous match;" he indicates that this is simply due to the rare scenario when a non-contributor "may share several alleles with the person of interest in the mixture." (T10 85-22 to 86-16).

The Lewis court went on to note that the absence of a precisely calculated error rate is not the same as saying there is no known error rate. While errors were possible, they held that the STRmix internal validation study established it as acceptably small. Lewis, 442 F.Supp. 3d at 1130; see also Special Master's Report at 42. The Special Master in Lewis stated that "while there were a few instances in which STRmix produced results that falsely linked non-contributors to the mixtures, these misleading results were rare and occurred no more often than would be expected

by chance due to adventitious (coincidental) similarity between DNA profiles of different individuals." Id. at 1130, Special Master's Report at 8.

One study has found that across all types of mixtures, there was a 1.35% non-contributor inclusion rate. Riman et al., Examining performance and likelihood ratios from two likelihood ratio systems using the PROVEDIt dataset, 16 PLoS One 9, 10 (2021). 2.07% of true contributors were incorrectly excluded. Ibid. The non-contributor inclusion rate for four-person mixtures was 2.76%. Id. at 11. The FBI's internal validation study found a 6.1% incorrect exclusion rate overall, a .1% non-contributor inclusion rate overall, and a 1.7% non-contributor inclusion rate when an additional contributor was incorrectly assumed to be included. Moretti et al., Internal Validation of STRmix For the Interpretation of Single Source and Mixed DNA Profiles, 126 Forensic Science International: Genetics 126, 138, 141 (2017). These inclusion rates, given the explanation from Drs. Buckleton and Coble that non-contributor inclusion is somewhat immutable, are not alarming. In fact, the FBI validation study explains that it would be more alarming to see a total lack of non-contributor inclusion during a validation study.

> Given probabilistic modeling within the stochastic range, LRs >1 are expected for some non-contributors. In validation testing, failure to demonstrate false support would indicate that the system is either not functioning properly or has not been queried with sufficiently challenging specimens. In fact, a high LR for a simulated non-contributor may even result from a high template single source profile, since simulation of a large number of non-contributor genotypes will eventually produce one that matches the profile.
>
> [Id. at 138]

Developmental validation is about foundational validity and acquiring data to ensure that limitations set by laboratories have adequate corroboration. (6T 84-1). This court finds that false exclusion or inclusion rates, or any doubts as to those rates, are a valid point for cross-examination, but not for the scientific reliability of the program.

STRmix also contains error-mitigation protocols in the form of diagnostics that alert an analyst to irregularities in the deconvolution process. While figures like the Gelman-Rubin (GR) diagnostic cannot tell a human what, if anything, the problem was, they can alert a human being to proceed with caution. See, e.g., D-52, email from Stuart Cooper to Brett Hutchinson, subject: Gelman-Rubin (describing different approaches to interpreting Gelman-Rubin diagnostic value and noting that "a high GR can still be suggestive of an issue, so if any weights don't feel intuitive then we have an issue and don't report that."). Since the human analyst is ultimately the one testifying to their work during a trial, the human element represents a significant error mitigation step. And, while the ground truth of a casework can never be known with certainty, humans can corroborate results independently in some cases. For example, when a STRmix calculation can be checked against traditional analysis or another iteration of a PGS, then the possibility for mitigation exists. In this case, many of the STRmix calculations are of the type that can be checked "by hand", leading to further confidence in the results. Drs. Buckleton and Coble both testified regarding Steven Myers, from the California Department of Justice DNA laboratory system and how he replicated many of these calculations by hand. (T7-52-17 to 53-4); (T10-46-15 to 47-5). This was confirmed by Keith Inman, who previously worked at the California DOJ laboratory system. (T19-143-4 to 143-18). Ms. Thayer described the section of the New Jersey State Police internal validation study that demonstrated how "genotypes derived from non-probabilistic analyses of profiles above those stochastic threshold[s] should be in complete concordance with the results of probabilistic methods." (8T 79-10). In other words, that STRmix should produce a similar result as a manual tabulation. For some of Thayer's calculations, the manual tabulation and STRmix were "exactly the same". (8T 80-18).

The defendant's contention that it is always an error "when a PGS gives an inclusionary LR for someone who did not contribute to a mixture or an exclusionary LR for someone who did" is explicitly rejected. Def. Br. at 64. There is no authority to support that statement, and it directly contradicts New Jersey caselaw. Due to the nature of DNA analysis in general, inclusionary support for a non-donor cannot categorically be labeled as a false positive. The fact that STRmix or any other PGS generates an LR above 1 for a non-contributor to a sample cannot be automatically characterized as a false positive or taken as proof of unreliability. The fact that inclusionary LRs tend to increase in samples with more contributors is also expected and can be accounted for. It "does not mean that mixed DNA profiles containing more contributors are less reliable, just that they are *less informative* with respect to potential contributors." 31 labs at 20 (emphasis added).

STRmix cannot distinguish between a person's DNA and DNA from another contributor that is similar. STRmix cannot distinguish between people at all; no method of DNA analysis can. Once an allele is reduced to a peak on an EPG, it is identical to any other allele that produces the same peak at the same locus. Any random amount of allele sharing between two individuals, related or not, has the potential to generate inclusionary support. What matters is whether the LR is appropriately high or low when compared to how many alleles are shared between a sample and an individual. Buckleton et al, The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity, 64 J. Forensic Sci. 393, 398 (2019) ("In a large set of mixtures compiled from 31 laboratories, all large (over 10,000) LRs for nondonors were investigated; in all instances, the nondonors had high allelic overlap with the profile. This is the correct result."). This is also the underlying principle behind the RMP given in traditional DNA analysis: there is *always* a possibility, however small, that a person aligns with a DNA profile by mere chance.

Probabilistic genotyping systems will therefore give inclusionary LRs for non-contributors (for the reasons explained above) and can also give exclusionary LRs for contributors (because of drop-out or other reasons). There is nothing inherently unreliable about this fact; other types of DNA analysis—such as mitochondrial or Y-STR DNA, which are shared by all matrilineally related persons or patrilineally related males, respectively—will provide inclusionary support for a vast swath of persons, yet they are reliable when used for their proper purpose. See State v. Calleia, 414 N.J. Super. 125, 150 (App. Div. 2010), rev'd on other grounds, 206 N.J. 274 (2011) (holding that Y-STR DNA is reliable even when it "cannot unequivocally establish" a suspect's presence in a sample but can show a person "cannot be excluded").

d. General Acceptance

Although each of the Defendant's experts had criticism, valid or otherwise, for STRmix's documentation and design aspects, as well as the testing process that STRmix underwent, there was no conclusive evidence that STRmix was failing to meet the generally accepted coding standards of the software community, or that they should be held to a more stringent one. Many of Mr. Adam's opinions that STRmix does not meet general acceptance in the software community focus on the accompanying documentation rather than the program itself, such as whether a specification should have been defined as "not zero" or "more than zero":

> Some of the "response" entries appear to be specifications – text accompanying the number of contributors (NOC) input parameter states that NOC may not be zero, but a screenshot of a STRmix error message indicates that NOC must be a positive integer. While the error message indicates that the correct behavior is implemented in STRmix, the discordance between the specification acceptance criteria (constrained to a non-zero value) and actual software behaviors (constrained to integer > 0) is troublesome. This discordance is apparent for multiple additional input parameters where acceptance criteria are specified as non-zero values (or any non-blank value) but which must mathematically constrained to positive integers. Obviously, a negative number of humans cannot contribute DNA to a sample. That no review process corrected this obvious error should raise questions as

> to whether less-obvious failures to properly specify system behaviors have also occurred.
>
> [D-17, Declaration of Nathan Adams 5-6 (July 2023).]

It is true that a negative number of humans cannot contribute to a sample, but Mr. Adams acknowledges that the actual software does not accept negative numbers. Observations such as these cannot be taken to show a lack of general acceptance in the software community, because observations such as these address the paperwork and not the software.

STRmix's general acceptance amongst forensic scientists is a more straightforward inquiry. According to Dr. John Buckleton, STRmix is currently in laboratories in all eight of the state and territory laboratories in Australia and laboratories elsewhere including the United States Army Criminal Investigation Laboratory; the Federal Bureau of Investigation; DNA Labs International; the Federal Bureau of Alcohol, Tobacco, Firearms and Explosives; and the Union County Prosecutor's Office; as well as laboratories in New York, California, Idaho, Michigan, Texas, Arizona, Oregon, Wyoming, Connecticut, Illinois, Florida, Kansas, and Indiana, *inter alia*. An updated list of laboratories was supplied by the State. See exhibit S-140.

The New Jersey Supreme Court has consistently held that the general acceptance of scientific evidence can be established by expert testimony, authoritative scientific literature, or persuasive judicial opinions. Harvey, 151 N.J. at 170 (quoting Kelly, 97 N.J. at 210). In Harvey, the Supreme Court looked to other jurisdictions in the context of a Frye hearing where "at the time of the R. 104 hearing, both the State and the defense were unaware of any judicial opinion discussing a new form of scientific evidence, specifically in that case regarding polymarker evidence." Id. at 175. The court cited Wilkerson v. Pearson, 210 N.J. Super. 333, 336 (Ch. Div. 1985), which held that absence of judicial opinions demonstrating acceptance by other courts of a particular type of scientific technique should not, by itself, foreclose a finding of general scientific acceptance and

reliability. The Harvey Court concluded that 6 courts had recently adopted polymarker DNA, which "support[ed their] conclusion that the trial court correctly admitted the evidence in the present case." Harvey, 151 N.J. at 175–76. In Marcus, the Appellate Division held that the trial court properly admitted results of a method of DNA analysis, reasoning that "[a]lthough there is no reported appellate decision in New Jersey dealing with the admissibility in a criminal trial . . . there is overwhelming authority in other jurisdictions sustaining the admissibility of such evidence." Marcus, 294 N.J. Super. at 282-83. The method of DNA analysis used in that case was "clearly established by authoritative scientific literature, the overwhelming weight of judicial authority throughout the country, and the testimony of experts at the Frye hearing[.]" Id. at 291. Similarly, this court notes the overwhelming trend towards acceptance nationwide.

Although a case of first impression in New Jersey, Federal courts and our sister jurisdictions have had opportunity to examine STRmix's reliability and have found it generally admissible, with some caveats. Defense counsel has argued that there are "no judicial opinions" that "persuasively demonstrate" STRmix's reliability. Def. br. at 154. In making that assertion, defendant has misunderstood the New Jersey Supreme Court's jurisprudence: courts shall not *limit* their reliability analysis to another jurisdiction's interpretation of the Daubert factors. Olenowski I, 253 N.J. 154.

In fact, Olenowski II expressly held that case law is an indicator of general acceptance, 255 N.J. at 606 (Citing Kelly, 97 N.J. at 210). The Court thoroughly applied that principle and extensively listed the opinions of sister jurisdictions that had accepted DRE evidence as reliable, and making a point to mention that "[o]nly a handful of courts -- none of which are a state's highest court -- have held that DRE evidence is inadmissible." Olenowski II, 255 N.J. at 606.

The defendant misinterprets Pickett, 466 N.J. Super. 270,[18] Doriguzzi, 334 N.J. Super. 530,[19] and J.L.G., 234 N.J. 265,[20] as establishing authority that overturns Kelly's holding that "authoritative scientific and legal writings [and] judicial opinions that indicate the expert's premises have gained general acceptance" are still, even post-Olenowski, proper for a court to consider. Kelly, 97 N.J. at 210.

None of the cases cited suggest that caselaw from other jurisdictions should be rejected in favor of starting from scratch here in New Jersey, and although defense counsel argues otherwise, there is no binding precedent to suggest that New Jersey jurisprudence agrees. If anything, the cases cited—which support the proposition that holding an evidentiary hearing is more preferable to not—all stand for the proposition that an appropriate part of a Rule 702 hearing is to consider other courts, weigh their reasoning, and be persuaded accordingly.

---

[18] It is important to note two things about the authority "house of cards" rejected in Pickett. First, the issue in Pickett was *not* whether TrueAllele was reliable, but whether defendant was entitled to a source code review and a reliability hearing. The court in Pickett acknowledged that "[p]ublished out-of-state judicial decisions, although persuasive rather than binding, carry great weight." Since none of the cases considered had addressed the concerns raised by the defendant, there was a justifiable reason to examine TrueAllele's code for errors. Pickett, 466 N.J. Super. at 316. There was no acceptance or rejection as to the ultimate conclusions of reliability, but rather a finding that those conclusions should not prevent a defendant from a hearing if their challenges exceeded the scope of what had been considered before. Second is that none of the other cases cited in Pickett as examples were published or precedential. There was only one authority in Pickett that could even serve as a persuasive foundation, which is why the court was dealing with a "house of cards" and not a "game of telephone."

[19] The quote that the defense attributes to the Doriguzzi court actually comes from People v. Kirk, 681 N.E.2d 1073 (Ill. App. Ct. 1997), where the Appellate Court of Illinois held that the reliability of the Horizontal Gaze Nystagmus test required a Frye hearing and that relying *solely on* other court's decisions was a slippery slope of uncritical acceptance, a "yellow brick road." Kirk, 681 N.E.2d at 1073.

[20] The Court in J.L.G. did *not* find that 40 other states were uncritically allowing Child Sexual Assault Accommodation Syndrome Testimony into their courts; they found that 40 states allowed such testimony "for some purpose" and that "[m]any, *like New Jersey*" allowed it for a limited purpose. 234 N.J. at 288-89 (emphasis added).

Not only are the judicial opinions of other states that address STRmix relevant, the court finds that they are persuasive and that persuasive does not—as defense counsel repeatedly insinuates—mean uncritically or "reflexively" taken at face value. To the extent that other courts have addressed the same features of STRmix as are before this court, their persuasiveness can be judged for its own merits. The extent to which other opinions have not addressed specific concerns before this court does not prevent those opinions from assisting on the points they have addressed.

To say that those opinions are "simply not good enough" and to write them off wholesale is not compelling to this court. It would be reckless to not consider everything, including judicial decisions, in making this decision.

In Gissantaner, the 6th Circuit reversed the trial court's suppression of a STRmix-generated report and found that "[m]easured by Evidence Rule 702 and a proper framing of the Daubert factors, the DNA evidence should be admitted on this record." 990 F.3d at 463. In Gissantaner, STRmix was used to test "touch DNA" collected from the grip of a firearm and offered to show that the weapon had been in defendant's possession. Id. at 460-61. The court recognized that such evidence "can be highly probative" and made a thorough inquiry into STRmix's reliability, reasoning that treating even a high likelihood of a particular person's DNA being present as dispositive "is not a good idea . . . if STRmix has a high error rate, if it has trouble avoiding false positives, and if there are no standards or guidelines to avoid or lessen these risks." Id. at 463, 465.

The court examined STRmix's testability, peer review, error rates and standards, and general acceptance and found it both reliable and reliably applied but cautioned that a positive LR could only demonstrate that Gissantaner's DNA was present on the weapon, and not how it ended up there. Id. at 470. As to testability, the court found that "STRmix can be tested . . . [u]sing laboratory created mixtures in which the actual contributors of DNA Samples are known" id. at

464. Its peer-reviewed status was considered sufficient. See id. at 464-65. Its overall accuracy was calculated at above 99%, and importantly, the court found that false positives were correlated with low LRs. “A likelihood ratio of 100 to 1 is more likely to produce a false inclusion than a likelihood ratio of 1 million to 1.” Id. at 465. “Importantly, the question is whether a method can be assessed for reliability, not whether it always gets it right.” Id. at 464.

In reaching this conclusion, the Gissantaner court highlighted that STRmix is the "market leader in probabilistic genotyping software." Id. 466-467. The Gissantaner Court continued:

> Consistent with this reality, numerous courts have admitted STRmix over challenges to its general acceptance in the relevant scientific community. *See United States v. Lewis*, 442 F.Supp. 3d 1122, 1155 (D. Minn. 2020) ("[T]here is no doubt that STRmix has gained general acceptance."); *United States v. Washington*, No. 8:19CR299, 2020 U.S. Dist. LEXIS 105447, 2020 WL 3265142, at *2 (D. Neb. June 16, 2020) ("Authority and evidence demonstrate that STRmix is generally accepted by the relevant community.") (Attached as Pa20); *People v. Blash*, No. ST-2015-CR-0000156, 2018 V.I. LEXIS 86, 2018 WL 4062322, at *6 (V.I. Super. Ct. Aug. 24, 2018) (Attached as Pa24); *People v. Muhammad*, 326 Mich. App. 40, 931 N.W.2d 20, 30 (Mich. Ct. App. 2018); *People v. Bullard-Daniel*, 54 Misc. 3d 177, 42 N.Y.S.3d 714, 724-25 (N.Y. Co. Ct. 2016); *United States v. Christensen*, No. 17-CR-20037-JES-JEH, 2019 U.S. Dist. LEXIS 24623, 2019 WL 651500, at *2 (C.D. Ill. Feb. 15, 2019) ("STRmix has been repeatedly tested and widely accepted by the scientific community.") (Attached as Pa25); *United States v. Oldman*, No.18-CR-0020-SWS, 2018 U.S. Dist. LEXIS 232762, ECF No. 227 at *16 & n.5 (D. Wyo. Dec. 31, 2018) (collecting cases) (Attached as Pa26); *United States v. Russell*, No. CR-14-2563 MCA, 2018 U.S. Dist. LEXIS 232864, 2018 WL 7286831, at *7-8 (D.N.M. Jan. 10, 2018) ("[STRmix's] analyses are based on calculations recognized as reliable in the field.") (Attached as Pa27); *United States v. Pettway*, No. 12-CR-103S (1), (2), 2016 U.S. Dist. LEXIS 145976, 2016 WL 6134493, at *1 (W.D.N.Y. Oct. 21, 2016) (discussing "exhaustive[] research[]" concluding that "the scientific foundations of the STRmix process are based on principles widely accepted in the scientific and forensic science communities")(Attached as Pa21). The Second Circuit determined that the scientific community accepted a different (but similar) DNA-sorting software, Forensic Statistical Tool, even though just one laboratory had used it. *Jones*, 965 F.3d at 156, 162.
>
> [Ibid. (citations and parentheticals in original)].

In June of 2024, the Southern District of California granted defendant’s motion to exclude STRmix evidence after determining that it was unreliably applied to touch DNA taken from a

handgun found during a traffic stop. United States v. Ortiz, 736 F.Supp.3d 895 (S.D.Cal. 2024). The court excluded this evidence after finding a bright-line rule for STRmix: that it "may only be used where NOC is five or less." Id. at 899. With respect to probabilistic genotyping software, "judges should ascertain whether the published validation studies adequately address the nature of the sample being analyzed." Id. at 900. In Ortiz, there were likely six contributors to the sample, which was found to be beyond the limit of STRmix. Ibid.

Mr. Ortiz conceded that STRmix satisfied F.R.E. 702's requirements and was "a product of reliable principles and methods." Ortiz, 736 F.Supp.3d at 900. His challenge rested solely on "the process by which STRmix was applied to a complex DNA mixture that likely contained six contributors." Id. at 901. The analyst who had studied the sample in question "[couldn't] say for certain it's not a six-person mixture" and their "best estimate [was] that it's a five-person mixture." Id. at 902. The court found that apparent 5-person mixtures had the highest risk of being underestimated, but that 2- and 3-person mixtures were at a lesser risk of underestimation. See id. at 903 (finding that "almost all" underestimations occurred on 4 or 5-person mixtures).

"Most labs have only validated STRmix's use of samples where NOC is four or less." Id. at 899. The district court in Ortiz noted that "risks are heightened . . . where evidence in the record demonstrates that the sample may have contained DNA from related individuals." Id. at 903 Like Ortiz, the samples in this case have been hypothesized to contain DNA from related individuals. Unlike Ortiz, those risks are mitigated: there is no evidence that allele sharing between the defendant and the four victims in this case could mask 5 or 6-person mixture as one that contains only two or three. There has been no testimony proposing such a scenario is even possible—only a

general acknowledgement that relatives share some alleles. Many of the samples tested also contained an allele only possessed by [redacted] due to her mutation, which means that at least one allele can never be hidden and can never be undercounted.

The facts and circumstances of Ortiz make it inapposite for comparison to the instant matter. Mr. Ortiz challenged STRmix under a completely different theory of reliability, the only experts who testified were the original DNA analyst and a single rebuttal witness, and the district court acknowledged that even "scientifically sound methods have been applied improperly ordinarily should be left for the jury to resolve" Id. at 904. The combination of several complicating factors and a lack of expert testimony explaining those factors led the court in Ortiz to a sound conclusion on a different issue than the one before the court today.

The Middle District of Pennsylvania has found TrueAllele—STRmix's main competitor which employs a similar underlying algorithm—to be reliable by applying the analysis used in Gissantaner. United States v. Anderson, 673 F.Supp.3d 671 (M.D. Pa. 2023).

TrueAllele was also found reliable under a Daubert-like framework by the Supreme Court of Nebraska, which adds cumulative but not controlling support that the underlying concept of probabilistic DNA analysis has reached a general state of acceptance. See State v. Simmer, 935 N.W.2d 167 (2019).

The Supreme Court of Delaware also found STRmix to be reliable under their parallel rule of evidence, and found the evidence of reliability was extensive to the point that a Daubert hearing was unnecessary. Hudson v. State, 312 A.3d. 615 (Del. 2024).

Other published cases from outside New Jersey considered by this court include United States v. Lewis, 442 F.Supp.3d 1122 (discussed passim); People v. Bullard-Daniel, 163 N.Y.S.3d 726 (App. Div. 2022) (STRmix generally accepted as reliable under Frye standard); People v.

Davis, 290 Cal.Rptr.3d 661 (App. Dep't Super. Ct. 2022) (STRmix generally accepted under Frye); People v. Muhammad, 931 N.W.2d 20 (Mich. Ct. App. 2018) (STRmix reliable under Daubert standard).While remaining cognizant of Rule 1:36-3, and noting that the following opinions are not only non-binding and non-precedential but also unpublished, the court takes judicial notice of the following opinions not for their reasoning but to illustrate the extent that STRmix has been discussed:

In United States v. Christensen, the District Court found STRmix to be reliable even if there are more precise tests available, due to STRmix having "been repeatedly tested and widely accepted by the scientific community." No. 17-CR-20037-JES-JEH, 2019 WL 651500, at *2 (C.D. Ill. Feb. 15, 2019).

In United States v. Russell, the United States District Court held STRmix to be reliable and denied defendant's motion to exclude this evidence. No. CR-14-2563 MCA, 2018 WL 7286831, at *8 (D.N.M. Jan. 10, 2018).

In United States v. Washington, the District Court found that STRmix has been "tested[,] subjected to peer review, . . . widely accepted by the relevant community, and . . . examined and admitted in other courts of competent jurisdiction,". No. 8:19CR299, 2020 WL 3265142, at 5 (D. Neb. June 16, 2020).

In United States v. Pettway, the District Court held that STRmix was scientifically valid and could be relied upon. No. 12-CR-103S, 2016 WL 6134493, at *2-3 (W.D.N.Y. 2016).

In State v. Hudson, the trial court found that that STRmix has been examined by other courts of competent jurisdiction, admitted into federal courts where the Daubert factors were applied to STRmix, and was scientifically valid. No. 1809009750, 2021 WL 4851971, at *6 (Del. Super. Ct. Oct. 15, 2021).

In People v. Seepersade, the court found that the raw data for DNA analysis must be produced to the defendant for STRmix to be admissible as "the People could not refuse to produce during discovery the final conclusions of STRmix analysis of a DNA mixture." 58 Misc.3d 1227(A) (N.Y. Sup. Ct. 2018).

In State v. Santos Javier Nunez, the court held that STRmix is software is valid, is admissible under the rules of evidence, and the software is generally accepted in the relevant scientific community. State of Connecticut v. Santos Javier Nunez, No. HHD CR17-0690894T (decided Aug. 15, 2019). Connecticut adopted the Daubert factors in State v. Porter and noted that it is the "proper threshold standard for the admissibility of scientific evidence." State v. Porter, 241 Conn. 57, 59 (1997).

In People v. Smith, the court noted if a "trial court determines that recognized scientific, technical, or other specialized knowledge will assist the jury in understanding evidence or determining a fact in issue, a witness qualified as an expert may testify." People v. Smith, No. 340845, 2018 WL 4926977, at *7 (Mich. Ct. App. Oct. 9, 2018). In Smith, the court found that the trial court did not err in admitting STRmix due to the testimony of Dr. John Buckleton. Id. at *8. Dr. John Buckleton testified, in 2018, that STRmix "had never caused a false inclusion," he only "knew of only two errors causing false exclusions" which "occurred during specific testing exercises," "STRmix™ was used in 17 laboratories in the United States," "[s]ixty-five additional laboratories had purchased the software, and "STRmix™ had been favorably reviewed in 19 peer-reviewed scientific journals." Id. This testimony showed the appellate court that STRmix was reliable and that the trial court did not err in admitting it. Id.

Overall, these cases suggest a general acceptance of STRmix but, more importantly, they show it is not a blind acceptance. STRmix is accepted as reliable when it is used reliably, which is

consistent with this court's analysis. It has never been found to be foundationally unreliable. There have been limits imposed, in cases involving a high NOC or in cases questioning the probative value of a low-template sample, but PGs generally and STRmix specifically are generally accepted as reliable science. As stated before, the fact that other jurisdictions have found STRmix and the underlying science of probabilistic genotyping to be reliable is helpful to show cumulative evidence of support, but in no way binds this court or allows for corners to be cut in its analysis. See Doriguzzi, 334 N.J. Super. at 540.

The overwhelming trend of acceptance to STRmix is not limited to courts in sister jurisdictions conducting evidentiary hearings. Laboratories across the country are steadily shifting to these types of systems. Dr. Coble was asked by the court about the number of DNA laboratories in the United States. He estimated that 212 laboratories conduct DNA testing and estimated that 130 or 140 of those have brought on Probabilistic Genotyping software. (11T-84-9 to 84-12). Of these, STRmix is used in the clear majority. (10T 49-50). As Dr. Coble succinctly and accurately stated, "I basically say there are two types of labs in the U.S. There are labs that are using probabilistic genotyping and there are labs that will soon be using probabilistic genotyping." (10T 82-13).

As of November 10, 2024, there are currently over one hundred individual laboratories who have internally validated and use STRmix daily. (S-140). These laboratories belong to eighty-nine laboratory systems. (S-140). Dr. Coble testified that "there's way more than 89 laboratories that are using STRmix . . . for example in Texas, we have [the] Texas DPS system, the Department of Public Safety," which consists of more than one laboratory. (10T 50-6). The number of laboratories using STRmix continues to grow. Dr. Coble testified that STRmix was first used by a United States laboratory in 2014. (10T 48-14).

STRmix has several competitors that employ the same fundamental concept. Laboratories and the forensic DNA community, in general, have been using and teaching about PGS before STRmix was even developed. (11T 63-14). Professionals in the forensic DNA community, like Dr. Coble, were excited for STRmix to be developed. (11T 63-23). STRmix and TrueAllele, a PGS competitor, are the two most used programs in criminal laboratories. (11T 84-10). The first case that used TrueAllelle was in 2009, while STRmix was not live in case work till 2012. (6T 16-20). As a whole, probabilistic genotyping "has been accepted and adopted by the forensic science community." (10T 99-5). Based off Dr. Coble's domestic and international experience, "STRmix is generally accepted in the United States." (10T 99-1).

e. Olenowski Conclusion

STRmix is reliable. It has sufficient standards to govern its development and its use. It is also important to note that the ANSI/ASB standard does not require hard limitations to be put on STRmix at the developmental stage, but rather that *known* limitations should be demonstrated. If STRmix and the underlying science of probabilistic genotyping have unrealized potential, then not knowing its limitations does not violate the standard. Each laboratory, through its internal validation study, needs to determine the limits of STRmix for themselves, and determine how it fits into the rest of their laboratory's systems and processes. See S-133.

A review of the aforementioned factors in their totality is important to an overall assessment of reliability. Having done so, there can be no doubt that STRmix is reliable given the sheer amount of testing which has determined that it is fit for casework. This is not the first Probabilistic Genotyping Software. It will not be the last. Probabilistic genotyping is just another in a line of many advancements which have occurred since the advent of DNA in 1988. While STRmix was not the first continuous probabilistic genotyping system, it is undoubtedly the preeminent in the

United States. Since the PCAST report was issued in 2016, the number of laboratories using STRmix has grown eight-fold, from 11 to 89 laboratory systems. A decade of test results, multiple opportunities for code review, and exhaustive published commentary and analysis (both scientific and legal) all support a finding of foundational reliability. How the specific samples in this case were analyzed will be discussed in the following section.

IX. N.J.R.E. 702 ANALYSIS

The admission of expert testimony is governed by Rule 702 of the New Jersey Rules of Evidence. This court must also determine the admissibility of STRmix by following the Rule as follows: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise." N.J.R.E. 702.

This means that "(1) the intended testimony must concern a subject matter that is beyond the ken of the average juror; (2) the field testified to must be at a state of the art such that an expert's testimony could be sufficiently reliable; and (3) the witness must have sufficient expertise to offer the intended testimony. Landrigan, 127 N.J. at 413.

a. A Subject Beyond the Ken of the Average Juror

Neither the State nor the Defendant contend that DNA evidence, much less DNA evidence subjected to complex probabilistic calculations by a computer program, is a subject matter that the general public could easily grasp without expert assistance. While DNA in and of itself is not a novel technique, it is not simple to understand. "In this jurisdiction a subject matter that is so esoteric that it is beyond the ken of the average person typically qualifies as an appropriate subject for expert testimony." Kelly, 97 N.J. at 209. DNA evidence is a science, and the calculations run

by STRmix require explanation to a layman. Prior courts have found that the analysis of DNA, based on scientific and mathematical principles, is beyond the ken of the average juror. Doriguzzi, 334 N.J. Super. at 538. Given that the policy favoring expansive admission of helpful testimony is "imbedded in our jurisprudence," and the average juror is not "sufficiently familiar" with cutting-edge DNA sequencing, all factors favor admissibility. Berry, 140 N.J. at, 290-93.

b. Reliability as Applied

After applying the non-exhaustive Olenowski factors to STRmix, both as a software program and as an analytical tool employed by the laboratories that analyzed samples in this case, this court has found it to be reliable. Because this court finds STRmix to be sufficiently reliable under the Olenowski standard, it can form the basis of expert testimony in this trial. Scientific acceptability need "not be predicated upon a unanimous belief or universal agreement in the total or absolute infallibility of the techniques, methodology or procedures that underlie the scientific evidence." Romano, 96 N.J. at 80. For a program like STRmix, an applicability determination requires inquiry into how many contributors were present in a sample analyzed, how much DNA was available for analysis, and similar variables.

Methods found reliable under Olenowski must be reliably applied in order to assist a trier of fact. See Generally Muise v. GPU, Inc., 371 N.J. Super. 13, 55-57 (App. Div. 2004) (customer survey data, while a reliable methodology, not reliable when used to draw conclusions about dissimilar store locations). For an opinion to be admissible under N.J.R.E. 702, the expert must utilize a technique or analysis with "a sufficient scientific basis to produce uniform and reasonably reliable results so as to contribute materially to the ascertainment of the truth." State v. J.R., 227 N.J. 393, 409 (2017) (quoting Kelly, 97 N.J. at 210).

If a methodology is reliably applied, however, and the conclusion is merely opposed by a competing expert's different conclusion, this properly affects the weight and credibility of the evidence, not its admissibility. See State v. McGuire, 419 N.J. Super. 88, 126-27 (App. Div. 2011).

In Gissantaner, the court found that "reliable principles and methods underpinning STRmix were reliably applied" and that "doubts about the reliability of STRmix at low levels of DNA can be hashed out through cross-examination or testing by the defendant." Gissantaner, 990 F.3d 468 (internal quotation marks omitted). This does not mean, however, that STRmix is appropriate for all situations. It should not be used on "a complex DNA mixture that likely contained six contributors." Ortiz, 736 F. Supp. 3d at 901; but see Lewis, 442 F. Supp. 3d at 1147 (finding that STRmix was reliably applied when used to analyze DNA mixtures with two to five contributors).

Since ground truth cannot be known, and the ultimate finding of fact is for a jury, the court cannot comment on whether these samples are truly made up of related individuals. The 31-labs study shows that allele overlap increases as NOC increases, regardless of relatedness. Therefore, the problem of allele sharing can be addressed without reference to relatedness.

The basics of DNA have already been explained. Probabilistic genotyping and traditional DNA are no different in the extraction, quantitation, amplification, and detection steps. (1T 49-8). The difference is in interpretation. The goal of forensic DNA is to ascertain the genotype—the combination of alleles—that belong to each person's DNA located in a sample.

When there are multiple contributors to a DNA mixture, the challenge is to identify the profile, or genotype, of each separate contributor. "In single-source samples, only a single genotype is possible at each locus"; however, [i]n a DNA mixture, it may not be clear which genetic components, called alleles, belong to which contributor." Mixture Interpretation at 12, 34. To interpret a mixture, analysts count allele peaks, identify the number of potential contributors, and

estimate the relative ratio of the individuals contributing to a mixture. Fundamentals of Forensic DNA Typing at 325. Because the mixture ratio is approximately preserved in PCR application across loci, a contributor would be expected to give a similar amount of DNA at each locus—so if there are consistently smaller peaks that are about a third of the bigger peaks, an analyst would be inclined to conclude that the smaller ones belong to one contributor and the bigger to another. Id. at 326. However, peak heights and contributor ratios need to be critically analyzed.

"Analysis of samples containing very small quantities of DNA tends to produce [electropherograms] with a higher proportion of artifacts due to stochastic variation or random sampling of DNA molecules." Mixture Interpretation at 33. "[W]ith low-quantity DNA samples, the resulting profile and EPG may vary in how accurately they reflect the original sample, which can lead to loss of genotype information from a true contributor to the mixture." Id. at 42. "Furthermore, in part due to stochastic variation, two low-quantity DNA samples collected from the same surface can produce DNA profiles with different peak heights and therefore different ratios of alleles and possible genotype combinations. Analyzing the same low-quantity DNA mixture two or more times can also produce dissimilar DNA profiles with different degrees of stochastic variation[.]" Id.

The importance of contributor ratios depends on the specific proportions. A 3-person sample where every person's alleles are present at a similar ratio and generate a similar peak height on an EPG can be very difficult to deconvolute. Without a detectable separation in peak heights, any particular allele could belong to any particular contributor—there is no way to narrow down the possibilities. This is particularly true in the case of a sample where two parents and their child are equal contributors to a sample: this type of triad will often hide one contributor from detection. (6T 90-21).

Not every case involving related contributors poses a triad. First and second-degree relatives present different levels of challenges, ratios of major/minor contributors can lessen or exacerbate the difficulty, and the presence of a de novo mutation in one family member can completely sidestep the problem of allele sharing. (20T 16-19). While the challenge of related contributors is "ineradicable" Def. Br. at 63, it also varies based on the facts of any particular case. In this case, Buckleton testified that the challenges were lessened because "No triads are postulated . . . only dyads were being proposed[.] [A] dyad is two people and the proposals were [redacted] and Paul and we do not have insurmountable troubles with dyads." (7T 88-11-23).

Shared alleles also increase the risk that the NOC will be underestimated. (11T 54-11). However, while increased allele sharing runs the risk of falsely excluding a true contributor to the mixture, it does not bring an inherent risk of increasing the LR for true contributors. Bode Validation at 19. It increases the likelihood of underreporting true contributor LRs, and it also increases the probability that STRmix will not run at all. See ibid. STRmix will not run if a 4-person mixture shows more than 8 alleles, if a 3-person mixture shows more than 6, or a 2-person mixture shows more than 4, since such a result would be biologically impossible. Ibid.

Many of these issues, as already established, are not with STRmix. They are immutable facts of biology. Probabilistic genotyping systems in general, and STRmix in particular, were designed to try to interpret complex mixtures that humans could not reliably interpret. (7T 167-8). STRmix was designed to be "a way forward" in addressing these challenges. (10T 68-13).

STRmix has been shown to be that way forward and *can* deconvolute mixtures with overlapping alleles reliably. The 31 labs study shows that even 100% allelic overlap can be deconvoluted if peak heights are distinct and varied. 31 Labs at 19.

A general scientific principle that must guide this case is that "[a]ll scientific methods have limits." Mixture Interpretation at 11. "Reliability is not a yes or no question, but a matter of degree. Understanding the degree of reliability of a method can help the user of that information decide whether they should trust the results of that method in any specific situation when making important decisions." Id. at 15.

The defendant argues STRmix's limits should be "established in relationship to the features that make different DNA samples more complicated to reliably interpret." Def. Br. at 54. That is true. But to the extent that the defendant wishes to establish that the only types of limits that should be placed on STRmix are hard limits, that has not been proven. Hard limits can exist, but boundary areas where caution is warranted can also exist.

When discussing whether laboratories need to stay "100 percent in line" with every variable considered in their validation studies, Dr. Coble testified that

> Again, this is all dependent upon the laboratory. I think the point that I was trying to pull out or to highlight was that using a threshold [such as contributor percentage] like we are going to ignore everything less than five percent because I don't know, that was the number that was picked out of a hat, I'm not sure where that came from. But I think that they are throwing away valuable information. I think you have to look at the totality. You can't just draw a line at five percent or 30 picograms, you have to look at the totality of the results to determine whether this sample is reliable.
>
> [(11T 18-5)].

The hard limitation that all the testifying experts appear to agree upon is the NOC; this makes sense, as each additional contributor increases the complexity exponentially. See Ortiz, 736 F. Supp. 3d at 899.

Another limitation is peak heights: a lack of clear peak heights above the analytical threshold limits the useful information in a given sample. When at least one contributor's peaks rise above that threshold, then a sample can be analyzed; however, a debate exists as to how useful

PGS is at deciphering information regarding any additional contributors. Dr. Buckleton expresses optimism that contributors down to near-zero can provide at least *some* information. (6T 100-25).

The New Jersey State Police laboratory has also evolved past using peak heights as a limitation. Instead of looking at the RFUs, they now base suitability for analysis on the presence of resolvable data at a minimum of 7 loci. (9T 30-12, 9T 68-7, 26-13). Ms. Naughton believes that peak *ratios* are informative, even below the previous analytical threshold. An "analyst should scrutinize the data to ascertain if the peak height ratios indicate allele sharing to the extent that the deconvolution may not be intuitive." (3T 101-10, 125-2).

Dr. Buckleton's excitement about the future potential of PGS is noted, but the court views it with healthy skepticism in light of the other experts holding more conservative views—and the fact that Dr. Buckleton supports using PGS more conservatively than its suggested potential. Without going so far as to set a bright-line rule, the court does find that peak heights are still a relevant limitation to PGS's capabilities, and information found below a laboratory's analytical threshold should not be given more weight than it can support. It still has useful applications, such as determining the NOC. (4T 9-21 to 10-4).

Other complicating factors, such as allele sharing, drop-in, potential contamination or degradation, and other stochastic effects, can still be cause for caution but are extremely fact-dependent issues that call for experience and discretion. See Lewis, 442 F. Supp.3d at 1139, 1159-60 ("the complexity of the DNA mixture analysis [is not] merely defined by the number of contributors" but requires inquiry into the facts and circumstances relevant to the sample tested).

There is disagreement between the defendant and State experts as to whether template (in picograms) and contributor proportion (expressed as a percentage of the total mixture) should be the guiding metric, or whether EPG peak heights (measured in RFUs) are a more important metric

for determining a sample's suitability for analysis. The defendant argues that RFU is neither a "superior" or "unusually accurate" metric. Def. Br. at 129. The 31-laboratory study shows that varying peak heights are very relevant to STRmix's performance. Bright et al, 31 labs, at 19-20. Furthermore, SWGDAM guidelines dictate that samples used for internal validation be prepared using template and contributor proportion as their variables, suggesting that this is the proper guideline for testing in casework.

Dr. Coble, however, has explained that peak heights correspond to SWGDAM's recommendation that laboratories determine the "stochastic threshold" for their equipment, and that peak heights below that threshold will suffer a higher risk of drop-out. (10T 21-14). Because STRmix is fully continuous in how it addresses peak heights, levels below the stochastic threshold are not useless data points, but they are less informative. While STRmix addresses peak heights and drop-out rates dynamically, there is still "a correlation . . . between the quantity of DNA and the peak height that you observe in the electropherogram." (10T 107-6).

This court therefore finds that it is appropriate for laboratories to conduct their internal validation using the variables outlined in SWGDAM, but this does not render the use of peak heights in casework improper. SWGDAM has already approved of RFU as a metric in DNA analysis, and to the extent that the forensic science community accepts RFU as a good indicator of the quantity and quality of a sample, its use is proper here.

As to the individual laboratories incorporating STRmix into their laboratories, the court finds they did so according to the standard practices of the forensics community. Both Bode and the New Jersey State Police set appropriate limits for STRmix's use, developed SOPs to guide those limits, and adhered to them. Arguments about the propriety of those limits, the differences between hard and soft limits, and whether SOPs have too much/too little discretion can all be

addressed on cross examination. Ms. Reed's testimony about visually excluding [redacted] shows that SOPs can evolve as understanding develops. (5T 52-17).

The defendant argues that the published summaries "did not present any of the information necessary for a full, independent review by this Court of the reliability of STRmix in these laboratories," Def. br. at 133. However, "[t]he critical determination is whether comparable experts accept the soundness of the methodology, including the reasonableness of relying on this type of underlying data and information." Rubanick, 125 N.J. at 451. There is no evidence to support a finding that the validation studies provided are inaccurate, untrustworthy, or not of the kind produced by any laboratory that conducts an internal validation. They comply with the SWGDAM guidelines.

The defendant and his experts argue that they were forced to "sift through" the raw validation data in an insufficient timeframe, and that it is not their burden to demonstrate accuracy on the part of these laboratories. But defendant's counsel has been in possession of this data since May 22, 2024, and received 500 more contributor profiles on August 5, 2024. There was sufficient time to undertake whatever analysis they chose and has yet to produce anything to support a finding that either study was flawed.

Both developmental and internal validation are necessary to demonstrate the reliability of probabilistic genotyping systems in casework. (6T 95-11). As Ms. Ghannam testified, developmental validation must establish the fundamental validity of a method, not its outermost boundaries. (1T 21-2). The defendant misunderstands his own sources on this point, and claims that developmental validation must establish the outermost bounds:

> Developmental validation is "the acquisition of test data and determination of conditions and limitations of a new or novel" method used on forensic samples. Federal Bureau Of Investigation [sic], Quality Assurance Standards for Forensic DNA Testing Laboratories 3 (2011) (D-204). In other words, developmental validation is

> supposed to determine if and when a new technique produces reliable results. As NIST explains, developmental validation is necessary to show that a forensic test method is "fundamentally valid."
>
> [Def. Br. at 68.]

The key phrase here, however, is "fundamentally valid." While developmental validation can establish theoretical bounds, each forensic laboratory, through its technical leader, must establish the boundaries for their laboratory, as discussed infra and testified to by Ms. Ghannam. (1T 21-2). Developmental validation *can* identify limitations, but its primary aim is to demonstrate the proof-of-concept and fundamental validity of a method. The SWGDAM guidelines address developmental validation. SWGDAM Guidelines at 5.

The importance of internal validation, meanwhile, is not in dispute. As Ms. Naughton testified, internal validation studies are required by the FBI quality assurance standards. (3T 65-4). Internal validation studies establish a laboratory's ability to reliably use a technique, including probabilistic genotyping systems. (1T 36-16 to 22). Standards, guidelines, and best practices require the use of internal validations to gather data on the kinds of samples a laboratory intends to analyze. (3T 64 to 65). The SWGDAM guidelines are an example.

Internal validations are the method a laboratory is supposed to use to define their standard operating procedures. (3T 64-22). Internal validation studies are then interpreted by the technical leader, who develops standard operating procedures supported by the validation data. As NIST explains, "[v]alidations attempt to test samples reflective of casework, and SOPs use this information to provide a framework for analysts' tasks and steps." Human Factors at 28. Clear SOPs not only make sure the laboratory consistently implements a methodology, but they also "arm analysts with the tools needed to make educated and empirically supported decisions and reduce inter- and intra-analyst variability." Id. Together, internal validation studies and SOPs find and

define the limits of what samples can be reliably analyzed in a specific laboratory. This is also consistent with the FBI Quality Assurance Standards. FBI QAS 9.1 ("The laboratory shall have and follow analytical procedures supported by the internal validations and approved by the technical leader"); FBI QAS 9.6 ("The laboratory shall have and follow written guidelines for the interpretation of data that are based on and supported by internal validation studies").

None of the analysts who testified in this hearing felt it was appropriate, based on their validation studies, to analyze a sample with an NOC higher than 4. Other variables were more discretionary, such as stutter or unresolved peaks: for those, the analysts viewed them as less of a hard limit and more as something where it would be "important to see how STRmix [evaluated]" the data. (2T 32-1). Internal validation is about limitations, yes, but not always about hard limits.

Establishing limits, therefore, means understanding how an individual laboratory performs with complex samples and determining the levels of complexity where laboratories are confident in their results. This involves "demonstrating that a method performs well in scenarios more complex than the case at hand (e.g., test cases with more contributors, less DNA template, or more degradation)." Mixture interpretation at 99. Each variable tested, together or in isolation, increases the confidence that STRmix performs reliably within that laboratory and, when taken in the aggregate, inspires confidence in STRmix's foundational reliability. "[T]esting a large number of ground-truth known samples[] can inspire confidence that the method may perform well in scenarios like the case at hand." Id.

The internal validation process was done in Bode and the New Jersey State Police's studies, as well as thousands of times in laboratories across the country. (2T 72-18); (8T 21-10). A labor-

atory's test results are not a stand-in for another laboratory's own internal validations, but SWGDAM's guidelines contemplate laboratory tests being used as data for developmental validation and for testing reliability in general. SWGDAM Guidelines at 5.

It is possible for internal validation studies to study the impact of relatedness on a probabilistic genotyping system's reliability. As Ms. Naughton testified, Bode's internal validation did "not directly [study] related contributors," so it is possible that the internal validation could have. (3T 101-25). Allele sharing can occur between unrelated and related individuals. 31 Labs at 20. The presence of related individuals will not render a sample *per se* unreliable. In the validation study cited by the defendant, none of the inclusionary LRs generated for the sibling non-contributors were higher than that of the true contributors. ex. D-108 at 29. Based on the fundamental principles of DNA, there is every reason to believe a non-contributing sibling of a contributor would potentially generate an inclusionary LR.

Applying those principles to the available data from relatedness studies, there is no evidence to suggest that the LRs generated where siblings are involved are erroneous, but any context involving allele sharing may be less informative. Again, this challenge would go to the weight of the evidence, and not to its admissibility

The defendant has argued that since STRmix calculates LRs for contributors one at a time, it "cannot account for the possibility that one of five brothers could have contributed to a mixture." This could raise a challenge in some situations. If only one of the brothers was able to provide a sample, it would be impossible to generate comparative LRs for each brother, which could impact the single tested sibling's LR of important context. A LR calculated without accounting for the

presence of siblings could provide criticism of the results. See In Defense Of, The Kinship Problem, https://indefenseof.us/issues/kinship-problem (last visited February 14, 2025). This area would be ripe for extensive cross-examination if presented unmoored from any context.

This does not mean that allele-sharing leads to inconsistent or inaccurate results, just to less informative ones. While allele-sharing between contributors can influence the calculated LR, "NOC . . . is the main driver to LR change." 31 Labs at 20. "[N]on-contributors are unlikely to yield large LRs even if they share many alleles with the true contributors." Ibid. This is partly because STRmix can rule out potential contributors by the peak heights on the EPG, and partly because "this apparent trend" of more readily assigning non-contributors based on allele sharing "is totally confounded by the NOC to the mixture." Put another way, STRmix has an easier time dealing with a high degree of allele sharing amongst fewer people (as would be expected with relatives) than it does dealing with more people and the random allele sharing that may occur. Id. at 20, fig. 11. As long as the peak heights are varied, even shared alleles can be assigned with some degree of confidence.

The 31 Labs study also suggests that—although false inclusion is always a risk when a person of interest shares alleles with a contributor to a mixture—STRmix is, in fact, better at handling that problem. "An individual that shares 100% of alleles with the other contributors to a mixture can still have their genotype resolved completely, based on peak heights, given the right circumstances[.] The ability to use peak heights in this way is one of the main drivers for the differences in LRs produced between fully and semi-continuous systems." Id. at 19-20 (emphasis added). While STRmix cannot handle every situation involving allele sharing, it can deconvolute mixtures involving 100% shared alleles (e.g., two parents and their child) if variations in peak heights are distinct enough to provide another way to differentiate between individuals present in

the sample. This separation of peak heights would also be detectable by visual inspection, meaning that even though an analyst may not be able to deconvolute a mixture unassisted, they can still perform some degree of quality control on the output.[21] In short, they can tell if it makes sense even if the calculated LRs couldn't be reverse engineered.

While the LRs reported were, as defendant brings to our attention, postulated based on the alternative proposition that either the person of interest or an unknown, *unrelated* individual contributed to the samples, this does not mean that the results reported are unreliable. They are consistently reported. All the persons tested against the samples are the first-degree relative of at least one other person. Paul is related to [redacted] [redacted] is related to Paul, [redacted] and [redacted] The alleles he shares with each of those persons varies. [redacted] is related to [redacted] and [redacted] Each of the children are related to their parents and sibling.

These relationships do not all pose intractable challenges. Second-degree relatives do not pose the same challenges. In samples where Paul and [redacted] were both analyzed, they both generated LRs based on the same alternative proposition. This means that the numerator, their personal LR number, is different, but the denominator, the possibility of the DNA coming from an unrelated person, is the same. This means their generated LRs are one-to-one comparable. They have context. They can be explained to a jury.

Due to [redacted] de novo mutation, she is always readily distinguishable from other family members: no one shares that allele with her. (5T 82-20). This means that her LRs can be explained without concerns that the presence of family members altered the calculations as they relate to her.

---

[21] See 31 Labs at 20 (a major contributor "may be clearly resolved by simply 'eyeballing' the electropherogram").

As explained supra, the greatest challenge for STRmix is related contributors who are present at roughly equal ratios. None of the samples analyzed met that criteria. The only sample with two contributors at near equal ratios was E05a, which had two minor contributors at roughly the same proportion. Those minor contributors were deemed unsuitable for comparison. The Major contributor made up more than 90% of the mixture in that case, and calculations for that contributor are not affected by uncertainty at the lower end.

Because the potential challenges presented by relatedness either did not present themselves in this case, were mitigated by mutations, or can be properly put into context for a factfinder, and because there is no evidence that any of these samples were 5-person contributors misidentified as 4-person, and because all the samples met their respective laboratories' SOPs for suitability, this court finds that STRmix was reliably applied in this case. Concerns about LRs generated for contributors with very low template amounts or very low RFU are well-founded but should be explored during cross-examination so that a jury may understand how to properly give them weight.

The court finds no potential complications posed by samples with related contributors. These samples do not involve 3-person mixtures with equal peak heights. They also do not involve apparent 5-person mixtures, and there is no reasonable hypothesis that there are any 6-person mixtures. The risk of three first-order relatives masking an apparent contributor is avoided by [redacted] mutation in her DNA profile.

The risk of unreliable results in these samples has been further mitigated because the samples are being viewed under the assumption that the contributors are all related to one another. The danger of an undetected familial relationship skewing the results is not an issue. A chance of inflated inclusionary support may exist, but such a theory goes to the weight of the evidence.

i. Bode's Validation and analysis of the samples

Bode Technologies was required by ANSI/ASB Standard 018 to validate STRmix before it was used in casework. Standard for the Validation of Probabilistic Genotyping Systems at 3 (2020) (hereinafter ANSI/ASB Standard 018). In its validation study, Bode tested samples of similar complexity to the samples it analyzed in this case. Bode assembled simulated 2- 3- and 4-person mixtures using ground truth samples. Bode designed samples to reflect minor contribution amounts of varying levels, as low as 25 picograms of DNA. Bode Technologies, Internal Validation of STRmix Version 2.5.11 with GlobalFiler™ and 3500xL, October 2018 at 4 (hereinafter Bode Validation). Bode also designed samples to reflect minor contributions of varying ratios, down to 5% of the total template DNA. Ibid. Approximately 20 samples were created at or around 5%. 6 were removed. 14 were tested. One was shown to be at 3%. Three others were at or around 5%. (T2 119-19 to 120-16; 120-24 to 124-15; T3 25-6 to 26-25). For some samples, Bode analyzed peak heights down to 0 simulated RFU, but has explicitly stated that "[t]he data is also only validated to be interpreted at analytical threshold of 125 RFU at this time." Bode Validation at 32. A potential genotype must have at least three peaks above 125 RFU to be assigned to a contributor. (4T at 10-16).

Bode testing samples at low template levels is consistent with ANSI/ASB Standard 018. Standard 018 at 3. Standard 018 requires for internal validation studies to include "compromised DNA Samples" like low template, degraded, and inhibited samples. Standard 018 at 3. The internal validation shall include "the range of actual casework samples intended for analysis with the system at the laboratory." Id. As Bode expects to see samples with low template DNA at their laboratory, they met this requirement of Standard 018. Id. Bode's internal validation included an examination "of single source samples, 2 to 4 person known mixtures, and mock evidence samples across a range of input DNA and mixture ratios." Bode Validation at 1.

Because of the variability inherent in amplification, mixture proportions necessarily have a margin of error in their calculation. This margin of error is a fact of biology, not software, and does not necessarily undermine reliability.

Bode's validation study does show discrepancies between the mixture proportions deliberately created by Bode and STRmix's estimation of those proportions. See generally Letter from Kristen Naughton to Christopher Decker, April 26, 2024 (entered as exhibit D-20). As mentioned above, this is an inherent fact of biology. None of the samples in this case involve both less than 25 picograms and a <5% minor contributor, meaning that no sample exceeded both boundaries at once.

Testing mixture proportions is consistent with the ANSI/ASB Standard 018. This standard requires internal validations to evaluate appropriate sample types regarding NOC, mixture ratios, and template quantities that a laboratory will be evaluating. ANSI/ASB Standard 018 at 3. In Section D of Bode's internal validation, specificity and sensitivity were tested. Bode Validation at 13. Specificity and sensitivity were tested by calculating the LR for "25 two-person, 15-three person, and 15-four person mixtures," in which these profiles represent the type encountered by Bode with "varying DNA quantity and mixture proportions." Id. Section D also helps to satisfy another requirement with ANSI/ASB Standard 018 where internal validations shall address "accuracy, sensitivity, specificity, and precision." ANSI/ASB Standard 018 at 3. Accuracy was addressed by Sections A, B, and C of the internal validation. Bode Validation at 4. Precision was specifically addressed in Section M of Bode's internal validation. Bode Validation at 28. In Section M, data from LR of each replicate demonstrated that "running the same sample through STRmix 10 times generates the same statistical conclusion." Bode Validation at 29.

The defendant has argued that in its validation study, Bode did not validate any of the LRs used to hypothesize the real contributor is a relative of the person of interest. This does not, however, change the reliability of the LRs that were reported in this case. It would be inappropriate to run a feature of STRmix that is unvalidated, such as a related LR or peak heights below the analytical threshold. See Bode Validation at 32. Bode did, however, study a similar set of hypotheses centered on "conditioning" the analysis on the presence of a known individual. This means testing LRs generated under two different propositions. Proposition 1 is that the DNA sample came from a person of interest, a specific known individual, and *N-2* unknown individuals (where *N*=the total NOC), as compared to the DNA coming from a specific known individual and *N-1* unknown individuals. Bode Validation at 15. Proposition 2 is that the sample came from a person of interest and *N-1* unknown individuals, as compared to *N* unknown individuals. Ibid. When these two propositions were compared, and the identity of the major contributor was assumed to be known, LRs increased. See id. at 16 fig.7a (reproduced in appendix). This demonstrates "that the addition of more relevant information (such as the addition of assumed contributors) is shown to improve the performance of STRmix." Id. at 16. When *every* contributor to the sample was assumed to be known, LRs increased significantly. See id. at 17 fig.7b (reproduced in appendix).

The court finds that a related LR may give a different level of inclusionary or exclusionary support than an unrelated LR, just as conditioning an analysis with known individuals will uniformly increase LRs. There is evidence that this is a consistent and reliable trend that can be addressed during testimony, and meaning and import of that trend is a question of fact that is properly before the jury. This does not make the expression of either hypothesis inaccurate or unfounded.

ANSI/ASB Standard 018 requires internal validation studies to evaluate multiple propositions for cases which shall consider the effect of overestimating and underestimating the NOC.

Standard 018 at 3. Bode conducted experiments to determine what occurs when the NOC is underestimated for a mixture. Bode Validation at 19 fig.9 (included in appendix). Their summary findings indicated that

> when you underestimate the number of contributors, STRmix may deconvolute incorrectly resulting in false exclusions. For 17 of the known contributors there was no significant effect on the LRs. For 8 known contributors the LR was significantly reduced. Underestimating the number of contributors generally results in lower LRs for H2 true comparisons. Note that STRmix will not run if there are peaks present that cannot be explained using stutter modelling or drop-in and can only be explained via an extra contributor being present in the mixture.
>
> [Id. at 19.]

The validation also found that "For mixtures with a lot of allele sharing that do not exceed 4 alleles for 2 person mixtures, 6 alleles for 3 person mixtures or 8 alleles for 4 person mixtures it will run." Ibid.

Bode's validation study also included the analysis of 15 mock samples that had previously been used for manual interpretation training. Id. at 24.

> Three of the samples (#4, #5 and #15) were interpreted by the analyst as ≥2 contributor mixtures and only the major contributor was deduced from the mixtures. STRmix deconvoluted full profiles for the major contributor for all three samples and the profiles were the same as those deconvoluted in the manual analysis. Two samples (#5 and #15) were run as 3 person mixtures and no minor alleles were deconvoluted for one sample (#5) and limited minor alleles were deconvoluted for the other (#15). Sample #4 was run in STRmix as a 2 person mixture and interpreted ≥99% confidence only 6 partial loci for the minor contributor. However, it was noted prior to running that the minor alleles did not have sufficient data for comparison purposes. The profile deduced for the expected single minor component could potentially be from multiple donors or a different donor all together. This sample was a touch DNA sample collected for a training set and it could have contained DNA from unexpected donors.
>
> [Id. at 27.]

The court finds this to be highly probative of two facts: that STRmix not only produces results consistent to manual interpretation of ground truth samples, but that there are many instances where a majority of STRmix's output can be independently verified.

Bode has two hard limits established on the use of STRmix: (1) No more than four contributors (3T 68-13 to 19; 4T 8-22 to 23); and (2) "genetic representation at least three testing locations or loci in order to deem that component of the mixture interpretable." (3T 91-1 to 8). Bode will not analyze five-contributor samples with STRmix. (2T 39-23).

Quality assurance parameters, analytical procedures, and interpretation protocols shall be derived from internal validations studies. Standard 018 at 3. Bode's internal validation of STRmix was done to help meet the FBI QAS and SWGDAM guidelines for probabilistic genotyping software. Bode Validation at 1. Bode's has Standard Operating Procedures ("SOP"), and they note increased risks of false positives on the kinds of samples analyzed in this case. Bode's SOP recognizes that mixtures compromised of relatives are particularly hard to reliably interpret. The SOP explains that an "evidence profile may be unsuitable for interpretation if it appears there is a mixture profile from first-degree relatives." Bode Technologies, BTF00278-Statistical Interpretation-STRmix, March 2021 at 9 (hereinafter Bode Interpretation SOPs).

Bode conducted visual exclusions in this case instead of running all relevant people through STRmix. The question is whether that was a reliable and acceptable exercise of discretion at the time, not whether the court should second-guess the decision. While the defendant characterizes this choice as inappropriate, it does not impact STRmix's reliability. While Bode may not visually exclude a person from a sample based on their understanding of STRmix today, the fact that they did so years ago is not rendered improper.

Bode's Standard Operating Procedures also acknowledge the challenges of samples at the edges of their validation studies and appropriately advise caution. None of the samples analyzed by Bode Technologies in this case exceeded their hard limitation on NOC or minimum loci. The credible testimony of the witnesses also establishes that the analysts at Bode did properly exercise caution and discretion in analyzing and reviewing this case; Ms. Reed's initial confusion when discussing samples during the STRmix demonstration, her ability to recover her composure, and her clarifications of policy show that Bode analysts continue to test and reevaluate their understanding and interpretations of their SOPs. She acknowledged the possibility she could have been mistaken, took a careful look at her data, and explained her reasoning. This reassures the court that these samples were not improperly analyzed. (5T 50-5 to 53-25).

The Bode Technologies samples that are the subject of defendant's objection are summarized in the appendices. In this case, there were fifteen evidentiary samples submitted to Bode Technologies.[22] At Bode, they were assigned to forensic DNA analyst Danielle Reed for testing and analysis.

Two additional samples were tested and analyzed by Reed without the use of STRmix. Sample E01b was the interior waistband of a pair of jeans, and it consisted of a partial profile for at least one male. E01b was deemed inconclusive. The second sample was E02a, the interior of glove 1. This sample was tested without STRmix and the results indicated two contributors. The major contributor was a female who matched the DNA profile of [redacted] No conclusions could be made for the minor contributor. This result is consistent with the STRmix deconvolution

[22] See Appendix, Table 1 (numbering from original).

of sample E02b, which came from the same glove and generated a LR for [redacted] as the major contributor of 1E+24:1.

Although the ground truth of a casework sample cannot be known, the samples from glove one were tested using two different methods of analysis and yielded complementary results. This demonstrates that STRmix's outputs are, in both ground truth samples and some parallel casework investigations, consistent with long-accepted scientific procedures. This demonstrates another way STRmix can be tested and corroborated and supports a finding of reliability as applied.

ii. The New Jersey State Police Validation and Analysis of Samples

In its validation study, the New Jersey State Police laboratory tested samples of similar complexity to the samples it analyzed in this case as required by ANSI/ASB Standard 018. Standard 018 at 3. They assembled simulated 2- 3- and 4-person mixtures using ground truth samples, and designed samples to reflect minor contribution amounts of varying levels, as low 2% of the total mixture and 6.25 picograms of DNA. New Jersey State Police Office of Forensic Sciences DNA Laboratory, Internal Validation of STRmix V2.8 14 (2021) (hereinafter NJSP Validation).

The New Jersey State Police laboratory determined that, in accordance with SWGDAM guideline 4.2.1.2, "[t]emplate amounts of 0.125 ng and above resulted in unambiguous genotype weightings (i.e. all 100% weight)." NJSP Validation, at 4-5. This indicates that, for samples above 125 picograms, STRmix performed not only as expected, but also performed as a traditional analysis would for a single-source profile of similar weight.

Although the New Jersey State Police did not conduct a separate relatedness study, they did include related contributors in their sensitivity and specificity section. They prepared two- to four-person mixtures with known contributors that were intended to

> cover a broad range of template amounts and mixture proportions and are likely to be representative of DNA profiles recovered during casework analysis. The contributors include homozygous and heterozygous alleles and there is varying amounts of allele sharing across the different loci, which was accomplished by using relatives within mixtures as well as maximizing allele sharing amongst contributors.
>
> [NJSP Validation at 13-14; See also id. at 14, table 3 (reproduced in appendix)].

They conducted over 80,000 LR comparisons using these prepared mixtures. Id. at 15. The New Jersey State Police also studied the effects of degradation on STRmix's sensitivity and specificity. "A subsequent sensitivity study was completed with 24 samples (five single source, five two-person mixtures, ten three-person mixtures and four four-person mixtures) that purposefully contained degraded components." Id. at 18. ANSI/ASB Standard 018 requires internal validations to address sensitivity and specificity by using samples that are like the actual casework the system will be testing at the laboratory. Standard 018 at 3.

Similar to Bode Technologies, the New Jersey State Police conducted studies on the effects of alternative hypotheses on the generated LR. By "conditioning" the hypothesis and generating a LR that evaluates the ratio between a sample coming from a specific known individual and *N-2* unknown individuals as compared to the DNA coming from a specific known individual and *N-1* unknown individuals, the New Jersey State Police validation study found that LRs for true contributors uniformly increased. Id. at 24, figure 7 (reproduced in appendix).

The New Jersey State Police validation expressly addressed the function of STRmix's secondary diagnostics, including the GR diagnostic. Id. at 72. The study found that higher numbers and a stronger suggestion of non-convergence were more likely to occur in 4-person samples. "Notionally, values above 1.2 indicate that the chains may not be nearing convergence. It is important to note that the GR diagnostic output by STRmix™ is a summary statistic: values less than

1.2 do not guarantee that all parameters have converged whilst values greater than 1.2 do not necessarily indicate that the results are unreliable." Ibid.

When comparing STRmix results to the results of traditional analysis on previously interpreted known two person mixtures yielded consistent and promising results. "In all scenarios, the STRmix software returned a stronger LR. This is to be expected when moving from a binary to a probabilistic model, where the additional information utilized will strengthen the match to a true contributor." Id. at 46.

Internal validations are meant to help design quality assurance parameters, analytical procedures, and interpretation results. Standard 018 at 3. Informed by this validation study, the New Jersey State Police SOPs establish two limits on suitability for STRmix analysis. The first is a NOC of 4 or less, the second is that the sample contains information on at least 7 loci. (9T 26-22). The New Jersey State Police Standard Operating Procedures rely on human experience when analyzing and reporting mixtures comprised of related individuals. The Standard Operating Procedures for the New Jersey State Police does not have a requirement for how many loci a specific mixture component is supposed to have. (9T 46-24). Altogether, defense counsel and Ms. Thayer agree that the Standard Operating Procedures are objective. (9T 22-5).

As Ms. Thayer testified, the New Jersey State Police will rely on the DNA analyst's intuition when interpreting STRmix results. (8T 97-15). One diagnostic an analyst "can intuitively assess" is the STRmix weights that are in the interpretation results. (8T 97-15). For intuitive diagnostics, if an analyst sees a result they were not expecting, further investigation is warranted. (8T 98-21).

This court finds that the New Jersey State Police conducted their internal validation study in accordance with SWGDAM guidelines and ANSI/ASB Standard 018, and that the results from

the study were consistent, reliable, and gave a clear understanding of STRmix's capabilities. At no point did STRmix generate a contraindicatory result when compared to traditional analysis. There were two samples from the New Jersey State Police Laboratory discussed in this hearing. Both are summarized in the appendices. For both samples, STRmix was used consistently with the New Jersey State Police's SOPs and well within the boundaries of their validation study. The New Jersey State Police prepared both two- and four-person mixtures with an expected minor contributor as low as 2% and contributing only 6.25 pg of DNA. NJSP Validation at 47, 66.

Defendant argues that for sample 6-1-4-1, it is improper for an analyst to "convert a genotype probability into a definite statement of source attribution." Def. Br. at 169. The court does not find this has occurred. Ms. Schlenker did testify that she attributed [redacted] Caniero as the major contributor to sample 6-1-4-1, (9T 84-5) she did not do so by improperly converting an LR from STRmix into a definitive statement. She did so by calculating an RMP using traditional analysis of an allele chart and bypassing STRmix entirely, and then used STRmix to analyze potential minor contributors. (9T 84-14 to 85-15).

The court does not find RMPs to be the subject of this hearing. However, the court does find that, for expert testimony regarding LRs generated from STRmix to be helpful and not confusing or misleading to a jury, it must be clear which conclusions are based on information derived from STRmix, and which conclusions are based on traditional DNA analysis. An RMP is making a different assertion than an LR, and they are generated independently. Great care must be taken to ensure the capabilities and limitations of the two are not conflated.

c. Qualifications and Expertise of the Witnesses

Although not all the experts that testified in this hearing possessed the same degree of credibility, all of them were sufficiently qualified in their respective fields to offer an opinion. The

bar for qualification as an expert is low. "[C]ourts take a liberal approach when assessing a person's qualifications. Our case law is replete with examples of the generous approach taken by our courts when qualifying experts based on training and experience." Jenewicz, 193 N.J. at 454; See also State v. Moore, 122 N.J. 420, 458 (1991) (police officer with two years of experience as a crime-scene investigator was qualified as an expert on blood-splatter, despite only two days of experience with blood-splatter analysis).

The baseline for the admissibility of expert testimony is that the witnesses "have sufficient expertise to offer the intended testimony." Accutane, 234 N.J. at 349. DNA analysis, the finer points of engineering and developing software, and internal and developmental validation studies are all beyond the ken of the average juror, and all the experts who qualified in their respective field therefore satisfy that requirement. The witnesses' understanding and insight into STRmix as it related to their respective fields, however, varied. Those who showed a deeper knowledge and provided consistent testimony on direct and cross-examination provided opinions that were more helpful to the Court.

The defendant raised a preliminary objection Dr. Buckleton's qualifications as an expert in software development, subject to further voir dire on their cross-examination. Though the defendant correctly points out that Dr. Buckleton was offered as an expert in software "development" and not software "engineering," this is not the critical distinction. It is clear from his testimony that Dr. Buckleton's day-to-day involvement in STRmix has little to do with writing the actual code and far more to do with the bigger-picture tasks of determining features for new versions, promoting the scientific reliability of the program, and overseeing ESR's general scientific mission.

Although Dr. Buckleton has no degree in computer science and has only been involved in the creation of one software program in his career, he is sufficiently experienced in the coding, development, and engineering of STRmix. An "expert may be qualified on the basis of his experience, even when it is limited." State v. Torres, 183 N.J. 554, 572 (2005). The fact that Dr. Buckleton has not written code in 20 years does not mean he is incapable of understanding or offering an opinion on software.

Similarly, although Mr. Adams possesses only a bachelor's degree in computer science, his work history and his experience with STRmix and competing software platforms is enough to render him competent to offer an expert opinion. See Moore, 122 N.J. at 458. The fact that Mr. Adams did not successfully complete his code review does not diminish his qualifications as an expert or his capability to form an opinion; rather, it goes to the underlying factual basis and thus the weight of his opinion.

The defendant also stated in their post-hearing submission that they reserve the right to challenge the qualifications of Dr. Buckleton, Ms. Schlenker, and Ms. Reed at trial. The court finds them to be sufficiently qualified now. All the experts were subject to voir dire and their qualifications were discussed and found sufficient to support the testimony they offered. This is not to say that the defendant cannot further question any testifying experts on their qualifications or expertise if they testify at trial, only that they are qualified to form opinions at this stage.

> That the strength of an individual's qualifications may be undermined through cross-examination is not a sound basis for precluding an expert from testifying . . . even if it likely will affect the weight that the jury will give the opinion. Rather, a court should simply be satisfied that the expert has a basis in knowledge, skill, education, training, or experience to be able to form an opinion that can aid the jury on a subject that is beyond its ken.
>
> [Jenewicz, 193 N.J. at 455.]

While the qualifications of an expert are generally viewed expansively, credibility is a separate determination. See Olenowski II, 255 N.J. at 573 (referring to the issues of qualification and credibility as both being within the discretion of the trial court). While this court finds that all experts were qualified to offer opinion testimony, their individual credibility determinations varied as discussed.

The court did not find that personal, professional, or financial ties to STRmix necessarily impacted credibility. Witnesses on both sides may or may not have motivations to see STRmix succeed or fail that are outside the scope of this hearing, and "the 'hired gun' argument . . . cuts both ways." Rubanick, 125 N.J. at 453. At bottom, this case is resolved based on the underlying scientific principles and their application to this case, and the most relevant factor to a witness's credibility is whether they supported their opinion with good scientific grounds, or whether their statements were cursory and conclusory.

d. The Fundamentals of DNA Analysis, Likelihood Ratios, Human Cognition, and Cognitive Bias are Not the Focus of This Hearing

"[P]robabilistic genotyping software[] marks a profound shift in DNA forensics." Pickett, 466 N.J. Super. 270 at 276. Although PGS is relatively new and unlocks the ability to analyze samples that were previously impractical to use in casework, it is still part of forensic DNA analysis and must be subject to the same empirical tests. As the defendant has acknowledged, "the reliability of traditional DNA analysis does not govern this case. But the principles underlying forensic DNA analysis apply." Def. br. at 53.

The defendant has submitted that the human decision-making process is essential to the reliable implementation of STRmix. Human analysts make decisions about STRmix inputs. All PGS, including STRmix, rely on human analysts to spot errors. STRmix is an investigative tool,

but DNA analysts must still use their experience and independent judgment in the overall analysis. DNA analysts use cutting edge software and technology throughout their career. STRmix software may be new, and it may have challenges, but this court finds that DNA analysts are using it responsibly, reliably, and in the limits of its capabilities.

The court agrees and does find that "human judgement is essential to the use of STRmix" within a laboratory and is relevant to STRmix's as-applied reliability determination. Def. Br. at 83. This court finds no competent evidence that "[h]uman analysts . . . are subject to cognitive bias and other limits" in their use of STRmix. Ibid. If STRmix is reliable and foundationally valid, then addressing the human component requires "similar considerations as for DNA analysis of single-source and simple-mixtures samples[.]" PCAST Report at 82. Those considerations already account for human error:

> Because errors due to human failures will dominate the chance of coincidental matches, the scientific criteria for validity as applied require that an expert (1) should have undergone rigorous and relevant proficiency testing to demonstrate their ability to reliably apply the method, (2) should routinely disclose in reports and testimony whether, when performing the examination, he or she was aware of any facts of the case that might influence the conclusion, and (3) should disclose, upon request, all information about quality testing and quality issues in his or her laboratory.
>
> [Ibid. at 75.]

All these considerations can be addressed during cross-examination of the individual analyst before a jury.

No evidence of cognitive bias has been shown by the defendant. DNA analysts are independent and there is no evidence of any bias, interest, or motive in the outcome of their findings. Defendant's argument is purely speculative, and the court finds that the concept of cognitive bias has no bearing on the reliability determinations before it. It is true that "[g]reater cognitive effort is required when results are complex and data are ambiguous, when there are time pressures, when

a large amount of information must be combined and processed, or when decisions are discretionary." Human Factors at 12. This is not a reason to remove human decision-making from the process, it is an example of when human decision-making is most needed.

The court found no evidence of bias due to financial interest. Dr. Buckleton testified in a limited capacity about licensing sales as he does not have firsthand knowledge of this information. (7T 124-18). STRmix pays 10 percent of their licensing sales to FSSA, which is around one million U.S. dollars per year. (7T 125-4 to 126-2). Dr. Buckleton is a civil servant and salaried, so the sales made from STRmix will not incentivize him. (6T 46-21 to 47-3). Although defendant has suggested that there is financial incentive for the laboratories who have purchased STRmix to recoup their investment, defendant has also argued that the cost to purchase and run STRmix is insignificant compared to the cost of a capillary electrophoresis machine. (20T 50-14).

Even though STRmix's outputs are reliable and admissible, a complete analysis requires a brief discussion of N.J.R.E. 402. Even relevant and probative evidence can be excluded if its value is "substantially outweighed by the risk of . . . confusing the issues or misleading the jury" N.J.R.E. 403. This court finds that LRs are complex and require careful explanation to jurors, but they are not new. LRs are an established method of communicating the relationship between two hypotheses and predate STRmix. (1T 63-20). ESR did not develop the concept of LRs or incorporate them into STRmix in a new or novel way. Like the well-established fundamentals of DNA analysis, LRs are not subject to reevaluation at this time. It is fair to say that weight given to LRs can and will be litigated, but their admissibility is accepted. Compare Spann, 130 N.J. 518 (acknowledging LRs as a relatively new method of reporting DNA evidence in 1993, and outlining challenges in reporting them to juries, the New Jersey Supreme Court was nonetheless "inclined to believe that appropriate jury instructions can cure all of them") with Rochat, 470 N.J. Super. at 440 (acknowledging,

in 2022, that the concept of LRs may be generally accepted even if a program that generated them was not) and Gissantaner, 990 F.3d 470 (LRs did not necessarily run afoul of F.R.E. 403, but concerns that a "jury might misunderstand what the likelihood ratio means could require advocates to describe it in a way that will not generate unfair prejudice or mislead the jury"); See also Lewis, 442 F.Supp.3d at 1140 (finding that a "likelihood ratio is not a statistic of inclusion or exclusion" and was admissible, but nevertheless it "is very important that the likelihood ratio be precisely communicated to the jurors lest it mislead them as to its significance"). Even defendant's expert Mr. Inman recognized that verbal scales and LRs were not something unique to reporting PGS results; his chief complaint was directed more to the practice of taking unique LRs and broadly categorizing them by putting them "in a bin." (19T 88-2). Mr. Inman would like to see a shift away from verbal scales and towards taking "a little more time to understand the concept that's really not that hard to understand." (19T 88-18).

The court therefore finds that LRs and their evidentiary weight have been recognized for over 30 years. They are relevant and probative, and their value is not "substantially outweighed" by the risk of misleading a jury under R. 403 so long as exactly which conclusions an LR supports—and which it cannot support—are explained.

X. OVERALL CONCLUSION

The Olenowski factors are not a "rigid" checklist. Olenowski II, 255 N.J. at 584. They inform an inquiry that, while flexible, must nevertheless be rigorous. The court finds that STRmix has withstood this inquiry. There are sufficient standards that guide its development, its validation, and its application. The fact that many standards are not binding may matter from a regulatory standpoint, but they do not prevent the relevant scientific communities from being guided by them. STRmix's publications have been peer-reviewed, and the court has found no precedent to inform

a finding that independent authorship is required. If it were not permissible for scientists to report their own discoveries or inventions, scientific progress would grind to a halt. This has never been the case before, and it is not the case now.

STRmix has been tested. It has been tested over 9 billion times (6T 98-20). Its performance across a vast range of samples has been extensively charted. See generally 31 Labs. Every laboratory that brings STRmix online tests its performance for itself. How LRs trend towards the uninformative range as the available information in a sample decreases is predictable, and, while it would be time-consuming to manually check all outputs from STRmix, it is possible to fully calculate any given output. "The fact that a possibility of error exists does not preclude a conclusion that a scientific device is reliable." Romano, 96 N.J. at 80. The numerous articles, the discussions and spirited debate between forensic experts as to PGS's role in their industry, and the thoughtful and thorough judicial opinions from other jurisdictions all lead to the same conclusion: probabilistic genotyping is reliable. STRmix implements it reliably. STRmix works, and it appears to work very well.

For these foregoing reasons, defendant's motion to exclude STRmix v. 2.5.1. and v. 2.8.0's DNA results is **DENIED**. The challenges to the STRmix evidence raised by the defendant do not go to the admissibility of the evidence, but to the weight afforded. Specific doubts as to the impact that related contributors have on the calculation of LRs, questions about the specific adherence to SOPs at the laboratories in question in this case, and general criticisms of PGS as a whole and STRmix in specific are properly put before the jury and are appropriate fodder for cross-examination.

XI. APPENDICES

a. Samples tested by Bode Technologies

The results of Reed's findings are as follows:

E01a: *Jeans, interior thigh above knee*. The NOC was three. Contributors 1 and 2 were suitable for comparison. Combination was possible.

Contributor 1: 82% contribution, 4038 RFU. LR for [redacted] Assuming there was a mixture of three, this mixture DNA profile obtained is at least 470 sextillion times more likely to occur if it originated from [redacted] and two unknown, unrelated contributors than if it originated from three unknown, unrelated contributors.

Contributor 2: 17% contribution, 844 RFU. LR for Paul: Assuming the sample had a mixture of three, this mixture DNA profile obtained is at least 3.7 quintillion times more likely to occur if it originated from Paul and two unknown, unrelated contributors than if it originated from three unknown, unrelated contributors.

Contributor 3: 1% contribution, 50 RFU. [redacted] [redacted] and [redacted] visually excluded from suitable contributors.

E01c: *Jeans, exterior front right thigh*. The NOC for this sample was one. Contributor 1 is suitable for comparison.

Contributor 1: 100% contribution, 2300 RFU. LR for [redacted] Assuming one contributor, this DNA profile obtained is at least 2.1 septillion times more likely to occur if it originated from [redacted] than if it originated from an unknown, unrelated individual. [redacted] [redacted] [redacted] and Paul were visually excluded.

E02b: *Glove 1, interior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison and combination was possible.

Contributor 1: 94% contribution, 1325 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 1 septillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 6% contribution, 91 RFU. Assuming a mixture of two, this mixture DNA profile obtained is at least 43 thousand times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. [redacted] [redacted] and [redacted] were visually excluded from suitable contributors.

E03a: *Glove 2, exterior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison and combination is possible.

Contributor 1: 59% contribution, 1922 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 100 quadrillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 41% contribution, 1325 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 45 quadrillion times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. [redacted] [redacted] and [redacted] were visually excluded from suitable contributors.

E03b: *Glove 2, interior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison and combination was possible.

Contributor 1: 86% contribution, 600 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 750 quintillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 14% contribution, 100 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 400 thousand times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is 480 times more likely to occur if it originated from two unknown, unrelated individuals than from [redacted] and one unknown, unrelated individual. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is 37 thousand times more likely to occur if it originated from two unknown, unrelated individuals than from [redacted] and one unknown, unrelated individual. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is 630 times more likely to occur if it originated from two unknown, unrelated individuals than from [redacted] and one unknown, unrelated individual.

E04a: *Glove 3, exterior*. The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison. Combination possible. [redacted] [redacted] and [redacted] visually excluded from suitable contributors.

Contributor 1: 97% contribution, 6519 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 980 sextillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 3%, 216 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 200 million times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

E05a: *Glove 4, exterior.* The NOC for this sample was three. Only contributor 1 is suitable for comparison. [redacted] and Paul visually excluded from the suitable contributor.

Contributor 1: 94% contribution, 2447 RFU. LR for [redacted] Assuming a mixture of three, this mixture DNA profile obtained is at least 720 sextillion times more likely to occur if it originated from [redacted] and two unknown, unrelated individuals than if it originated from three unknown, unrelated individuals.

Contributor 2: 4% contribution, 109 RFU. No suitable contributor could be determined.

Contributor 3: 2% contribution, 56 RFU. No suitable contributor could be determined.

E06a: *Glove 5, exterior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison. Combination possible. [redacted] and visually excluded from the suitable contributors.

Contributor 1: 95% contribution, 6066 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 870 sextillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 5% contribution, 313 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 21 thousand times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

E06b: *Glove 5, interior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison. Combination possible. [redacted] and [redacted] visually excluded from the suitable contributors.

Contributor 1: 58% contribution, 784 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 49 billion times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 42% contribution, 563 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 19 quadrillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

E07a: *Glove 6, exterior.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison. None of the reference profiles could be visually excluded from the suitable contributors. Inclusionary LR's were generated for all reference profiles but Bode has noted in their summary results that although every person of interest could be present as a minor contributor, only [redacted] and Paul could be present simultaneously.[23]

Contributor 1: 62% contribution, 256 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 48 billion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

[23] See note, Appendix c. Table 1.

Contributor 2: 38% contribution, 159 RFU. LR for Paul: Assuming a mixture of two, this mixture DNA profile obtained is at least 1.9 thousand times more likely to occur if it originated from Paul and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. LR for [redacted] ([redacted] was put in contributor number 2 spot). Assuming a mixture of two, this mixture DNA profile obtained is at least 9 thousand times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. LR for [redacted] ([redacted] was put in contributor number 2 spot). Assuming a mixture of two, this mixture DNA profile obtained is at least 85 times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals. LR for [redacted] ([redacted] was put in contributor number 2 spot). Assuming a mixture of two, this mixture DNA profile obtained is at least 2.6 thousand times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

E09a: *Swab, reddish stain, south side kitchen island*. The NOC for this sample was one. Contributor 1 is suitable for comparison. [redacted] and Paul were visually excluded from the suitable contributors.

Contributor 1: 100% contribution, 1788 RFU. LR for [redacted] Assuming one contributor, this DNA profile obtained is at least 34 quintillion times more likely to occur if it originated from [redacted] than if it originated from an unknown, unrelated individual.

E10a: *Swab, reddish stain, lower kitchen cabinet*. The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison. Combination possible. [redacted] and Paul were visually excluded from the suitable contributors.

Contributor 1: 73% contribution, 1300 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 71 septillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 27% contribution, 4913 RFU LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 45 septillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

E11a: *Swab, reddish stain, pullout drawer below sink.* The NOC for this sample was two. Only contributor 1 is suitable for comparison. [redacted] and Paul were visually excluded from the suitable contributor.

Contributor 1: 99% contribution, 13,797 RFU. LR for [redacted] Assuming a mixture of two, this mixture DNA profile obtained is at least 990 sextillion times more likely to occur if it originated from [redacted] and one unknown, unrelated individual than if it originated from two unknown, unrelated individuals.

Contributor 2: 1% contribution, 91 RFU. No suitable contributor was identified.

b. Samples Tested by the New Jersey State Police Laboratory

The New Jersey State Police Laboratory analyzed two samples using STRmix v. 2.8.0. These samples were discussed by Christine Schlenker. (9T 82-17 to 97-14)

6-1-4-1: *Jeans, staining front right shin*[24]*.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison.

[24] These are the same jeans that were submitted to Bode and reflected in Bode samples E01a, E01b and E01c.

Contributor 1: 93.82% contribution, 1569 RFU. When examining the deconvoluted profile of contributor one, Ms. Schlenker observed that there were results at 7 or more locations. She then conducted a visual comparison of contributor one's profile to the reference profile of [redacted] could not be visually excluded. As such, Ms. Schlenker calculated an RMP. The RMP was one in 2.73 septillion, exceeding the source attribution threshold of one in eight trillion, which led to [redacted] being identified as the source of the profile for contributor one.

Contributor 2: 6.18% contribution, 103 RFU. Ms. Schlenker testified that when she ran this sample through STRmix, she ran a simultaneous comparison to the defendant's profile. The STRmix output report reflected that the defendant was placed into the contributor two spot, but nonetheless a likelihood ratio strongly supporting exclusion was generated. Comparisons of the remaining reference profiles ([redacted] and Sean Edson) were run through STRmix. STRmix placed each of the above reference profiles into the contributor two spot, but ultimately generated likelihood ratios supporting exclusion for [redacted] and Sean, and an uninformative likelihood ratio for [redacted]

41-1-1: *Inside collar scraping of long-sleeved shirt.* The NOC for this sample was two. Contributors 1 and 2 are suitable for comparison.

Contributor 1: 79.62% contribution, 343 RFU. Ms. Schlenker testified that when she input sample 41-1-1 to STRmix, she also ran a simultaneous comparison to the defendant's profile. The STRmix output report reflected that the defendant was placed into the contributor one spot and that a likelihood ratio of 110 million supporting inclusion was calculated.

Contributor 2: 20.38% contribution, 88 RFU. Comparisons of the remaining reference profiles ([redacted] and Sean) were run through STRmix, and that each was placed

into the contributor two spot. LRs supporting exclusion were generated for [redacted] [redacted] and Sean; and an uninformative LR was generated for [redacted] (9T 95-13 to 97-14).

c. Charts and Diagrams Referenced

Bode Case #: CCA2076-0119
Agency Case #: MCP1800780

Date: November 20, 2020

Table 1: STRmix Results

Bode Item #	Assumed # of Contributors	Suitable Contributors	1 ([redacted]	2 ([redacted]	3 ([redacted]	4 ([redacted]	20 (Paul Caneiro)	Combination Possible?
CCA2076-0119-E01a1	3	1, 2	Visually excluded.	Visually excluded.	470 sextillion (very strong support for inclusion)	Visually excluded.	3.7 quintillion (very strong support for inclusion)	Yes
CCA2076-0119-E01c1	1	all	Visually excluded.	Visually excluded.	2.1 septillion (very strong support for inclusion)	Visually excluded.	Visually excluded.	N/A
CCA2076-0119-E02b1	2	all	Visually excluded.	Visually excluded.	1 septillion (very strong support for inclusion)	Visually excluded.	43 thousand (strong support for inclusion)	Yes
CCA2076-0119-E03a1	2	all	Visually excluded.	Visually excluded.	100 quadrillion (very strong support for inclusion)	Visually excluded.	45 quadrillion (very strong support for inclusion)	Yes
CCA2076-0119-E03b1	2	all	480 (moderate support for exclusion)	37 thousand (strong support for exclusion)	750 quintillion (very strong support for inclusion)	630 (moderate support for exclusion)	400 thousand (strong support for inclusion)	Yes
CCA2076-0119-E04a1	2	all	Visually excluded.	Visually excluded.	980 sextillion (very strong support for inclusion)	Visually excluded.	200 million (very strong support for inclusion)	Yes
CCA2076-0119-E05a1	3	1	18	Visually excluded.	720 sextillion (very strong support for inclusion)	60	Visually excluded.	N/A
CCA2076-0119-E06a1	2	all	Visually excluded.	Visually excluded.	870 sextillion (very strong support for inclusion)	Visually excluded.	21 thousand (strong support for inclusion)	Yes
CCA2076-0119-E06b1	2	all	Visually excluded.	Visually excluded.	19 quadrillion (very strong support for inclusion)	9,730	49 billion (very strong support for inclusion)	Yes
CCA2076-0119-E07a1	2	all	2.6 thousand (moderate support for inclusion)	85 (limited support for inclusion)	48 billion (very strong support for inclusion)	9 thousand (moderate support for inclusion)	1.9 thousand (moderate support for inclusion)	* See note
CCA2076-0119-E09a1	1	all	Visually excluded.	34 quintillion (very strong support for inclusion)	Visually excluded.	Visually excluded.	Visually excluded.	N/A
CCA2076-0119-E10a1	2	all	Visually excluded.	71 septillion (very strong support for inclusion)	45 septillion (very strong support for inclusion)	Visually excluded.	Visually excluded.	Yes
CCA2076-0119-E11a1	2	1	Visually excluded.	Visually excluded.	990 sextillion (very strong support for inclusion)	Visually excluded.	Visually excluded.	N/A

* Note: Under the stated assumption that the mixture is of two individuals, the only two references submitted that can together be contributors to this mixture DNA profile are [redacted] and Paul Caniero.

Figure 7a: STRmix likelihood ratios with and without conditioning on a known contributor

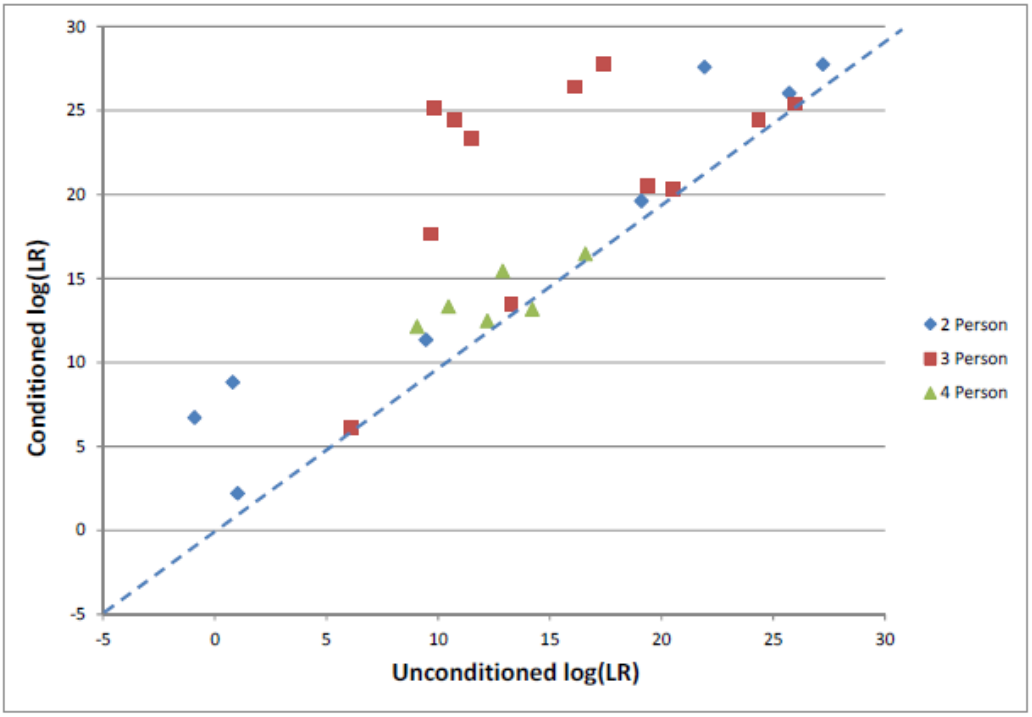

Figure 7b: STRmix likelihood ratios without conditioning – knowns in H_1 only

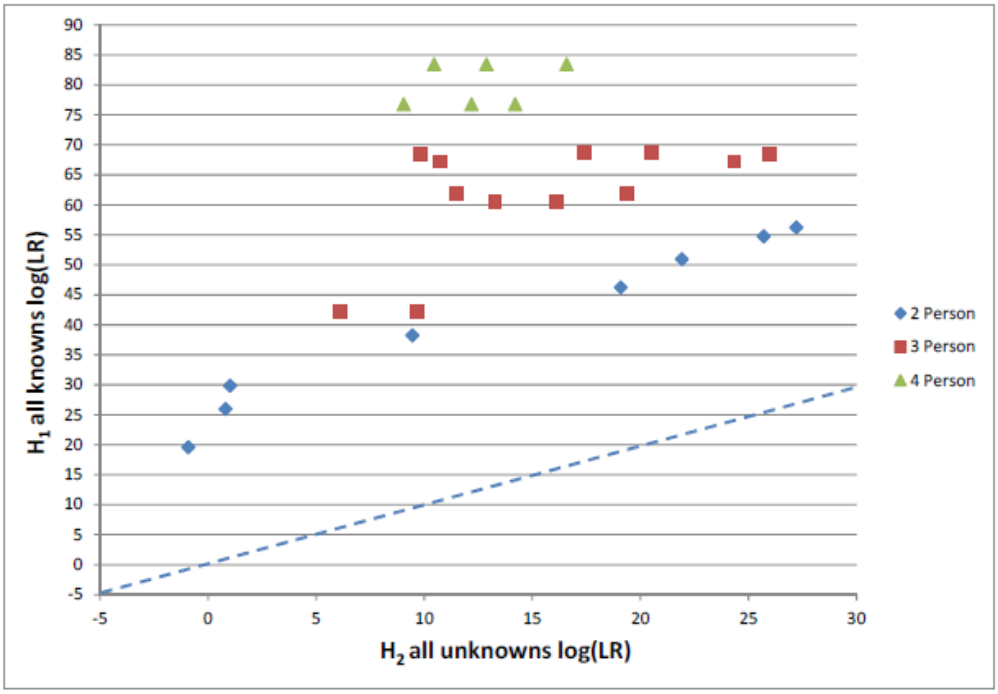

Figure 8: Comparison of log(LR) for true and non-contributors under the assumption of N and N+1

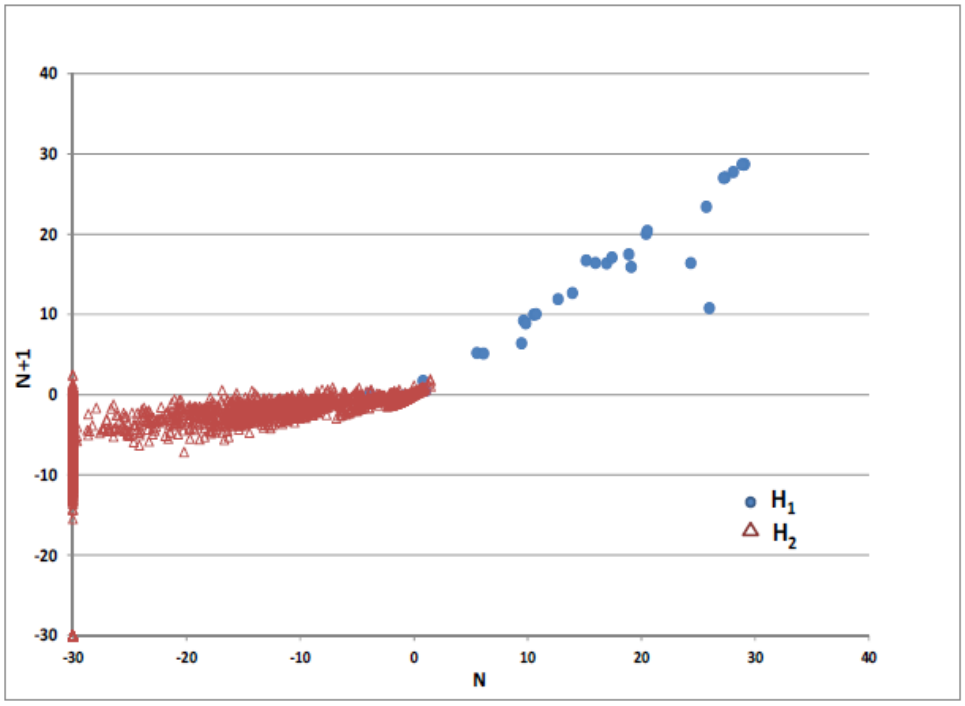

Figure 9: Comparison of log(LR) for true and non-contributors under the assumption of N and N-1

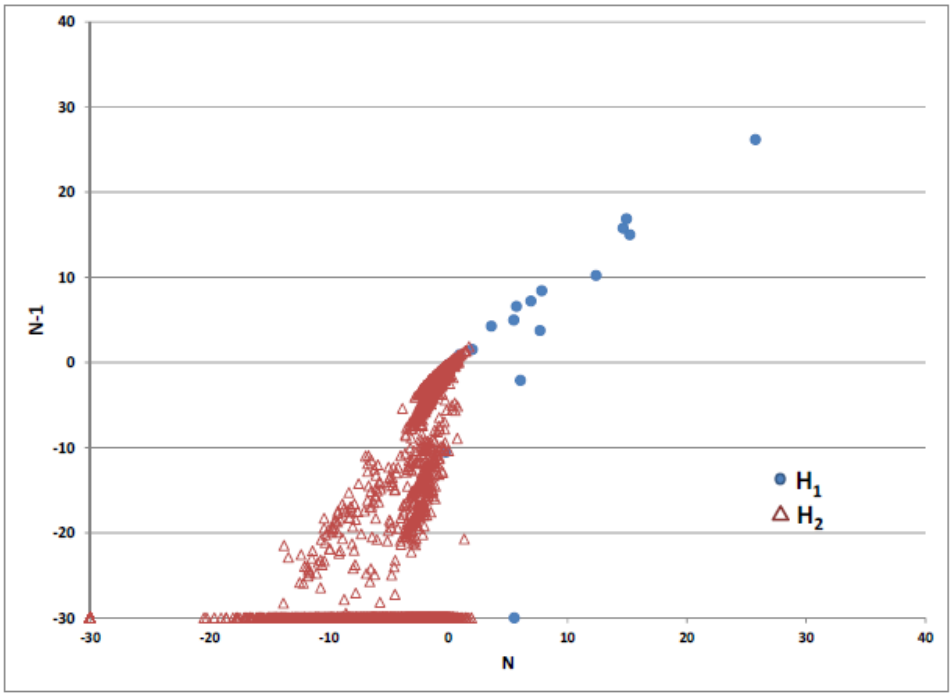

Table 3: Summary of mixtures prepared by the NJSP Laboratory to examine STRmix™ sensitivity and specificity.

Four-person mixtures	Three-person mixtures	Two-person mixtures	DNA amount of smallest contributor (pg)
-	-	2:1[1]	333
-	-	3:1[1]	250
-	-	5:1[1]	167
-	-	10:1[1]	91
-	2:2:1 and 3:1:1	-	200
-	4:2:1	-	142
4:3:2:1 and 10:5:2:1	10:5:1, 5:4:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	100
-	10:1:1	-	83
-	10:6:1	-	59
4:3:2:1 and 10:5:2:1	10:5:1 and 3:2:1	20:1, 10:1, 5:1, and 3:1	50
-	20:5:1	-	38
-	20:10:1	-	32
4:3:2:1, 10:5:2:1, and 100:100:100:6	10:5:1, 3:2:1, and 100:100:4	20:1, 10:1, 5:1, 3:1, and 100:2	25
4:3:2:1, 10:5:2:1, and 100:100:100:6	10:5:1, 3:2:1, and 100:100:4	20:1, 10:1, 5:1, 3:1, and 100:2	12.5
4:3:2:1, 10:5:2:1, and 100:100:100:6	10:5:1, 3:2:1, and 100:100:4	20:1, 10:1, 5:1, 3:1, and 100:2	6.25
Four-person mixtures	**Three-person mixtures**	**Two-person mixtures**	**DNA amount per contributor (pg)**
-	-	1:1[1]	500
1:1:1:1	1:1:1	1:1	400
1:1:1:1	-	-	250
1:1:1:1	1:1:1	1:1	200, 100, 50, 25, 12.5

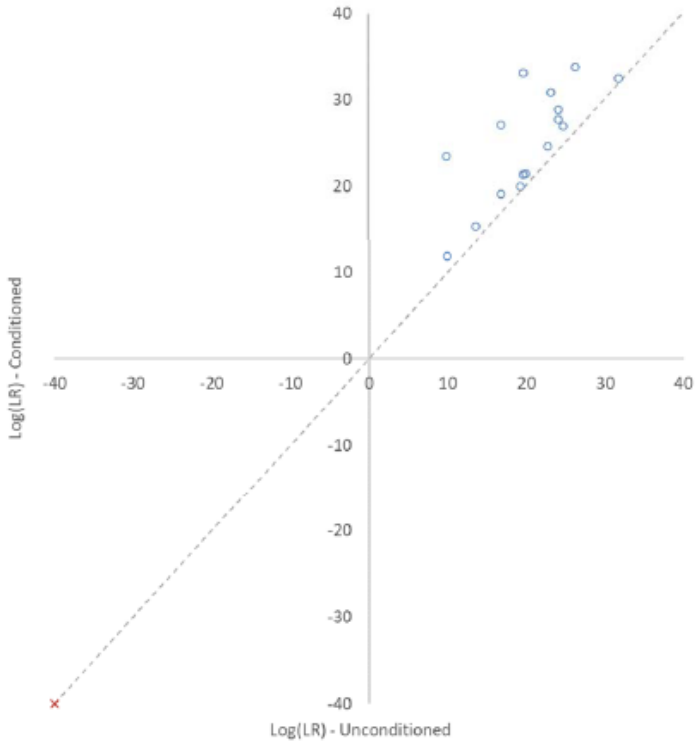

Figure 7: Comparison of sub-source log(*LR*) assigned following interpretation with and without use of a conditioning profile. Log(*LR*)s assigned for known donors are plotted using circles whilst those assigned for non-contributors are plotted using crosses. Exclusions (*LR* = 0) have been plotted as log(*LR*) = -40. A dashed line at $x = y$ has been added to assist with interpretation.