CERTIFIED FOR PUBLICATION

IN THE COURT OF APPEAL OF THE STATE OF CALIFORNIA

FIFTH APPELLATE DISTRICT

THE PEOPLE,	F030754
Plaintiff and Respondent,	(Super. Ct. No. 8517)
v.	
MICHAEL ANTONIO PIZARRO,	Madera County
Defendant and Appellant.	

APPEAL from a judgment of the Superior Court of Madera County. Edward P. Moffat, Judge.

Lynne S. Coffin, State Public Defender, Jeffrey J. Gale and Valerie Hriciga, Deputy State Public Defenders, for Defendant and Appellant.

Bill Lockyer, Attorney General, David Druliner and Robert R. Anderson, Chief Assistant Attorneys General, Robert R. Anderson and Jo Graves, Assistant Attorneys General, Stephen G. Herndon and Paul E. O'Connor, Deputy Attorneys General, for Plaintiff and Respondent.

-ooOoo-

In 1990, defendant and appellant Michael A. Pizarro was convicted of murder, forcible lewd or lascivious act on a child under age 14, and forcible rape. The case, now on appeal for the second time, presents an unusual procedural posture. In the first appeal, Pizarro contended the DNA (deoxyribonucleic acid) evidence against him was inadmissible because the prosecution had failed to demonstrate that the DNA restriction fragment length polymorphism (RFLP) testing conducted by the FBI was generally accepted in the scientific community. At that time, the admissibility of DNA evidence was still being debated, evaluated, and resolved by appellate review. We remanded the case for a thorough evidentiary (*Kelly*[1]) hearing. (*People v. Pizarro* (1992) 10 Cal.App.4th 57 (*Pizarro I*).) That *Kelly* hearing is the basis of this opinion. In 1998, the trial court again ruled that the DNA was admissible.

During the years since trial, significant case law has developed. In 1991, *People v. Axell* (1991) 235 Cal.App.3d 836 was decided, followed by *People v. Barney* (1992) 8 Cal.App.4th 798 in 1992. And, shortly after the trial court's *Kelly* ruling in this case in 1998, the Supreme Court published *People v. Venegas* (1998) 18 Cal.4th 47.

It is in this procedural context that defendant appeals again, contending the DNA evidence was inadmissible for various reasons.[2] We now address several issues, most of which were not originally recognized, either because of the state of the record in the first proceeding or because they now arise as a result of the present proceeding and developments in DNA analysis post-1990. We will reverse the judgment.

Due to the heavy burden placed on judges and attorneys who grapple with sophisticated, technical, and often subtle scientific issues, we publish a detailed opinion that we hope will provide some guidance. We recognize this opinion is exhaustive in

---

[1] *People v. Kelly* (1976) 17 Cal.3d 24.

[2] For consistency and clarity, throughout this opinion we generally refer to Pizarro as "defendant" or "Pizarro," rather than "appellant."

both length and detail; however, the exceptionally compelling nature of DNA evidence requires us to demand a high degree of accuracy and accountability in its use.

In this case, we hold the following:

(1) The frequency of the perpetrator's genetic profile (the random match probability) calculated from the Hispanic database was admitted without adequate foundation because there was insufficient evidence that the perpetrator is Hispanic.[3] *To make the ethnic database relevant, the prosecution was required to present sufficient foundational evidence to show that the perpetrator is within that database's ethnicity.* (Evid. Code, § 403.) In the absence of sufficient proof that the perpetrator is Hispanic, the Hispanic database was irrelevant, and the Hispanic profile frequency was irrelevant and created substantial danger of confusing the issues and misleading the trier of fact. The trial court abused its discretion both in finding sufficient evidence of the perpetrator's Hispanic ethnicity and in not ruling that use of the Hispanic database was improper scientific procedure.

This error was compounded when the prosecution and the FBI improperly relied on *defendant's* ethnicity to justify use of the ethnic database. First, since there was inadequate foundational proof of the *perpetrator's* ethnicity, *defendant's* ethnicity was irrelevant and reference to it as an incriminating trait was error. Second, and of greater consequence, the jury was directly informed that the FBI used the Hispanic database *because defendant is Hispanic* -- and thus the jury was indirectly informed that *defendant's* ethnicity served as proof of the *perpetrator's* ethnicity and was relied upon to render the ethnic database relevant. In other words, this bootstrap logic allowed

---

[3]   We use the terms "Hispanic" and "Caucasian" for consistency because they were used in this case and because they are used by scientific sources we cite. The term "perpetrator" is used specifically to designate the person who *committed* the criminal act, as opposed to "defendant," which designates the person who is *accused* of committing the criminal act.

defendant's ethnicity to justify calculation of an ethnic frequency, which when presented to the jury effectively operated as proof of the *perpetrator's* ethnicity -- which in turn served as evidence of *defendant's* guilt. Reliance on defendant's ethnicity was founded on the improper assumption that *defendant is in fact the perpetrator*, and that assumption was conveyed by implication to the jury.

Under the facts of this case, the individual errors connected to the improper use of the ethnic database may or may not have created sufficient prejudice to compel reversal, but the combination of these errors with the other errors committed in this case does, in our view, constitute prejudice requiring reversal.

(2) The perpetrator's genotype at one of the genetic loci was also admitted without adequate foundation because there was insufficient evidence of the perpetrator's genotype at that locus. *The relevance of data from that genetic locus, including defendant's genotype and the conclusion that defendant matches the perpetrator at that locus, required that the prosecution present sufficient foundational proof of the perpetrator's genotype at that locus.* (Evid. Code, § 403.) Without such proof, data from that locus were irrelevant and inadmissible. The trial court abused its discretion both in finding sufficient evidence of the perpetrator's genotype and in failing to find use of the data from that locus improper scientific procedure.

Proof of the perpetrator's genotype at that locus was insufficient because the evidence demonstrated that the perpetrator's genotype was *not discernable* from a mixed perpetrator/victim DNA sample except by two methods, neither of which was permissible. First, reference to *defendant's* genotype was not permissible to establish the perpetrator's genotype. Just as defendant's ethnicity was irrelevant to the determination of the perpetrator's ethnicity, defendant's genotype was irrelevant to the determination of the perpetrator's genotype. Without sufficient foundational proof of the *perpetrator's* genotype, reference to *defendant's* genotype as an incriminating trait was error, and reliance on defendant's genotype was based on the improper assumption that *defendant is*

in fact the perpetrator. Second, use of band-intensity analysis to discern the perpetrator's genotype from the autoradiograph (autorad) was not permissible because that method is subject to *Kelly* scrutiny and has not yet undergone such scrutiny. Had the data from that locus been properly excluded, the frequency of the overall genetic profile would have been more common and less compelling evidence as to the guilt of defendant.

(3) The evidence established that *the FBI's match window was a ± 5% window, and that the FBI's statistical window was therefore required to be at least ± 5% in order to span all the alleles in the population (and the fixed bins into which they are grouped) that could be the same as the perpetrator's allele.*[4] In this procedure, the original description of matching alleles must remain the same (or be broadened) when the number of those matching alleles is estimated. Although the fixed bin method does not directly count those alleles, it attempts to conservatively approximate a direct count by referring to predefined, precounted groups of alleles. Fewer bins may be overlapped by an undersized statistical window and the resulting frequency, which is chosen from the highest of the overlapped bins, may be significantly underestimated. The trial court abused its discretion by not ruling that the FBI's use of an undersized statistical window was improper scientific procedure.

(4) The evidence established that the FBI's statistical window may have been centered on the average of the perpetrator's and defendant's alleles or drawn around the outline of the perpetrator's and defendant's overlapping uncertainty windows. *Either*

---

4   "Statistical windows" are discussed in part III.E.2.c., *post.* An allele, defined in more detail *post*, is a particular segment of DNA.

We recognize that "percent" is considered the more appropriate form in legal writing, but in this opinion we choose to use the scientific form of "%" due to the scientific nature of this material. Similarly, we use "±" rather than "plus or minus." We believe this usage not only suits the subject matter, but also enhances the consistency, effectiveness, and readability of this opinion.

method improperly took into account defendant's allele measurement¾again impermissibly presupposing a degree of identity between the perpetrator and defendant. The statistical window should have been centered on the *perpetrator's* allele measurement because the window's purpose is to estimate the frequency of the *perpetrator's* allele in the population. Defendant was irrelevant to this determination and use of a statistical window affected by defendant's allele constituted improper scientific procedure. The trial court abused its discretion by failing to so rule.

(5) The propriety of the H2 Hispanic database is moot.

(6) The trial court did not abuse its discretion by not ruling that the failure to present the possibility of laboratory error with the profile frequency was improper scientific procedure.

(7) The trial court did not abuse its discretion by not ruling that the failure to present a confidence interval with the profile frequency was improper scientific procedure.

STATEMENT OF THE CASE

The following statement of the case is taken from our opinion in *Pizarro I*:[5]

> "On August 11, 1989, an information was filed alleging [defendant] Michael A. Pizarro had committed the following crimes: count I, murder of [the victim] (Pen. Code, § 187) with the special circumstances that the murder was committed while [defendant] was engaged in the crime of rape (Pen. Code, § 190.2, subd. (a)(17)), and that the murder was committed while [defendant] was engaged in the crime of a lewd or lascivious act upon a child under age 14 (Pen. Code, § 190.2, subd. (a)(17)); count II, forcible lewd or lascivious act on a child under age 14 (Pen. Code, § 288, subd. (b)); and count III, forcible rape (Pen. Code, § 261, subd. (a)(2)).

---

[5]   Footnotes are included and sequentially renumbered.

For consistency and clarity, throughout this opinion we generally refer to Pizarro as "defendant" or "Pizarro," rather than "appellant."

"On August 17, 1989, [defendant] was arraigned and pleaded not guilty.

"On May 22, 1990, jury selection commenced. On May 31, 1990, during trial, a *Kelly/Frye*[6] hearing was held to determine the admissibility of the results of DNA identification evidence and the trial court ruled the results were admissible.

"On June 6, 1990, the jury returned verdicts finding [defendant] guilty of all counts and also finding the charged special circumstances to be true.

"On July 3, 1990, [defendant] was sentenced to life in prison without the possibility of parole on count I, to be served consecutively to the upper term of eight years on count II. The sentence on the rape count was stayed pursuant to Penal Code section 654.

"On July 6, 1990, [defendant] filed his notice of appeal." (*Pizarro I, supra,* 10 Cal.App.4th at pp. 60-61.)

On appeal, we remanded to the trial court for a full-blown evidentiary hearing to determine the general scientific acceptance of the FBI's DNA profiling procedure and the FBI's Hispanic database. (*Pizarro I, supra,* 10 Cal.App.4th at pp. 95-96.) On March 19, 1998, after a hearing conducted in 1994 and 1995, the trial court found the procedure and the database generally accepted and the evidence admissible. Defendant filed a timely notice of appeal.

STATEMENT OF FACTS

The following statement of facts is also taken from *Pizarro I*:[7]

"On June 10, 1989, [defendant], along with his wife, Sandy, and his five-month-old son, drove from Clovis to North Fork, California, to visit his family. They arrived around noon and, soon thereafter, [defendant] went to a schoolyard to play basketball with a friend. Following the basketball game, [defendant] visited the home of his friend and also spent

---

6   "*People v. Kelly* [*supra,*] 17 Cal.3d 24 and *Frye v. United States* (D.C. Cir. 1923) 293 F. 1013" (hereafter *Kelly/Frye*).

7   Footnotes are included and sequentially renumbered.

time at Manzanita Lake. [Defendant] then returned to his mother's house and, later that evening (about 8 p.m.), he and his wife went to a party at a mobilehome park in town. [Defendant]'s 13-year-old half sister, [the victim], was also at the party.

"[Defendant] had consumed beer throughout the afternoon and he continued to drink at the party. Because Sandy wanted to leave before [defendant] was ready to go, she and [defendant] argued and Sandy left without him -- then returned to try to persuade [defendant] to join her. Eventually, [defendant] began walking toward his mother's house. Sandy followed in their truck and repeatedly asked [defendant] to get inside with her. [Defendant] ignored the requests and behaved erratically, crisscrossing the road, lying in front of the truck and, occasionally, hiding from Sandy. After approximately a half hour, Sandy left [defendant] in the road and drove to the home of her mother-in-law, Chris Conston.

"Sandy arrived at the Conston house about 1 a.m. [The victim], who had returned from the party earlier, agreed to accompany Sandy back to the area where she had left [defendant]. [The victim]'s mother gave her a flashlight before she left with Sandy and the Pizarros' baby in their truck.

"Thereafter, Sandy and [the victim] saw [defendant] walking towards town but when they approached him, [defendant] ran. When Sandy turned around to follow, [defendant] ran up an embankment and Sandy shined the flashlight on him. [Defendant] then came down from the embankment and, again, began running for town. Sandy stopped the truck and [the victim], who had been holding the baby, put the child down on the seat and got out, taking the flashlight with her. Sandy watched [the victim] cross the street towards the area where [defendant] had gone. Sandy picked up her baby and closed the passenger door. When she looked up, [the victim] was gone.

"Sandy called out for [defendant] and [the victim] but there was no response. She circled her truck around and yelled for them to turn on the flashlight or say something to let her know they were all right. She then saw a flash of light coming from the area where she had last seen [the victim]. She then heard a scream and, immediately following the scream, a slight muffled sound. Frightened, she returned to the Conston house and told her mother-in-law what had happened. It was then almost 2:30 a.m.

"Chris Conston called 911 and Sandy arranged to meet sheriff's deputies at Sierra Automotive which she believed was near the area where [defendant] and [the victim] had last been seen. At 2:51 a.m., within 20 minutes after the 911 call, Madera County Sheriff's Deputy Weisert met Sandy and was directed to the place where Sandy thought [defendant] and

[the victim] had gone.[8] Another deputy and Chris Conston also went to the area and they drove up and down the road calling for [the victim] over a public address system. There was no response and, soon after 4 a.m., the officers left the area. After waiting for Sandy's parents to come for Sandy, Chris Conston also went home.

"About 5:50 a.m., [defendant] showed up alone at his mother's house. He was dirty, sleepy and appeared to his mother to be drunk. [Defendant] told his mother that, on his way home, a man had confronted him and accused him of kidnapping his sister.[9] Mrs. Conston then left to search for [the victim] at a friend's house and [defendant] went to sleep.

"Shortly after 7 a.m., officers again began searching the area which Sandy Pizarro had pointed out. When they were unable to find [the victim], Deputy Lidfors went to the Conston home at about 8 a.m. to talk to [defendant]. [Defendant] was awakened and he told the officer to look at another location approximately one-tenth of a mile farther west from the area where they had been searching. During this conversation, [defendant] did not appear intoxicated or 'hung over' to the officer.

"Deputy Lidfors, along with Deputy Nelson, went to the area described by [defendant] and there they found [the victim]'s body. [The victim]'s pants had been removed and her underpants were down around her right foot; her T-shirt and bra were pushed up above her breasts. Deputy Lidfors noticed bruises on [the victim]'s face and blood smears on her stomach and leg. Her flashlight was lying by her feet.

"An autopsy was performed and the pathologist, Dr. Gerald Dalgleish, determined that suffocation was the cause of death. He also noted the presence of bruises on the right side of the victim's face as well as swelling and discoloration around her lips and a mark on her nose. [The victim] had been alive when the injuries to her face were inflicted and the pathologist believed that the flashlight could have been the instrument

---

8   "The following day Sandy realized she had made an error and had actually last seen [the victim] and [defendant] a short distance up the road (under one-tenth of a mile away). It was in that area that [the victim]'s body was found."

9   "[Defendant] later told Madera County Sheriff's Detective Kern that the sheriff had stopped him and made the accusation. He told Deputy Weisert that 'some cops' had met him and accused him of kidnapping."

which caused some of the injuries. Semen was present in [the victim]'s vagina.

"On the morning [the victim]'s body was found, [defendant] was taken to the sheriff's substation and interviewed by Sergeant Gauthier. [Defendant] told Gauthier that, after [the victim] had followed him into the brush, he told her he was mad at his wife and did not want to return to the truck. He said he then started to walk up the hill but [the victim] was mad because he had taken her flashlight. He said he was several paces away from her so he turned to toss the flashlight back to her and then left. According to [defendant], that was the last time he had seen [the victim]. At the time of the interview, Sergeant Gauthier examined [defendant]'s hands and found that the knuckles on one of [defendant]'s hands were red and swollen. Gauthier collected the clothes [defendant] was wearing and arranged to have samples of [defendant]'s blood drawn.

"[Defendant] was also interviewed 10 days later by Madera County District Attorney investigator Fred Flores. [Defendant] told Flores that, after he had thrown the flashlight back to [the victim], he continued running up the hill and passed out about 100 yards later. [Defendant] claimed he did not know what occurred from that point until the time he awoke and walked to his mother's house. When Flores asked [defendant] how he would feel about being arrested, [defendant] told Flores, 'it would be a big mistake because [Flores] did not have enough proof.' [Defendant] did not specifically deny having killed his sister in that conversation. He did deny that he had undressed.[10]

"Forensic tests determined that [the victim]'s blood type was O and she was a nonsecretor. [Defendant]'s blood is type B and he is a secretor. Approximately 8 percent of the population is comprised of type B secretors. The semen which was present in the victim's vagina was from a type B secretor. Additional vaginal swabs and reference blood samples from [defendant] and victim were sent to the Federal Bureau of Investigation's (FBI) laboratory in Washington D.C. for deoxyribonucleic acid (DNA) genetic analysis.

---

10   "Foxtails were found in the victim's hair, fist and hairband. Foxtails were also present inside and outside of [defendant]'s shorts and in his underwear."

"Dr. Dwight Adams, a special agent assigned to the FBI laboratory, performed DNA analysis on the evidence [in 1989].[11] Dr. Adams concluded the DNA from the semen on the vaginal swabs matched the known blood sample of [defendant]. Using a data base from a Hispanic population, Dr. Adams noted that the likelihood of finding another unrelated Hispanic individual with a similar profile would be approximately 1 in 250,000.[12]

"Defense

"[Defendant] testified at trial. He said that he had consumed beer throughout the afternoon and evening and, by the time he arrived at the party at the mobilehome park, he was fairly intoxicated. While there, he continued to drink beer and mixed drinks. He testified that he remembered his argument with Sandy and leaving the party with the intention of walking to his mother's house. He also recalled crisscrossing the road and lying down in front of the truck.

"When Sandy returned with [the victim], he attempted to hide and ran into the brush. He testified that [the victim] followed him but he told her that he and Sandy were having problems and that she should go home. According to [defendant], he took [the victim]'s flashlight and started walking away. He said that when [the victim] asked for the light, he turned and tossed it to her.

"Throughout his testimony, [defendant] maintained he remembered nothing from the time he threw the flashlight until he woke up in the brush. [Defendant] said that, when he awoke, he did not walk back to North Fork along the dirt road but instead cut through an area of brush and trees. [Defendant] claimed to have met a man in tan pants and a white shirt who he assumed was a law enforcement officer and who accused him of

---

11　"The qualifications of Dr. Adams and the methods used in conducting the analysis will be discussed, in detail, in the portion of this opinion addressing [defendant]'s contentions regarding DNA analysis admissibility."

12　"In the White population, the likelihood would decrease to 1 in 10,000,000. When a subject is half White and half Hispanic, the FBI would use the more conservative statistic applicable to the Hispanic population (here 1 in 250,000) to favor a defendant."

kidnapping his sister. He also said that he saw a full-size pickup on the road when it was fairly light out.[13]

"[Defendant] testified that the injury to his hand had occurred at work. [Defendant] denied telling investigating officers that he had not removed his underwear or clothes, and claimed that he had actually told them he did not 'believe' he had undressed. He also said investigator Flores had mischaracterized his response to the question of how he would feel about being arrested. Rather than stating to Flores that it would be a mistake because there 'wasn't enough proof,' [defendant] testified that he told Flores that Flores would be making a mistake 'because [he] didn't kill [the victim].'

"[Defendant] also testified that he had, in the past, suffered blackouts and loss of consciousness after drinking excessively and that such episodes began to occur more frequently after he suffered a head injury in 1985. He also admitted that he told an investigator that alcohol made him violent.

"[Defendant]'s mother also testified for the defense. She said [defendant] and [the victim] had been close. Although [defendant] had scratches on him when he appeared at her home in the morning, the scratches did not appear to her to have been made by a person; she assumed he had been scratched by bushes. Mrs. Conston recalled that, when [defendant] learned his sister was dead, he put his head in her lap and cried.

"Guy Clements was the final defense witness at trial. Mr. Clements was working as a newspaper delivery person on June 11, 1989. He testified that he was driving near the area where [the victim]'s body was found, about 1:30 a.m., when he saw a red Datsun pickup stopped in the middle of the road. It appeared to him that there was a man inside the truck.[14]" (*Pizarro I, supra,* 10 Cal.App.4th at pp. 61-66.)

---

13   "[Defendant] said that Detective Gauthier was mistaken in reporting that [defendant] had previously stated that he saw the truck after he had run into the brush away from [the victim]."

14   "On the morning that the crime was discovered, a year before trial, Mr. Clements reported that he had seen a yellow gold, newer model Nissan truck with a young White male inside."

12.

DISCUSSION

Defendant does not argue the scientific methods used for DNA analysis in his case were not generally accepted by the scientific community, but instead contends proper scientific methods were not followed in this particular case. He raises his contentions under the third prong of *Kelly*, in which the Supreme Court articulated this three-step test for the admission of evidence generated by a new scientific technique: (1) the reliability of the technique must be sufficiently established to have gained general acceptance in the relevant scientific community; (2) the witness providing the evidence must be properly qualified as an expert; and (3) the evidence must establish that, in the particular case, the correct and accepted scientific technique was actually followed. (*People v. Kelly, supra,* 17 Cal.3d at p. 30.)[15]

Specifically, defendant argues proper scientific procedures were not followed in this case because (1) evidence of a Hispanic profile frequency was improperly admitted without sufficient evidence that the perpetrator is Hispanic; (2) all possible genotypes in a mixed sample were improperly unaccounted for; (3) the statistical window was too small; (4) the statistical window was improperly centered on the average of the perpetrator's and defendant's allele measurements; (5) the H2 Hispanic database was defective; (6) evidence of the possibility of laboratory error should have been presented in addition to the profile frequency; and (7) evidence of a confidence interval should have been presented in addition to the profile frequency. Defendant also argues that, in the event we find the evidence admissible, he should receive a new trial so his evidentiary

---

[15]   Although the federal *Frye* analysis has been superceded by the Federal Rules of Evidence (28 U.S.C.), as held in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) 509 U.S. 579, the California Supreme Court reaffirmed the *Kelly-Frye* test in this state (*People v. Leahy* (1994) 8 Cal.4th 587, 611). The foundational requirement is now referred to as the *Kelly* test. (*People v. Leahy, supra,* at p. 612; *People v. Soto* (1999) 21 Cal.4th 512, 515, fn. 3.)

13.

challenges can be heard by the jury that determines his guilt, and, lastly, that his counsel was ineffective for failing to properly contest the DNA evidence.

We first address our scope of review following our remand to the trial court for a *Kelly* hearing, then the law and science relevant to DNA evidence, and finally each of defendant's specific contentions.

I.   SCOPE OF REVIEW

In an appeal following a limited remand, the scope of the issues before the reviewing court is determined by the remand order. (*People v. Deere* (1991) 53 Cal.3d 705, 713; *People v. Murphy* (2001) 88 Cal.App.4th 392, 396.) In *Pizarro I*, we concluded:

> "[W]e are of the opinion that remand is appropriate to allow the trial court the opportunity to conduct a 'full blown' *Kelly/Frye* hearing. At the conclusion of the hearing, the trial court must decide whether there is general acceptance in the scientific community of the DNA testing method and the data base utilized in the instant case by the FBI." (*Pizarro I, supra,* 10 Cal.App.4th at p. 95.)

Having remanded for the sole purpose of a complete *Kelly* hearing to determine the admissibility of the DNA evidence, we limit our review to issues raised within *Kelly*'s parameters.

II.   LAW

In *Kelly*, the Supreme Court spoke to the dangers of scientific evidence and its power to mystify and impress a jury. The court formulated a test composed of three prongs, the first and third of which specifically address the scientific procedures used to generate the scientific evidence against the defendant.[16] The first prong requires that the scientific procedures be *reliable*, as shown by their general acceptance by scientists in the

---

[16]   We consider the terms "procedure," "technique," and "methodology" interchangeable in this context.

14.

relevant field. The third prong requires that the reliable, generally accepted procedures were *actually followed* or *complied with* by the scientists in the particular case before the court. (*People v. Kelly, supra*, 17 Cal.3d at p. 30.) The party offering the evidence has the burden of proving its admissibility by a preponderance of the evidence. (*People v. Ashmus* (1991) 54 Cal.3d 932, 970.)

The *Kelly* test is an evidence-screening device that targets highly sophisticated scientific evidence that to the average juror would be not only incomprehensible but also irresistibly convincing. The test requires that such evidence pass the court's scrutiny before it is submitted to the jury -- it "is intended to forestall the jury's uncritical acceptance of scientific evidence or technology that is so foreign to everyday experience as to be unusually difficult for laypersons to evaluate. [Citation.] In most other instances, the jurors are permitted to rely on their own common sense and good judgment in evaluating the weight of the evidence presented to them. [Citations.] [¶] DNA evidence is different." (*People v. Venegas, supra,* 18 Cal.4th at p. 80.) "Lay jurors tend to give considerable weight to 'scientific' evidence when presented by 'experts' with impressive credentials." (*People v. Kelly, supra,* 17 Cal.3d at p. 31.) "'[S]cientific proof may in some instances assume a posture of mystic infallibility in the eyes of a jury ....' [Citation.]" (*Id.* at p. 32.) "Unlike fingerprint, shoe track, bite mark, or ballistic comparisons, which jurors essentially can see for themselves," questions concerning sophisticated scientific concepts, procedures, and laboratory compliance require educated expert testimony. (*People v. Venegas, supra,* 18 Cal.4th at p. 81.)

"'It is our duty … , where the life or liberty of a defendant is at stake, to be particularly careful that there is not only substantial evidence to support the implied finding of [defendant's] identity but that the finding is based upon admissible and nonprejudicial evidence.'" (*People v. Kelly, supra,* 17 Cal.3d. at p. 36.) Because of the immense power of scientific evidence, the *Kelly* test goes to the admissibility, not the weight, of the evidence. (*Id.* at pp. 30-32.)

## A.	KELLY'S FIRST PRONG

The trial judge serves as gatekeeper, allowing only evidence that is sufficiently reliable and trustworthy to reach the jurors. In performing this function in the context of scientific evidence, the judge must rely on the educated testimony of scientific experts. Thus, the first prong of the *Kelly* test -- the general acceptance of the procedure by the relevant scientific community -- is intended to confirm the reliability of a procedure too sophisticated or technical for the average lay person to readily understand. (See *People v. Kelly, supra,* 17 Cal.3d at pp. 30-32; *Frye v. United States*, *supra*, 293 F. 1013.) The first prong "assures that *those most qualified to assess the general validity of a scientific method will have the determinative voice*." (*People v. Kelly, supra,* 17 Cal.3d at p. 31.) It is "intended to interpose a substantial obstacle to the unrestrained admission of evidence based upon new scientific principles…. [A] '... misleading aura of certainty … often envelops a new scientific process, obscuring its currently experimental nature.' [Citations.] … [¶] Exercise of restraint is especially warranted when the identification technique is offered to identify the perpetrator of a crime. "'When identification is chiefly founded upon an opinion which is derived from utilization of an unproven process or technique, the court must be particularly careful to scrutinize the general acceptance of the technique.'" [Citation.]" (*Id.* at pp. 31-32.)

The question of general scientific acceptance may be answered by prior case law: "once a trial court has admitted evidence based upon a new scientific technique, and that decision is affirmed on appeal by a published appellate decision, the precedent so established may control subsequent trials, at least until new evidence is presented reflecting a change in the attitude of the scientific community." (*Kelly, supra,* 17 Cal.3d. at p. 32; *People v. Venegas, supra,* 18 Cal.4th at p. 54 ["trial court could properly rely on [a published appellate decision] as establishing general scientific acceptance"].) However, the published decision does not serve as precedent when there is proof of a "material scientific distinction" between the methodology approved by the published case

16.

and that used in the case before the court. (*People v. Venegas, supra,* 18 Cal.4th at p. 54.) Materially distinct procedures must pass first-prong scrutiny independently.

B.   KELLY'S THIRD PRONG

The third *Kelly* prong is a case-specific inquiry that asks: were the proper scientific procedures (those that have been deemed generally accepted under the first prong) *followed in this case*? (*People v. Venegas, supra,* 18 Cal.4th at p. 78.) Or, here, did the FBI scientists follow correct scientific procedures when they performed the DNA testing in Pizarro's case?

The *Venegas* court comprehensively explained *Kelly*'s third prong:

> "The *Kelly* test's third prong … assumes the methodology and technique in question has already met [the general acceptance] requirement. Instead, it inquires into the matter of whether *the procedures actually utilized in the case* were in compliance with that methodology and technique, as generally accepted by the scientific community. [Citation.] The third-prong inquiry is thus case specific; 'it cannot be satisfied by relying on a published appellate decision.' [Citation.]

> "... 'Due to the complexity of the DNA multisystem identification tests and the powerful impact that this evidence may have on a jury, satisfying *Frye* [i.e., satisfying *Kelly*'s first prong] alone is insufficient to place this type of evidence before a jury without a preliminary critical examination of the actual testing procedures performed....' [Citation.]" [¶] … [¶]

> "[Q]uestions concerning whether a laboratory has adopted correct, scientifically accepted procedures for [DNA testing] or determining a [profile] match depend almost entirely on the technical interpretations of experts. [Citations.] Consideration and affirmative resolution of those questions constitutes a prerequisite to admissibility under the third prong of *Kelly*.

> "The *Kelly* test's third prong does not, of course, cover all derelictions in following the prescribed scientific procedures. Shortcomings such as mislabeling, mixing the wrong ingredients, or failing to follow routine precautions against contamination may well be amenable to evaluation by jurors without the assistance of expert testimony. Such readily apparent missteps involve 'the degree of professionalism' with which otherwise scientifically accepted methodologies are applied in a

17.

given case, and so amount only to '[c]areless testing affect[ing] the weight of the evidence and not its admissibility' [citation].

"The *Kelly* third-prong inquiry involves further scrutiny of a methodology or technique that has already passed muster under the central first prong of the *Kelly* test, in that general acceptance of its validity by the relevant scientific community has been established. The issue of the inquiry is whether the procedures utilized in the case at hand complied with that technique. Proof of that compliance does not necessitate expert testimony anew from a member of the relevant scientific community directed at evaluating the technique's validity or acceptance in that community. It does, however, require that the testifying expert understand the technique and its underlying theory, and be thoroughly familiar with the procedures that were in fact used in the case at bar to implement the technique. [Citations.]" (*People v. Venegas, supra,* 18 Cal.4th at pp. 78-81.)

"The third-prong hearing 'will not approach the "complexity of a full-blown" *Kelly* hearing. [Citation.] "All that is necessary in the limited third-prong hearing is a foundational showing that correct scientific procedures were used." [Citation.]' [Citation.] Where the prosecution shows that the correct procedures were followed, criticisms of the techniques go to the weight of the evidence, not its admissibility. [Citations.]" (*People v. Brown* (2001) 91 Cal.App.4th 623, 647.) Similarly, where there is substantial evidence showing both that the procedures were followed and that they were not followed, the question is one for the jury to resolve. (*People v. Venegas, supra,* 18 Cal.4th at p. 91.) But where defense evidence establishes a failure in procedure, and that failure is not contradicted by substantial evidence, then the scientific evidence produced as a result of that incorrect procedure is inadmissible. (See *id.* at pp. 91-92.)

C. STANDARDS OF REVIEW

1. First Prong: De Novo

When the trial court relies on a published appellate decision finding general scientific acceptance of a scientific procedure, if there is no proof of any material scientific distinction between the accepted procedure and that used in the case before the court, the appellate court upholds its ruling. (*People v. Venegas, supra,* 18 Cal.4th at pp.

53-54.) But when the trial court independently concludes that a new scientific technique has been generally accepted, we independently review that conclusion. (*Id.* at p. 85.) "The preliminary showing of general acceptance of the new technique in the relevant scientific community is a mixed question of law and fact. [Citations.]" (*People v. Axell, supra,* 235 Cal.App.3d at p. 854.) "[I]n reviewing the scientific acceptance of [a methodology] de novo under *Kelly*, we are not required to decide whether such methodology is 'reliable as a matter of "scientific fact," but simply whether it is generally accepted as reliable by the relevant scientific community.' [Citation.] '"General acceptance" under *Kelly* means a consensus drawn from a typical cross-section of the relevant, qualified scientific community.' [Citation.] The *Kelly* test does not demand 'absolute unanimity of views in the scientific community .... Rather, the test is met if use of the technique is supported by a clear majority of the members of that community.' [Citation.]" (*People v. Venegas, supra,* 18 Cal.4th at p. 85.) Conversely, the test fails if "'"scientists significant either in number or expertise publicly oppose [a technique] as unreliable."' [Citations.]" (*People v. Axell, supra,* 235 Cal.App.3d at p. 854.) "In determining the question of general acceptance, courts 'must consider the quality, as well as quantity, of the evidence supporting or opposing a new scientific technique. Mere numerical majority support or opposition by persons minimally qualified to state an authoritative opinion is of little value ....' [Citation.]" (*People v. Venegas, supra,* 18 Cal.4th at p. 85.)

"Because the technical complexity of many new scientific procedures may prevent lay judges from determining the existence, degree, or nature of a scientific consensus without the testimony and interpretation of qualified experts in the field, *Kelly/Frye* properly emphasizes the record made at trial. [Citation.]" (*People v. Axell, supra,* 235 Cal.App.3d at p. 854.) In addition to reviewing the trial court record, the appellate court may also independently survey the scientific literature and case law to determine whether acceptance of the procedure does indeed exist. (*Ibid.*)

2.   Third Prong: Abuse of Discretion

In contrast to first-prong issues, the trial court's third-prong conclusions that proper procedures were followed in the particular case are reviewed for abuse of discretion. (*People v. Venegas, supra,* 18 Cal.4th at p. 91.) The appellate court is "required to accept the trial court's resolutions of credibility, choices of reasonable inferences, and factual determinations from conflicting substantial evidence. [Citation.]" (*Ibid.*) We thus consider whether there is substantial evidence in the record to support the conclusion that the procedures were in fact performed in a manner fully consistent with the underlying science such that they produced reliable results. (*Id.* at pp. 91-92.)

"'This standard is *deferential*. [Citations.] But it is not empty. Although variously phrased in various decisions [citation], it asks in substance whether the ruling in question *falls outside the bounds of reason*' under the applicable law and the relevant facts [citations].'" (*People v. Garcia* (1999) 20 Cal.4th 490, 503.) "Abuse may be found if the trial court exercised its discretion in an arbitrary, capricious, or patently absurd manner, but reversal of the ensuing judgment is appropriate only if the error has resulted in a manifest miscarriage of justice. [Citations.]" (*People v. Coddington* (2000) 23 Cal.4th 529, 587-588, overruled on other grounds in *Price v. Superior Court* (2001) 25 Cal.4th 1046, 1069, fn. 13.) "The governing canons are well established: 'This discretion ... is an impartial discretion, guided and controlled by fixed legal principles, to be exercised in conformity with the spirit of the law, and in a manner to subserve and not to impede or defeat the ends of substantial justice. [Citations.]' [Citation.] 'Obviously the term is a broad and elastic one [citation] which we have equated with "the sound judgment of the court, to be exercised according to the rules of law." [Citation.]' [Citation.] Thus, '[t]he courts have never ascribed to judicial discretion a potential without restraint.' (*Ibid.*) … '[A]ll exercises of legal discretion must be grounded in reasoned judgment and guided by legal principles and policies appropriate to the

particular matter at issue.' [Citation.]" (*People v. Superior Court* (*Alvarez*) (1997) 14 Cal.4th 968, 977.)

"A trial court abuses its discretion when the factual findings critical to its decision find no support in the evidence.… '[I]t would seem obvious that, if there were no evidence to support the decision, there would be an abuse of discretion.'" (*People v. Cluff* (2001) 87 Cal.App.4th 991, 998.) Thus, when the defense establishes that proper scientific procedures were not followed, and the prosecution fails to present "substantial evidence upon which to base a contrary conclusion," the prosecution has failed to carry its burden and the trial court's admission of the evidence constitutes an abuse of discretion. (*People v. Venegas, supra,* 18 Cal.4th at p. 93.)

D.   RELEVANT HISTORY

In *People v. Axell, supra,* 235 Cal.App.3d 836 , filed in October 1991, the court ruled that the general RFLP methodology used by Cellmark had gained general scientific acceptance. (*Id.* at pp. 853-863.) In August 1992, the court in *People v. Barney, supra,* 8 Cal.App.4th 798, relying primarily on *Axell*, rejected challenges to the scientific acceptance of the RFLP procedures conducted by both Cellmark and the FBI in two companion cases. (*People v. Barney, supra,* 8 Cal.App.4th at pp. 811-814; see also *People v. Venegas, supra,* 18 Cal.4th at p. 77.)

In October 1992, we filed our first opinion in the present case (*Pizarro I, supra,* 10 Cal.App.4th 57), in which defendant claimed the FBI's RFLP methodology had not been deemed scientifically accepted. Concerned by the differences between the protocols used by Cellmark in *Axell* and by the FBI in this case, and by the lack of evidence that the protocols were the same, we held the evidence insufficient to establish general scientific acceptance of the FBI's technique (*id.* at pp. 79-80), and remanded the case for a complete *Kelly* hearing. That hearing took place in 1994 and 1995. In its 1998 ruling, the trial court stated that we remanded the matter for a *Kelly* hearing to determine (1) whether the DNA testing method used by the FBI in this case was generally accepted

by the scientific community, and (2) whether the database used by the FBI in this case was generally accepted by the scientific community. The trial court found the evidence admissible, ruling as follows: "There is general acceptance in the scientific community of the DNA testing method used by the F.B.I." and "The data base used by the FBI to calculate statistical probability estimates was, and is, accepted in the scientific community." The court found that "the fixed bin product rule statistics are very conservative estimates of frequency," but did not directly mention third-prong issues regarding whether the FBI followed correct scientific procedures. The court denied the motion to exclude the DNA evidence and confirmed the conviction.

Two months after the trial court's ruling, the Supreme Court published *Venegas*, which concluded "the *Axell* and *Barney* opinions clearly established the general scientific acceptance, under *Kelly*'s first prong, of the basic RFLP methodology utilized by the FBI …." Unless there was proof the FBI's procedure was materially distinct from the basic RFLP procedure deemed approved by *Axell* and *Barney*, these opinions served as precedent for a first-prong challenge. (*People v. Venegas, supra,* 18 Cal.4th at pp. 53, 78-79.)[17] In effect, *Venegas* determined that, once the basic procedure was deemed accepted, the burden fell on the opponent of the evidence to show that the procedure in the case before the court differed materially from the accepted basic procedure. If the opponent could not do so, then the first prong remained satisfied by precedent.

We review this case in light of these developments.

---

[17] We note that this conclusion by *Venegas* calls into question the principle that one appellate court's decision is not binding on another appellate court. (See, e.g., 9 Witkin, Cal. Procedure (4th ed. 2001) §§ 934-935, pp. 971-974 and cases cited therein.)

III.   SCIENCE[18]

A.   *INTRODUCTION*[19]

Put simply, forensic DNA profiling is intended to demonstrate two facts -- first, that *the defendant could be the perpetrator because his DNA profile matches the perpetrator's*, and, second, that *the chance of finding a person in the population with the same DNA profile as the perpetrator's is a specific numerical probability.* The first fact allows the prosecution of the defendant to continue (a profile nonmatch would exonerate him); the second allows the jury to weigh the value of the first. (National Research Council, DNA Technology in Forensic Science (1992) p. 51 (hereafter NRCI).)

A genetic profile is much like a physical profile or composite sketch -- it is a compilation of traits to describe the perpetrator. The profiler or sketch artist attempts to include as many of the perpetrator's traits as possible because the more traits described, the more specific the sketch of the perpetrator and the more limited the pool of possible perpetrators. A physical profile that describes a male perpetrator as having black hair, blue eyes, and 5-foot-8-inch stature limits the pool of possible perpetrators to men with these three traits. If a fourth trait -- prominent ears, for example -- is added to the profile,

---

[18]   For what we hope will be greater clarity, we generally refer to the parties and their DNA samples as defendant (rather than suspect), perpetrator (rather than evidence or evidentiary), and victim. We recognize that what we refer to as the perpetrator's sample is more accurately referred to as the evidentiary sample because it may contain DNA from someone other than the perpetrator. But, for clarity and simplicity, and to stress the distinction between the perpetrator and the defendant, we generally adhere to this scheme. Also, because the perpetrator in this case is likely male, we occasionally use the masculine form.

[19]   Our reference to scientific literature is to provide the background necessary for the understanding of the issues in this case, not to resolve those issues. Although we cite various scientific sources, our discussion of the science and procedure of RFLP is derived in great part from a report entitled The Evaluation of Forensic DNA Evidence (hereafter NRCII), prepared in 1996 by the National Research Council.

the description becomes more specific and the pool of possible perpetrators decreases

further. In the same way, a genetic profile that describes a perpetrator as having certain

genetic characteristics at three DNA sites (loci) limits the pool of possible perpetrators to

people with those three traits. Again, if more loci are added to the profile, the

description's specificity increases and the pool of possible perpetrators decreases.

There are three basic theoretical determinations in RFLP genetic profiling:

(1) what is the perpetrator's profile? (2) does the defendant match that profile? and

(3) how rare is that profile in the population? (NRCI, *supra*, at p. 51.) The first and

second steps involve molecular biology, the third statistics and population genetics.

B. THEORETICAL SUMMARY

Returning to the physical sketch scenario, we summarize the theoretical steps of

RFLP genetic profiling, mindful that the genetic loci used for DNA profiling have

nothing to do with physical features; the comparison is strictly illustrative.

(1) Profiles -- What does the perpetrator "look like"?

Metaphorically: the perpetrator has black hair, blue eyes, and
5-foot-8-inch stature.

Genetically: the perpetrator possesses certain alleles at three
particular DNA loci.

(2) Matching -- Does the defendant "look like" the perpetrator?

Metaphorically: does the defendant also have black hair, blue eyes,
and 5-foot-8-inch stature?

Genetically: does the defendant's genetic profile match the
perpetrator's at each allele of the three loci?

If so, the defendant "looks like" the perpetrator and cannot be
excluded as a possible perpetrator; the case against the defendant
may proceed.

If not, the defendant does not "look like" the perpetrator and is
excluded as a possible perpetrator; the defendant is exonerated.

(3)   Statistical Probability -- How many people in the population "look like" the perpetrator?

Metaphorically: how often would we expect to find a person with black hair, blue eyes, and 5-foot-8-inch stature?

Genetically: how often would we expect to find a person whose alleles match the perpetrator's alleles? how common/rare is the perpetrator's genetic profile in the relevant population?

If the perpetrator's traits occur together *commonly*, the pool of possible perpetrators is not decreased significantly. A common profile such as this *benefits* the defendant (who shares this profile). He will say, "Lots of people look like the perpetrator. The fact that I look like him too is nearly meaningless."

If the perpetrator's traits occur together *rarely*, the pool of possible perpetrators is decreased dramatically. A rare profile such as this *incriminates* the defendant. The prosecution will say, "Almost no one looks like the perpetrator. The fact that you look like him means you probably *are* him."

C.   *PROCEDURAL SUMMARY*

In RFLP, these three theoretical steps are implemented with three procedural

steps: a molecular biology protocol to process the DNA and produce the genetic profiles;

a matching protocol to determine whether, accounting for measurement imprecision, the

perpetrator's and defendant's profiles match; and a statistical protocol to determine the

rarity of the profile and the probability of a match. (NRCI, *supra*, at p. 51.)

(1)   Profiles -- Molecular Biology[20]

1.   Extraction and isolation of the DNA samples (perpetrator,
        victim, and defendant)

2.   Cutting (digestion) of the DNA with a site-specific enzyme to
        create an enormous number of fragments

3.   Separation of the DNA fragments according to size by gel
        electrophoresis

4.   Transfer (blotting) of the separated DNA fragments from the
        gel onto a nylon membrane for convenience

5.   Sequential probing (hybridization) of the separated DNA
        fragments attached to the membrane with various radioactive
        probes that attach to only two VNTR regions on the
        fragments (one region from each parent)

6.   Autoradiography of each hybridization to memorialize the
        results on X-ray film

---

20   These well-established steps are described on pages 3-19 of the FBI's protocol
received into evidence as Exhibit 7. (See also NRCII, *supra*, at pp. 15-18, 42, 65-67;
Congress of the United States Office of Technology Assessment, Genetic Witness:
Forensic Uses of DNA Tests (1990) pp. 44-46 (hereafter OTA); DNA in Forensic
Science: Theory, Techniques and Applications (Robertson, et al. eds., 1990) pp. 68-70
(hereafter Robertson); Easteal, et al., DNA Profiling: Principles, Pitfalls and Potential
(1991) pp. 149-161 (hereafter Easteal); Coleman & Swenson, DNA in the Courtroom
(1994) pp. 36-41 (hereafter Coleman).)

When this procedure is completed, the autorads are analyzed to determine whether the defendant's profile matches the perpetrator's.

(2)   Matching[21]

1.   Preliminary visual examination of the autorads to determine whether each of the defendant's alleles appears to be the same size as each of the perpetrator's alleles (to eliminate obvious mismatches)

2.   Computerized examination to measure the size of each allele

3.   Calculation of $\pm 2.5\%$ "uncertainty windows" around each allele measurement

4.   Determination of whether, for each allele, the defendant's uncertainty window overlaps the perpetrator's uncertainty window (so that the alleles could actually be the same size)

5.   Declaration of a matching profile if overlap of uncertainty windows is found to occur at each allele

Last, the statistical probability of the perpetrator's profile in the population is calculated.

(3)   Statistical Probability[22]

1.   Calculation of a "statistical window" for each of the perpetrator's alleles

2.   Reference to database frequencies to assign a frequency to each of the perpetrator's alleles

3.   Calculation of the overall frequency of the perpetrator's complete DNA profile in the database population (also called the random match probability)[23]

---

[21]   Matching steps are described on pages 20 and 21 of the FBI's protocol (Exhibit 7). (See also NRCII, *supra*, at pp. 18-19, 43-44, 68-69; Easteal, *supra*, at pp. 161-163.)

[22]   The FBI's protocol (Exhibit 7) does not address statistical probability. (But see NRCII, *supra*, at pp. 20-21, 44-45, 68-69; Coleman, *supra*, at p. 45; part III.E., *post*.)

We now address in more detail these three steps -- profiles, matching, and statistical probability -- discussing both theory and procedure.

D.   PROFILES & MATCHING

1.   Theory

a.   Profiles

Determination of a person's genetic profile using RFLP relies on the differences in *length* of certain DNA regions. (NRCII, *supra*, at p. 65.) Because further discussion requires some familiarity with a few basic genetic principles and definitions, we repeat our summary in *People v. Brown, supra,* 91 Cal.App.4th 623, in which we analogized DNA to text:[24]

> "The genetics of a human cell can be compared to a library, the *genome*, composed of 46 'books,' each a single *chromosome*. The 'text' contained in the books is written in *DNA*, the chemical language of genetics. The 'library' is compiled by the owner's parents, each of whom contributes 23 books, which are then matched up and arranged together in 23 paired sets inside the sacrosanct edifice of the *nucleus*. During embryonic development, the original library is copied millions of times so that each cell in the human body contains a copy of the entire library.[25]

---

23   We recognize there is commentary stating that in complicated cases there may be a distinction between the frequency of the perpetrator's profile and the probability of a random match. (See, e.g., Weir, DNA Match and Profile Probabilities: Comment on Budowle et al. (2000) and Fung and Hu (2000) (2001) Forensic Science Communications.) However, this case apparently does not present such complications.

24   Footnotes are included and sequentially renumbered.

25   "There are a few exceptions, the two most significant being red blood cells and sex cells. Red blood cells contain no nucleus and therefore no chromosomes. Egg and sperm cells contain half the number of chromosomes of the rest of the body's cells, so that upon fertilization the complete number of chromosomes will be restored rather than doubled. Blood can be used to test a person's DNA because white blood cells contain DNA; sperm cells can be used because enough cells are tested that collectively the entire complement of DNA is represented. ([NRCII, *supra*, at] p. 12 ….)"

"Twenty-two of the twenty-three paired sets of books are entitled 'Chromosome 1' through 'Chromosome 22'; externally, the two paired books of each set appear to be identical in size and shape. However, the 23d set, which contains information on gender, consists of one book entitled 'Chromosome X' (given by the mother) and one book entitled either 'Chromosome X' or 'Chromosome Y' (given by the father and determining the sex of the library's owner). The 22 sets comprising 'Chromosome 1' through 'Chromosome 22' address an enormous variety of topics describing the composition, appearance, and function of the owner's body. In addition, they include a considerable amount of what appears to be nonsense. The two paired books of each set, one book from each parent, address identical topics, but may contain slightly different information on those topics. Thus, two paired books opened to the same page contain corresponding 'paragraphs,' but the text within those corresponding paragraphs may vary between the two books. For example, within the paragraph addressing eye color, one book may describe blue eyes while the other book of the set may describe brown eyes.[26]

"The two corresponding, but potentially variant, paragraphs in the two paired books are called *alleles*. If, for a particular topic (i.e., at a particular region or *locus* on the DNA), the allele from the mother is A and the corresponding allele from the father is B, the *genotype* at that locus is designated AB. The text of two corresponding alleles at any locus may be identical (a *homozygous* genotype, e.g., AA) or different (a *heterozygous* genotype, e.g., AB). Regardless, one person's genetic text is, in general, extremely similar to another person's; indeed, viewed in its vast entirety, the genetic text of one human library is 99.9 percent identical to all others. As a result, the text of most corresponding paragraphs varies only slightly among members of the population.

"Certain alleles, however, have been found to contain highly variable text. For example, alleles are composed of highly variable text when they describe structures requiring enormous variability. Also, some alleles appear to contain gibberish that varies greatly, or repeated strings of text that vary not in text but in repeat number. These variants

---

26   "The physical characteristic exhibited by the library's owner generally depends on the dominance or recessiveness of those two descriptions. Paragraphs describing a physical characteristic such as eye color, or describing a particular cellular product or function, are called *genes*. By definition, they contain a discrete amount of text sufficient to describe a particular thing or function."

(*polymorphisms*) found at certain *loci* render each person's library unique[27] and provide forensic scientists a method of differentiating between libraries (people) through the use of forensic techniques that rely on the large number of variant alleles possible at each variable locus…. Since each person receives two alleles for each locus, the number of possible combinations is further increased.

"When a sample of DNA -- usually in the form of hair, blood, saliva, or semen -- is left at the crime scene by a perpetrator, a forensic genetic analysis is conducted. First, DNA analysts create a genetic 'profile' or 'type' of the perpetrator's DNA by determining which variants or alleles exist at several variable loci. Second, the defendant's DNA is analyzed in exactly the same manner to create a profile for comparison with the perpetrator's profile. If the defendant's DNA produces a different profile than the perpetrator's, even by only one allele, the defendant could not have been the source of the crime scene DNA, and he or she is absolutely exonerated.[28] If, on the other hand, the defendant's DNA produces exactly the same genetic profile, the defendant could have been the source of the perpetrator's DNA -- but so could any other person with the same genetic profile. Third, when the perpetrator's and the defendant's profiles are found to match, the statistical significance of the match must be explained in terms of the rarity or commonness of that profile within a particular population -- that is, the number of people within a population expected to possess that particular genetic profile, or, put another way, the probability that a randomly chosen person in that population possesses that particular genetic profile.[29] Only then can the jury weigh the value of the profile match. [Citation.]."[30] (*People v. Brown, supra,* 91 Cal.App.4th at pp. 627-629; see NRCII, *supra*, at pp. 12-14, 60-65; NRCI, *supra*, at

---

27   "Identical twins, however, share essentially identical DNA."

28   "This, of course, assumes there was no error in handling of evidence or in laboratory procedure and analysis."

**29**   "This probability is often called the random match probability."

30   "'A determination that the DNA profile of an evidentiary sample matches the profile of a suspect establishes that the two profiles are consistent, but the determination would be of little significance if the evidentiary profile also matched that of many or most other human beings. The evidentiary weight of the match with the suspect is therefore inversely dependent upon the statistical probability of a similar match with the profile of a person drawn at random from the relevant population.' [Citation.]"

pp. 1-3, 32-33; OTA, *supra*, at pp. 3-6, 41-43; Kirby, DNA Fingerprinting (1992) pp. 7-34 (hereafter Kirby); Robertson, *supra*, at pp. 1-8, 31-33.)

The RFLP procedure used in Pizarro's case exploits genetic polymorphisms called variable number of tandem repeats (VNTRs), repeated sequences that abut one another without interruption. These DNA regions, which have no known product or function, vary greatly in repeat unit number and hence in length. The repeat unit is generally 15 to 35 base pairs (bp) long,[31] and the total length of the allele usually ranges from 500 bp to 10,000 bp.[32] The variation in allele length provides a method of comparison between the two alleles of a single person and between the alleles of different people. (NRCII*, supra*, at pp. 14-15, 65; NRCI, *supra*, at pp. 34-36, 38; OTA, *supra*, at pp. 43-44; Robertson, *supra*, at pp. 27-28.)

This concept can be explained schematically. Assume, for example, that the two alleles, one from each parent, possessed by the perpetrator at a particular locus are hypothetically referred to as 2 and 5 (rather than A and B, to denote their lengths). The perpetrator's genotype at this locus is 2,5. Schematically, the alleles, which have been enzymatically cut out of the long DNA molecule,[33] might appear as follows:
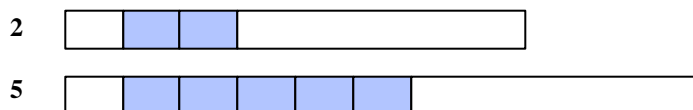
Perpetrator:

Since one locus or trait does not very specifically describe the perpetrator and thus does not narrow down the possible perpetrators significantly, just as describing the

---

[31]   Base pairs are the "letters" in the text of DNA.

[32]   The term "allele" technically refers to variants of a *gene*, but for convenience it is also used to refer to variants of a polymorphic locus.

[33]   The DNA has been cut with an enzyme that recognizes a specific sequence known not to exist within the VNTR sequence. Thus, the cuts always occur outside of and without disturbing those regions. (NRCII, *supra*, at p. 66.)

perpetrator as having black hair does not significantly reduce the field of possible perpetrators, additional genetic traits must be examined to flesh out the genetic sketch.

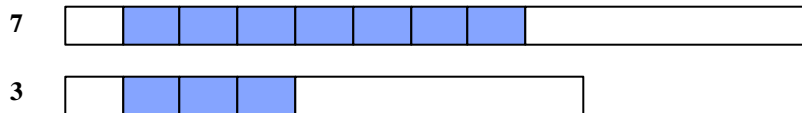The alleles at a second locus might appear as follows:

Perpetrator:

6

6
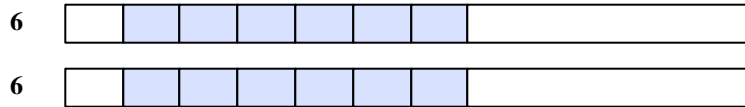
The perpetrator's genotype at this locus is 6,6.

The alleles at a third locus might appear as follows:
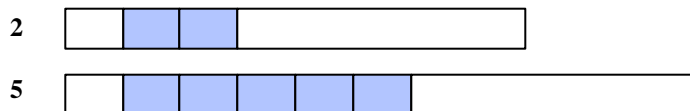
Perpetrator:

7

3

The perpetrator's genotype at this locus is 7,3. The genetic sketch of the perpetrator now consists of three loci and six alleles. (See NRCI, *supra*, at pp. 35-36; 45-46.)
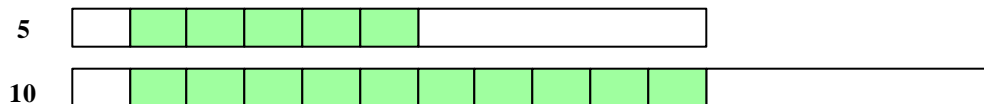
b. *Matching*

The matching step determines whether each of the defendant's alleles match the perpetrator's alleles -- that is, whether the defendant *could be* the perpetrator. Assume that the following sets of alleles are revealed at the first locus for the perpetrator and the defendant:

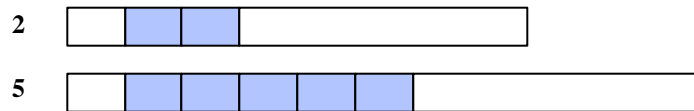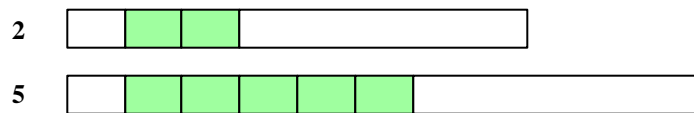**Perpetrator:**

2

5

Defendant:

5

10

Although the perpetrator and defendant share one allele (5), they do not share both, and therefore the defendant is excluded as a possible perpetrator. Identity between all of the alleles is required. When this lack of identity exists for even one allele at one locus, the defendant is exonerated. Stated metaphorically, the defendant's "hair color" (5,10) is not the same as the perpetrator's (2,5) and thus the defendant cannot be the perpetrator.

Assume instead that the following alleles are revealed for the perpetrator and the defendant at the first locus:

Perpetrator:

2

5

Defendant:

2

5

Now, both alleles at this locus match and the defendant is not excluded as a possible perpetrator. The defendant's "hair color" matches the perpetrator's. If all of the defendant's alleles at the remaining two loci match the perpetrator's, the overall profiles match, and the defendant is a possible perpetrator. If, on the other hand, even one of the defendant's alleles fails to match, the defendant is no longer a candidate. (NRCII, *supra*, at p. 18; NRCI, *supra*, at p. 4.)

While these diagrams suggest otherwise, the unfortunate reality of RFLP analysis is that it cannot display the actual alleles or measure their exact lengths.[34] Thus,

---

[34] If, however, each allele were directly sequenced to determine its exact length, an allele from the perpetrator and the corresponding allele from the defendant could be compared unambiguously. If the base pair lengths were identical, the allele lengths

determination of a match between two alleles is complicated by the system's measurement imprecision. (NRCII, *supra*, at p. 139; NRCI, *supra*, at pp. 38, 61.) We turn next to the procedure by which RFLP, despite its limitations, determines whether an allele from the defendant is the "same" length as the corresponding allele from the perpetrator.

2. Procedure

a. Profiles[35]

1. Extraction of DNA

In practice, comparison of allele lengths by RFLP begins with the extraction of DNA from the crime scene evidence -- from hair, a blood stain, a saliva stain, or, as in this case, a vaginal swab. This evidentiary sample will likely contain the perpetrator's DNA. For comparison, blood samples are taken from the victim and the defendant, and DNA is extracted from those samples also.

2. Enzymatic Digestion of DNA

Once purified, the DNA in the three separate samples is cut into millions of fragments of varying lengths by a restriction enzyme that cuts at a specific short sequence wherever it exists in the DNA. The spacing of these cutting sites on the DNA varies slightly from person to person; thus, the array of fragments produced by the cutting will vary slightly from person to person. If the array of fragments in two samples is the same

---

would be the same. If the base pair lengths were off by even a single base pair, the allele lengths would be different and the defendant could not be the perpetrator.

[35]   For an overview of the following molecular biology procedure, see NRCII, *supra*, at pages 15-18, 42-45, 65-69; NRCI, *supra*, at pages 36-40; and Robertson, *supra*, at pages 74-79. Unless otherwise noted, in this section (part III.D.2.a.) we rely on NRCII, *supra*, at pp. 15-18, 42-43, 66-68; NRCI, *supra*, at pp. 36-40; OTA, *supra*, at pp. 46-47; Kirby, *supra*, at pp. 51-73, 94-104, 110-116; Robertson, *supra*, at pp. 62-65, 69-70; Coleman, *supra*, at pp. 36-37, 40-41; and Easteal, *supra*, at pp. 85-87.

-- that is, if the *lengths* of the fragments are the same -- then the DNA in the two samples could be from the same person. Consequently, to determine whether the perpetrator and defendant could be the same person, the fragments in their DNA samples must be compared.

3. Gel Electrophoresis of DNA Fragments

To accomplish this, the DNA fragments are first separated by size using gel electrophoresis. A portion of each DNA sample is added to separate wells near the end of a horizontal slab of dense gel (the wells do not penetrate through to the bottom of the gel). The gel is then placed in an electrical field. The DNA fragments, which are negatively charged, travel through the gel toward the positive pole, their speeds depending on their size and ability to maneuver through the gel structure. The gel is something of a molecular obstacle course -- the shorter, more agile DNA fragments move through it more quickly and advance farther down the gel in a given amount of time than the longer, more cumbersome fragments. When the electrical current is turned off, the DNA fragments in a sample are spread down a lane extending from the well to the other end of the gel. The fragments of DNA form what are called bands. For reference, size standards (often called molecular weight markers), which are DNA fragments of known sizes, are also run on the gel. (Fig. 1.) This electrophoretic step serves two purposes: (1) it spreads out the invisible contents of each DNA sample, preparing the DNA fragments for further study, and (2) it allows estimation of the sizes of the DNA fragments.

Fig. 1. *Bands of DNA on Gel.*
NOTE: View looking down on the horizontal gel. The wells are at the top of t he gel.
The smallest fragments travel farthest and are at the bottom of the gel.
In all diagrams, Size stds = size standards; V = victim; Def = defendant; Perp = perpetrator.

4.   Blotting of DNA Fragments onto Membrane

Since the gel is fragile and short-lived, the DNA fragments are transferred

(blotted) from the gel onto a durable nylon membrane, the DNA retaining the same band

formation. But before the gel is blotted, it is soaked in a chemical that separates or

"unzips" the two strands of every DNA fragment within the gel (i.e., the double-stranded

DNA fragments are denatured into single-stranded fragments). Now the single-stranded

DNA fragments can be analyzed.

5.   Hybridization with Radioactive Probes

Recall that RFLP seeks to identify the polymorphic VNTR regions that vary in

length between people. After the DNA is cut into fragments, the two specific VNTR

regions possessed by a person are contained in two of the many fragments now spread

down the lane of the gel, but it is impossible to tell which fragments contain the VNTR

regions by simply looking at the DNA fragments on the gel or membrane. These regions

must be sought out and flagged by a molecular probe. The highly specific bonds formed between the two strands of DNA make this possible. The single-stranded DNA fragments immobilized on the membrane are available for bonding with other single-stranded DNA fragments, but only if the sequences of the two fragments are complementary. Thus, if a known sequence is being sought in the DNA (e.g., a specific VNTR region), a short, single-stranded DNA fragment (probe) with a complementary sequence can be created to seek out that sequence among the fragments attached to the membrane. Every probe molecule is radioactively tagged to allow visualization of the invisible DNA fragments later.

Many copies of the radioactive probe are added to liquid in a container, then the membrane is added and sloshed about for several hours. When a probe molecule happens to wash across a complementary DNA fragment attached to the membrane, it will bind tightly (hybridize) to it. Then, when the excess probe is washed off the membrane, the remaining probe molecules are bound only to the VNTR regions in the two alleles per DNA sample. The hybrids formed between the radioactive probe molecules and the complementary VNTR regions on the membrane are radioactive and will be visualized in the next step. (Other radioactive probe molecules specific to the size standards are also added to the hybridization liquid so the standards will also be identifiable.) (Fig. 2.)
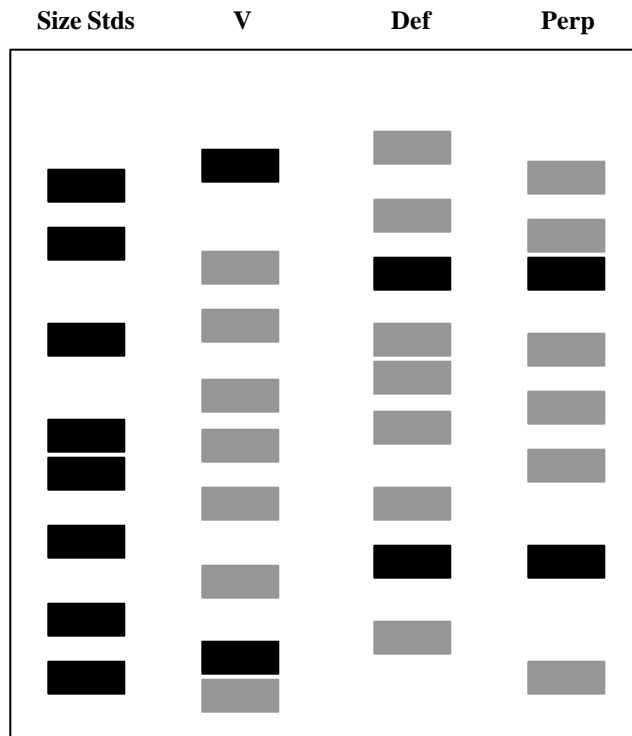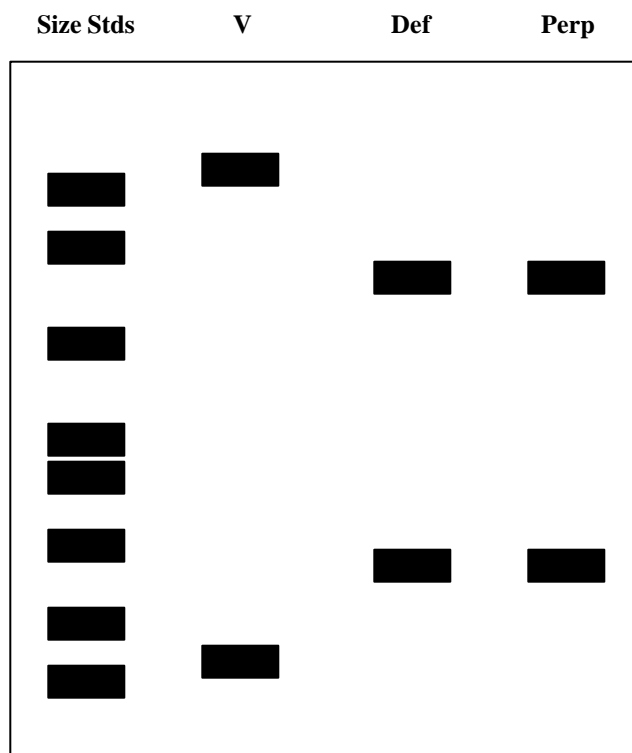
Fig. 2. *Bands of Probe-Bound DNA on Membrane.*
NOTE: The probe binds to two alleles per DNA sample, represented by the two black bands in the V, Def, and Perp lanes. Different probe binds to the size standards so they too will be visible later.

6.   Autoradiography

When an X-ray film is placed over the membrane, the radioactive tags reveal the positions of the invisible probe-bound alleles and size standards. The other nonradioactive DNA bands remain invisible. The resulting autorad becomes the DNA evidence in the case. (Fig. 3.) From the autorad, scientists determine the approximate sizes of each person's two VNTR alleles, based on comparisons with the size standards on the same autorad. Once an autorad has been made from the hybridized membrane, the probe is chemically stripped off the membrane, and the procedure is repeated with a different probe specific to another VNTR locus. The membrane can be reused for several different probes (but there is a limit because the DNA attached to the membrane is gradually stripped off).

Fig. 3. *Bands of Probe-Bound DNA on Autorad.*
NOTE: The bands of probe-bound DNA are finally visible on the X-ray film.

  This molecular biology procedure reveals the two VNTR alleles possessed by the perpetrator, the defendant, and the victim at each locus tested. If three loci are tested, three autorads are produced, each showing one or two bands for each person's DNA. Usually, the two alleles possessed by a person are different lengths and therefore appear as two bands (a heterozygous genotype). If the two alleles are the same or very similar in length, they will appear as a single band (a homozygous genotype). (NRCII, *supra*, at p. 69; Aitken, Statistics and the Evaluation of Evidence for Forensic Scientists (1995) p. 207. A person's bands from all the autorads together make up that person's genetic profile. From this point forward, these are the fragments, bands, or alleles we will be discussing.

b.   *Matching*

1.   Measurement Imprecision -- Uncertainty Windows [36]

The first three steps of matching -- visual examination, computerized measurement, and calculation of uncertainty windows -- are measurement steps. First, the scientist visually compares one of the defendant's bands with the corresponding perpetrator's band on an autorad to see if they appear to be the same size (i.e., are in the same position because they traveled the same distance on the gel). If the two bands are an obvious mismatch, the analysis ends. If the bands appear to match (as in figure 3, *ante*), they are measured by the computer, using the size standards on the same autorad for comparison. The RFLP system, however, is not capable of precise measurements, and accommodations must be made to account for its imprecision. It is therefore necessary to understand that in this system an allele's *measurement* is not always the same as its *true length*. (See NRCII, *supra*, at pp. 68, 139-140; NRCI, *supra*, at pp. 53-54; Easteal, *supra*, at pp. 87-88.) We attempt to explain the features of RFLP using the following scenario.

a.   *Height*

Suppose that rather than measuring and comparing the lengths of alleles using RFLP, scientists want to measure and compare the heights of people using a 12-inch ruler. Although the ruler measurement system is certainly not the most precise method for measuring height, it is convenient, economical, and widely accessible. The scientists, aware of the system's shortcomings, first test the system to determine the extent of its measurement imprecision -- its margin of error. They begin with a test person who, in this scenario, has been measured as exactly 5 feet 6 inches (66") by a different, very

---

[36]   Unless otherwise noted, in this section (part III.D.2.b.1.) we rely on NRCII, *supra*, at pages 7, 18-19, 43-44, and 139-142.

precise method. Then, to test their measurement system, the scientists repeatedly

measure that 5-foot-6-inch test person with a ruler, and they record the measurements to

see where they fall relative to the known true height of 5 feet 6 inches. They obtain the

following measurements, among others. (Fig. 4.)
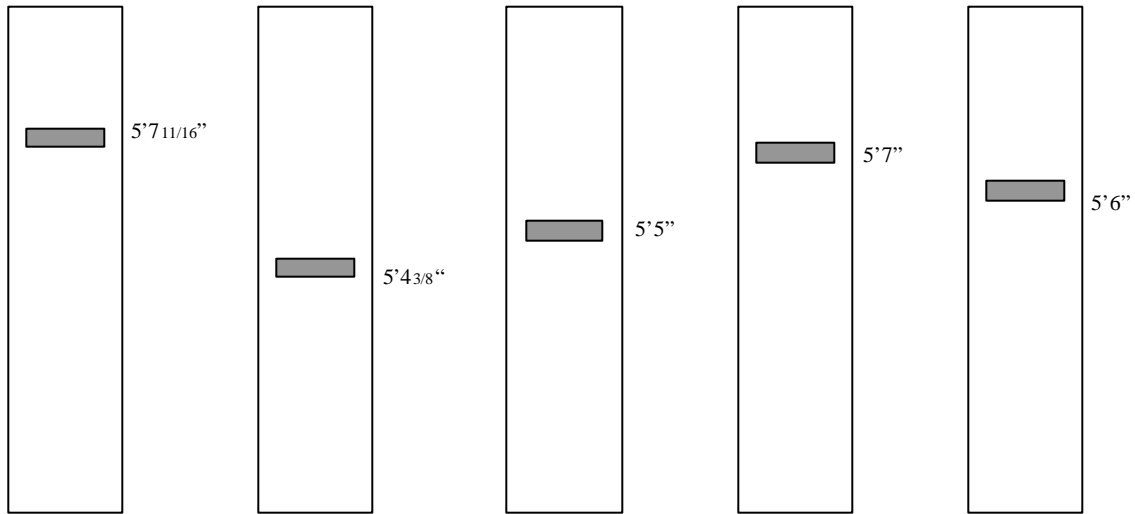
Fig. 4. *Repeated Measurements of a Test Person.*
NOTE: These five measurements represent the many repeated measurements taken of a 5-foot-6-inch person.

The scientists find that all their measurements, when compiled, happen to fall

between 5 feet 4 and 3/8 inches (64.37") and 5 feet 7 and 11/16 inches (67.69") -- within

a $\pm$ 2.5% range of the true 5-foot-6-inch measurement.[37] (Fig. 5.)

---

[37]   66" + 66"x = 67.69";   66x = 67.69 - 66;   x = + 1.69/66;   x = + 0.025.
66" + 66"x = 64.37";   66x = 64.37 - 66;   x = - 1.63/66;   x = - 0.025.
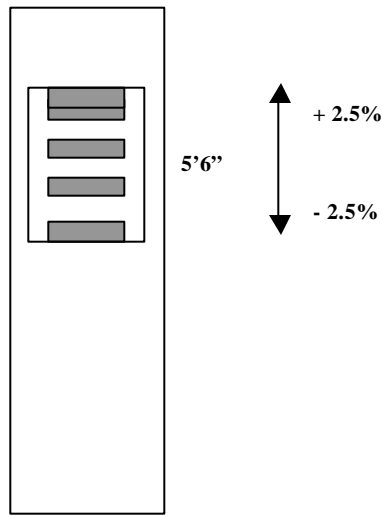
Fig. 5. *Compiled Measurements of a Test Person.*
NOTE: Repeated test measurements fall within a ±2.5% range of the true height.

Consequently, the measurement imprecision, margin of error, or "uncertainty window" for this system is ± 2.5%, and the scientists know that every measurement they take using the ruler method could be off by this much.[38] (See NRCII, *supra*, at p. 140.) Although a person may be *measured* as a certain height, *his or her true height could actually fall anywhere within the ± 2.5% uncertainty window* around that measurement. The system simply cannot define the measurement more precisely.

b.   RFLP

The same concept applies to the RFLP system. Laboratories using RFLP must first establish the imprecision of their RFLP systems by repeatedly measuring DNA fragments that have been exactly sequenced and measured. (NRCI, *supra*, at pp. 61-62.) In Pizarro's case, the FBI scientists, like the height-measuring scientists in our scenario, found that all their test measurements fell within a ± 2.5% range of the true length of a

---

38   We use the term "uncertainty window," which is cogently used by NRCII (e.g., NRCII, *supra*, at p. 140), because we find it clear and functional. See part III.E.2.a., *post*, for the distinction between "uncertainty window" and "match window," and part III.E.2.c., *post*, for an explanation of "statistical window."

DNA fragment. For example, if the FBI had repeatedly measured a test DNA fragment known to be exactly 5,000 bp, all its measurements would have fallen between about 4,875 and 5,125 bp. (Fig. 6.)
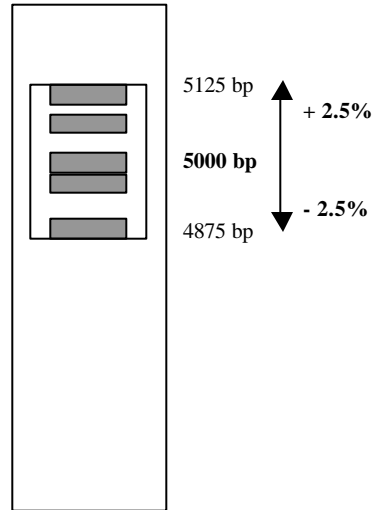
Fig. 6. *Compiled Measurements of a Test DNA Fragment.*
NOTE: Repeated test measurements fall within a ± 2.5% uncertainty window.

Results such as these reveal that all the FBI's RFLP measurements can be off by as much as 2.5% in either direction. Again, although an allele is *measured* to be a certain length, *its true length could exist anywhere within the ± 2.5% uncertainty window* around that measurement.

2.   Comparison Between People --
Overlapping Uncertainty Windows [39]

The remaining two steps of matching -- determination of uncertainty-window overlap and declaration of a matching profile -- concern the comparison between two people's measurements.

---

[39]   Unless otherwise noted, in this section (part III.D.2.b.2.) we rely on NRCII, *supra*, at pages 7, 18-20, 44-45, and 139-142.

In the height measurement hypothetical, the scientists would like to use the ruler to measure two particular people to determine whether they are the same height. But this is impossible to determine positively because each measurement is imprecise, and thus the question is actually whether the two people *could be the same height*, or, whether within the limitations of the measurement system the heights are indistinguishable and therefore *can be considered the same* -- a match.[40]

The scientists first measure Person 1 as 5 feet 8 inches, which means that his or her true height is somewhere between 5 feet 6 inches and 5 feet 10 inches (5'8" ± 2.5%). (Fig. 7.)

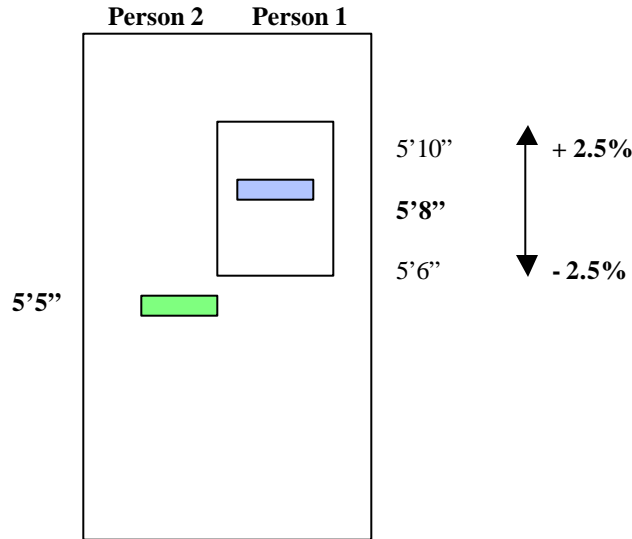**Person 1**

5'10"   + **2.5%**

5'8"

5'6"   - **2.5%**
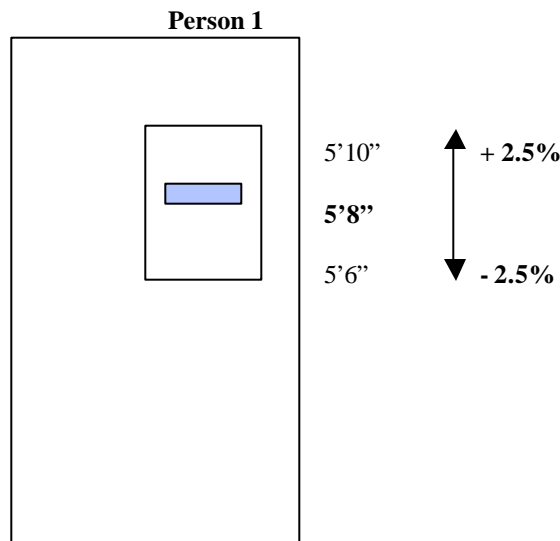
Fig. 7. *Person 1's Measurement Surrounded by ± 2.5% Uncertainty Window.*

Next, the scientists measure Person 2 as 5 feet 5 inches. Since Person 2's measurement falls outside the 5-foot-6-inch to 5-foot-10-inch uncertainty window around

---

[40]   Of course, it is necessary in this scenario to assume the scientists cannot stand the people side by side to see whether they are the same height.

Person 1's measurement, does this mean Person 2's height cannot be the same as Person 1's? (Fig. 8.)

**Fig. 8.** *Person 2's Measurement Falls Outside Person 1's Uncertainty Window.*

No; Person 2's height measurement is subject to the same $\pm 2.5\%$ imprecision because the same imprecise method was used to take the measurement. Thus, a $\pm 2.5\%$ uncertainty window must be drawn around *both* people's height measurements. As a result of the measurement imprecision, the scientists know only this: Person 1's height is somewhere between 5 feet 6 inches and 5 feet 10 inches (5'8" $\pm 2.5\%$), and Person 2's height is somewhere between 5 feet 3 inches and 5 feet 7 inches (5'5" $\pm 2.5\%$). Because Person 1 could actually be as short as 5 feet 6 inches and Person 2 could actually be as tall as 5 feet 7 inches (i.e., their uncertainty windows overlap), their heights are *indistinguishable by this system.* (Fig. 9.)
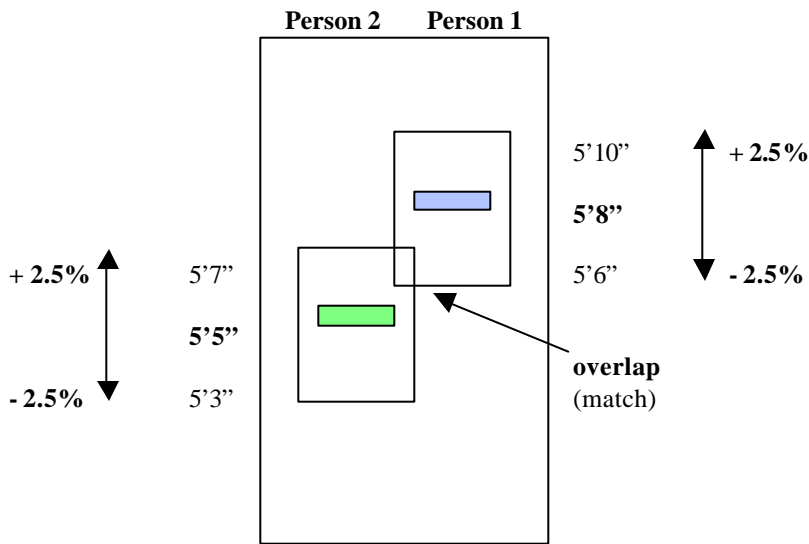
Person 2   Person 1

Fig. 9. *Overlapping ± 2.5% Uncertainty Windows.*
NOTE: Person 2's height matches Person 1's because they are indistinguishable within the measurement ability of the system. Person 1 and Person 2 are deemed to be the same height even though the measurements of their heights differ by three inches.

Note that these measurements match because the true heights of Person 1 and Person 2 could exist wi thin the overlap of the uncertainty windows, and therefore these two people *could actually be* the same height despite the differences in their measurements. (Fig. 10.)
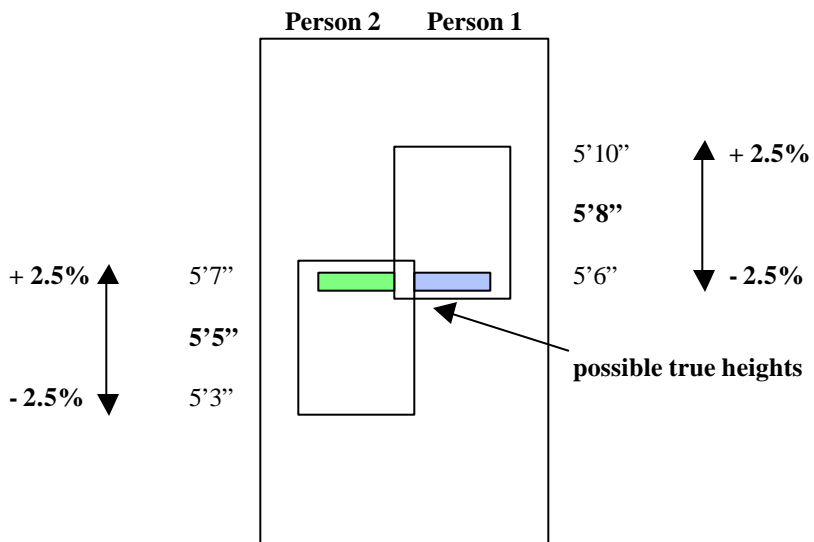
Person 2   Person 1

Fig. 10. *Possible True Heights.*
NOTE: The true heights of Person 1 and Person 2 could both lie within the overlap.

46.

b.   RFLP

Similarly, when the FBI scientists measure the length of a defendant's allele to determine whether it is the same length as the perpetrator's allele, they must surround both measurements with a ± 2.5% window of uncertainty. If these two windows overlap, the FBI declares the two alleles a match.

Consider the following alleles at a single locus. Assume that the perpetrator's two alleles are measured by the RFLP system as 1,000 bp and 800 bp; the defendant's alleles are measured as 960 bp and 870 bp. The autorad would appear roughly as follows:[41]
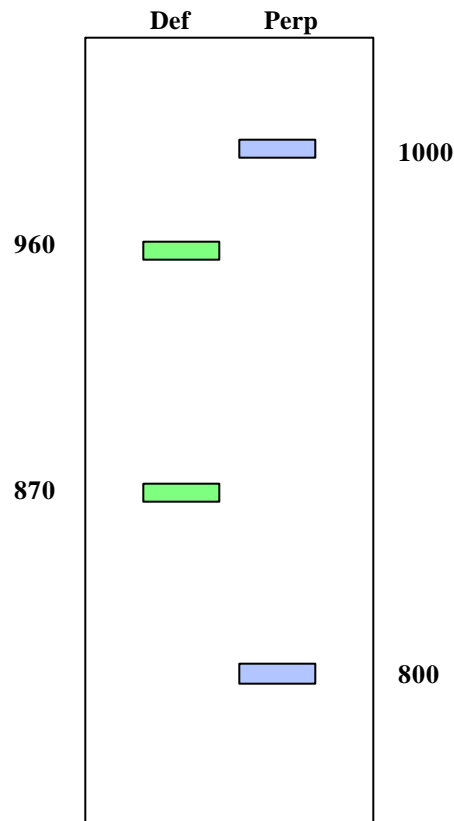
Fig. 11. *Autorad of Two Sets of Alleles at a Locus.*
NOTE: ☐ = Perp's 1st allele; ☐ = Perp's 2d allele;
☐ = Def's 1st allele; ☐ = Def's 2d allele.

---

[41]   We recognize that visual examination of this autorad would almost certainly have already established an obvious mismatch, but we proceed for the sake of explanation.

For the perpetrator's 1,000 bp allele, the ± 2.5% uncertainty window around it reveals that the allele's true length is somewhere between 975 bp and 1,025 bp (1,000 bp ± 2.5%). The defendant's corresponding allele measures 960 bp long, which means the actual size of that allele falls somewhere between 936 bp and 984 bp (960 bp ± 2.5%). Because the two uncertainty windows overlap (984 exceeds 975), the defendant's allele is said to match the perpetrator's -- the 1,000 bp and 960 bp alleles are considered to be the same length because they *could actually be* the same length (i.e., both fall within the overlap). (Fig. 12.)
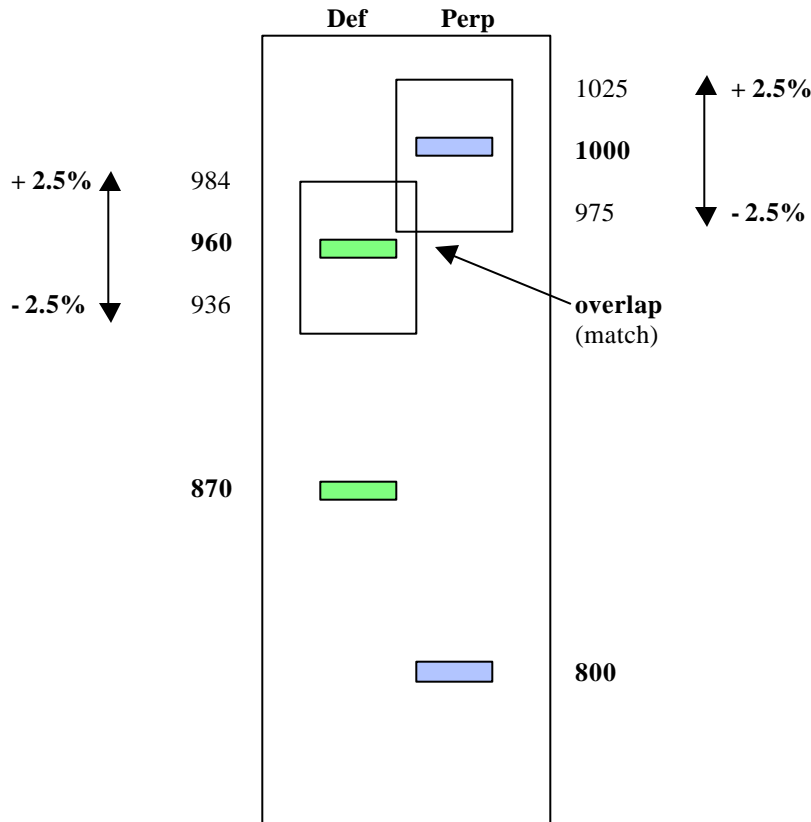
Fig. 12. *Overlapping ± 2.5% Uncertainty Windows.*
NOTE: The first set of bands match because their uncertainty windows overlap.

Uncertainty windows are also drawn around the other two corresponding alleles on the autorad to determine whether they match. Because their windows do not overlap, the alleles do not match. (Fig. 13.)

Def   Perp

1025   ▲ + 2.5%

1000

975   ▼ - 2.5%

+ 2.5% ▲   984

960

- 2.5% ▼   936

overlap
(match)

+ 2.5% ▲   891

870

- 2.5% ▼   848

no overlap
(no match)

820   ▲ + 2.5%

800

780   ▼ - 2.5%
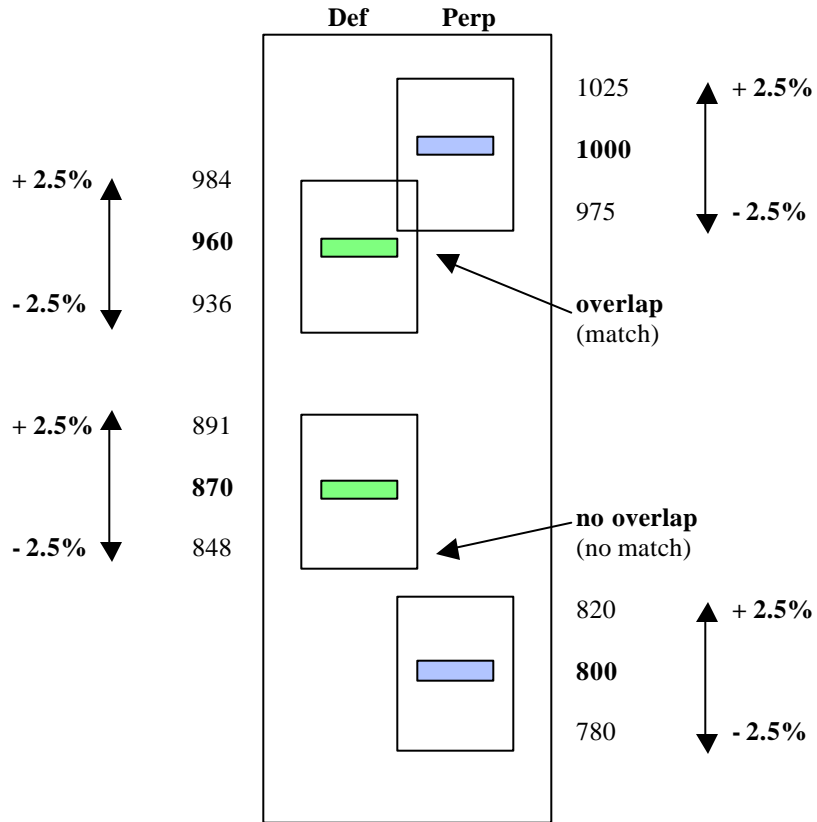
Fig. 13. *Two Sets of ± 2.5% Uncertainty Windows.*
NOTE: The first two bands match, but the second two do not.

Each of the defendant's alleles on each autorad is compared to each of the perpetrator's corresponding alleles in this manner. If even one pair of uncertainty windows fails to overlap, such that one allele is determined a mismatch, the defendant is excluded as a donor of the perpetrator's DNA and he is exonerated (as he would be in figure 13, *ante*). If all the alleles match, the defendant is a possible donor of the perpetrator's DNA and could have committed the crime. This is the method by which RFLP measures and compares the lengths of the perpetrator's and the defendant's alleles.

E.   STATISTICAL PROBABILITY

1.   Theory

It is important to understand that a match between all the defendant's alleles and all the perpetrator's alleles (i.e., between their profiles) does not signify an *absolute* match between the entirety of the perpetrator's DNA and the entirety of the defendant's DNA, which of course would absolutely prove the perpetrator and the defendant are the same person. The match is actually between only a few or several regions of an enormous amount of DNA, and therefore it cannot absolutely prove identity. What it does prove is that the defendant *could be* the perpetrator. However, this information standing alone is not particularly helpful to the jury; it is in fact unwieldy, overwhelming, even irresistible. If the jury is told simply that the defendant's genetic profile matches the perpetrator's profile and thus the defendant could be the perpetrator, the jury -- awed by the sophistication and incomprehensibility of the evidence -- will naturally respond by assuming the match absolutely proves identity. For this reason, courts have insisted that the prosecution provide comprehensible evidence regarding the meaning or significance of the match. (See NRCII, *supra*, at pp. 192-199; NRCI, *supra*, at pp. 9-11, 44; Easteal, *supra*, at pp. 90-91.)

The determination of what is often called the "significance of the match" is an assessment of *how incriminating* it is that the defendant's profile matches the perpetrator's. It quantifies the commonness or rarity of the perpetrator's profile in the population, thereby allowing the jury to weigh the evidence that the defendant possesses the same profile. It is a numerical assessment that asks, in essence, are there multitudes of people who possess the perpetrator's profile, or exceedingly few people who possess the perpetrator's profile? The rarer the profile, the more incriminating the defendant's possession of it. (See NRCII, *supra*, at p. 127; NRCI, *supra*, at p. 44.)

First, a numerical frequency is determined for each of the perpetrator's alleles, one at a time; then the genotype frequency (for an allele pair) at each locus (for each autorad)

is calculated; and finally the overall frequency of the perpetrator's DNA profile is calculated to determine how many people in the relevant population would be expected to possess or match the perpetrator's profile (or, stated differently, the probability that a random person in the population would possess that profile). (See NRCII, *supra*, at pp. 90-93, 122, 127; NRCI, *supra*, at pp. 44, 77-79.) The product is expressed as, for example, 1 in 10,000 or 1 in 500 million.

2. Procedure

a. Calculation of Match Window[42]

Calculation of the "match window" is the fundamental step in the determination of the statistical significance of the match between each of the defendant's alleles and each of the perpetrator's alleles. At the outset, it is critical to establish some terminology -- in particular, to differentiate between the uncertainty window and the match window.

The *uncertainty window*, which we have already discussed, surrounds a single allele measurement and represents the imprecision of that single measurement. It encompasses all the lengths that, due to measurement imprecision, that particular allele *could actually be*. It is applied to both the defendant's and the perpetrator's alleles alike. The subsequent determination that the defendant's allele matches the perpetrator's allele requires that the defendant's uncertainty window overlap the perpetrator's uncertainty window such that the true alleles within the windows could actually be the same length. (NRCII, *supra*, at pp. 18-19, 44, 139-142.) (See fig. 12.)

The *match window*, on the other hand, completely disregards the defendant's allele measurement. It defines and includes the *entire range* of allele measurements in the population that could be the same as the *perpetrator's* allele. It is also a product of the

---

42   Unless otherwise noted, in this section (part III.E.2.a.) we rely on NRCII, *supra,* at pages 18-20, 44-45, and 139-143.

measurement imprecision because only allele measurements within a distance permitting an overlap of uncertainty windows will match the perpetrator's allele measurement. Thus, the match window is about two times the size of the uncertainty window. (NRCII, *supra*, at pp. 20, 44-45, 139-143.)

<div align="center">1.   Height</div>

Recall that when Person 2 was compared with Person 1, their 5-foot-5-inch and 5-foot-8-inch height measurements were indistinguishable -- considered the same -- because their uncertainty windows overlapped. That was the matching step, which includes consideration of Person 2 and the *uncertainty window* around his or her height.

Calculation of the *match window* around Person 1's height, however, does *not* consider Person 2. It asks, which height measurements could possibly be the same as Person 1's true height? and, how far can a measurement be from Person 1's measurement such that their uncertainty windows will still overlap? In order to allow the two ± 2.5% uncertainty windows to minimally overlap, two height measurements must be within 5% of each other. Here, a measurement of just over 5 feet 4 inches (5'8" - 5%) is the minimum match. Any person measured between 5 feet 4 inches and 5 feet 8 inches could be the same height as Person 1. (Fig. 14.)
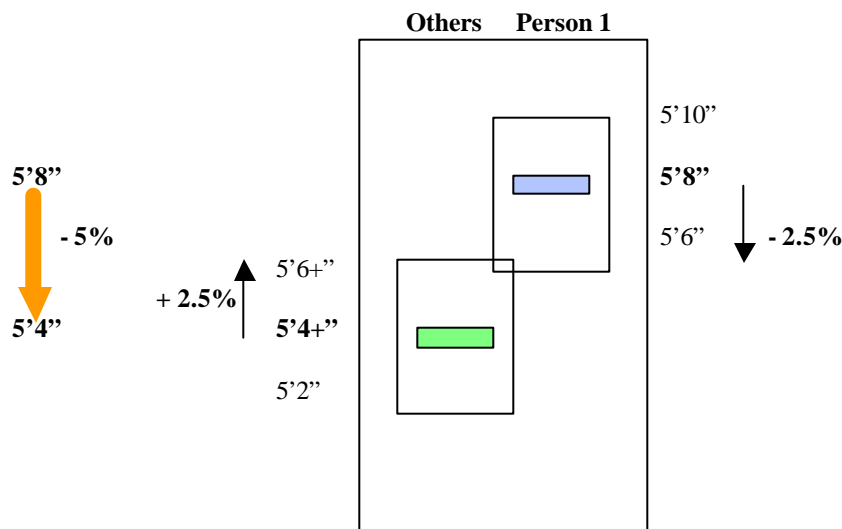
Fig. 14. *Distance Between Minimally Overlapping Uncertainty Windows.*
NOTE: [green box] = the height measurement of any other person. All heights measured to be within the - 5% range are considered to match the 5-foot-8-inch measurement.

Of course the measurement imprecision is not unidirectional and there is an approximately equal range of measurements *above* Person 1's measurement that would also be considered a match to the 5-foot-8-inch measurement. A measurement of just under 6 feet (5'8" + 5%) still overlaps Person 1's uncertainty window. Thus, the match window includes people who are measured as slightly shorter, as well as slightly taller, than Person 1. As a result, the entire range of measurements that would match the 5-foot-8-inch measurement range from 5 feet 4 inches to 6 feet 0 inches -- a ± 5% match window. Anyone measured as any height between 5 feet 4 inches and 6 feet 0 inches could be the same height as Person 1, and is by definition contained within Person 1's match window. (Fig. 15.)
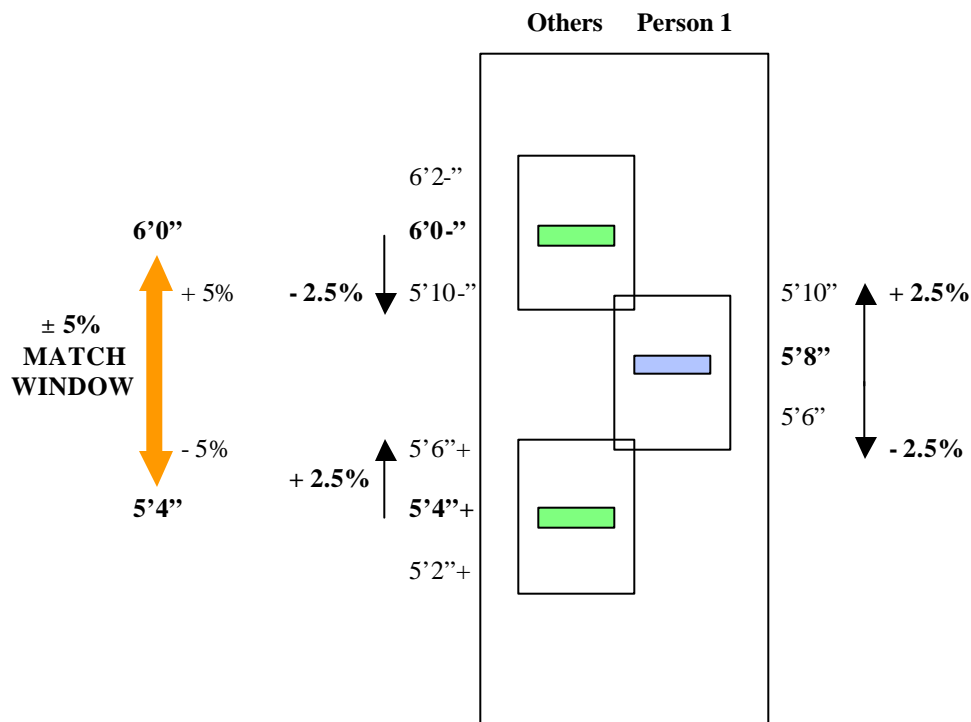
Fig. 15. *± 5% Match Window Around Person 1's Height Measurement.*
NOTE: [green box] = the height measurements of any other people. All heights measured to be within the ±5% window are considered to match the 5-foot-8-inch measurement.

2.   RFLP

In the RFLP example, the defendant's 960 bp allele matched the perpetrator's 1,000 bp allele because their uncertainty windows overlapped. (See fig. 12, *ante*.) Now,

the match window is calculated from the perpetrator's allele, without regard to the defendant's allele. As in the height analogy, all measurements within ± 5% of the perpetrator's allele measurement are considered to match the perpetrator's allele measurement. Here, all allele measurements between 950 bp and 1,050 bp could be the same length as the perpetrator's allele. (Fig. 16.)
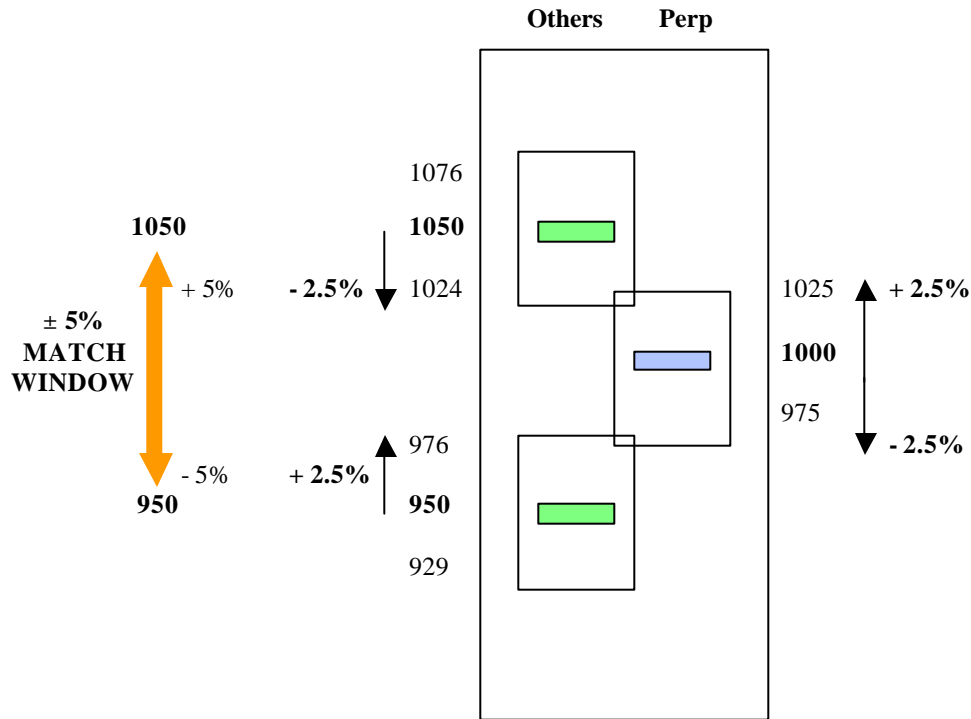
Fig. 16. _± 5% Match Window Around Perpetrator's Allele Measurement._
NOTE: [green box] = the allele measurements of any other people. All alleles measured to be within the ±5% window are considered to match the 1,000 bp measurement.

b.   _Creation of Database Allele Frequency Table_[43]

As we have discussed, the match window describes and encompasses all the possible measurements that match, or could be the same as, a perpetrator's allele. The next step -- frequency calculation -- estimates _how often that allele occurs in the population_ (i.e., how many alleles fall into the match window).

---

43   Unless otherwise noted, in this section (part III.E.2.b.) we rely on NRCII, _supra_, at pages 20-22, and 95; NRCI, _supra_, at pages 77 and 85-86.

It may be helpful to imagine that the match window determines *what* the thing is, and the frequency determines *how many people* possess that thing. The thing may be a trait such as reddish hair; the frequency may be 3 out of 100 people (0.03 or 3%) who have reddish hair. The thing may be a height of 5 feet 4 inches to 6 feet 0 inches; the frequency may be 77 out of 100 people (0.77 or 77%) who are 5 feet 4 inches to 6 feet 0 inches tall. In the RFLP context, the thing may be an allele length of 950 bp to 1,050 bp; the frequency may be 85 out of 1,000 alleles (0.085 or 8.5%) that are 950 bp to 1,050 bp long.

Frequencies such as these are easily estimated using population databases. We explained the underlying idea in *Brown*:

> "For example, if the victim reports that the perpetrator had blue eyes and abnormally short fingers (brachydactyly), forensic scientists will need to know how rare the combination of blue eyes and brachydactyly is in the population. That determination requires knowledge of the *separate* frequencies of these two traits in the population -- how many people have blue eyes and how many people have brachydactyly. But it is impractical to actually examine the entire population to count every person with blue eyes and every person with brachydactyly; instead, scientists create a database of randomly selected people, and use the frequencies of the traits of that group of people to represent the entire population. If among the people used to compile the database the occurrence of blue eyes is fairly common and the occurrence of brachydactyly is very uncommon, then the probability of the two traits occurring *together* will be extremely rare. That determination, derived from the database, is presumed to apply to the entire population the database was created to represent. Therefore, the reasoning goes, if very few people are expected to have both traits -- that is, if the profile is rare -- the probability is greater that a defendant who possesses both traits is in fact the perpetrator.
>
> "In reality, forensically important alleles do not manifest themselves in obvious physical traits, but the idea is the same. Because allele frequencies cannot be determined from external appearances, preparation of a database requires collection of DNA samples (usually blood) from unrelated individuals in the relevant population, genetic analysis of each DNA sample to determine the alleles present at each locus tested, tally of the various alleles at each locus, and statistical analysis of the tallied results to determine the frequency of each allele (the allele frequency) and then the

55.

frequency of every possible corresponding set of two alleles (the genotype frequency) at each locus. These database frequencies become standard values from which a perpetrator's profile can be given a numerical probability of existing in a population." (*People v. Brown, supra*, 91 Cal.App.4th at pp. 629-630, fns. omitted.)

1.   Height

In our height measurement hypothetical, the scientists might randomly select 100 people to create a database. They measure the height of each person, collect the measurements together, and arrange them in order. There might be five people whose measurements fall between 5 feet 3 inches and 5 feet 4 inches, seven people between 5 feet 4 inches and 5 feet 5 inches, 12 people between 5 feet 5 inches and 5 feet 6 inches, and so on. The hypothetical database height frequency table might look something like figure 17.

Height	# of People	Freq. (#/100)
6'4"	1	.01
6'3"	0	0
6'2"	1	.01
6'1"	2	.02
6'0"	1	.01
5'11"	4	.04
5'10"	6	.06
5'9"	10	.10
5'8"	14	.14
5'7"	8	.08
5'6"	21	.21
5'5"	12	.12
5'4"	7	.07
5'3"	5	.05
5'2"	0	0
5'1"	3	.03
5'0"	1	.01
4'11"	0	0
4'10"	2	.02
4'9"	0	0
4'8"	1	.01
4'7"	0	0
4'6"	0	0
4'5"	1	.01

Fig. 17. *Hypothetical Database Height Frequency Table.*

2.   RFLP

In the RFLP context, the scientists might randomly select 500 people (1,000 alleles). These people are tested using RFLP to measure the lengths of their two alleles at certain DNA loci. The 1,000 alleles at a locus are recorded together on a single table (a similar table of 1,000 alleles is created for each locus tested). A portion of the hypothetical database allele frequency table might look something like figure 18.

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

Fig. 18. *Hypothetical Database Allele Frequency Table.*

c. *Estimation of Allele Frequency*[44]

There are two ways to use these database frequency tables to estimate the frequency of a perpetrator's allele in the population -- the floating bin method and the fixed bin method. Both methods involve "bins," which are attached to the frequency table, grouping the allele measurements into sets. We will explain in detail.

1. Floating Bin Method

The floating bin method is the more accurate, statistically preferable method for estimating allele frequency. In the floating bin method, the frequency table has attached to it a single bin that "floats" up and down the table, taking a new position for each perpetrator's allele tested. (Fig. 19.)
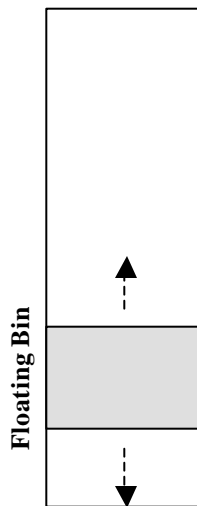
Fig. 19. *Floating Bin on Database Frequency Table.*
NOTE: The single floating bin slides up and down the frequency table to position itself around the perpetrator's allele.

---

44   Unless otherwise noted, in this section (part III.E.2.c.) we rely on NRCII, *supra*, at pages 7, 18-21, 142-145, 161-162 and 177.

Once the perpetrator's allele is measured from the autorad, that measurement is positioned next to the frequency table according to its size. The floating bin then slides into position around that measurement. (Fig. 20.)
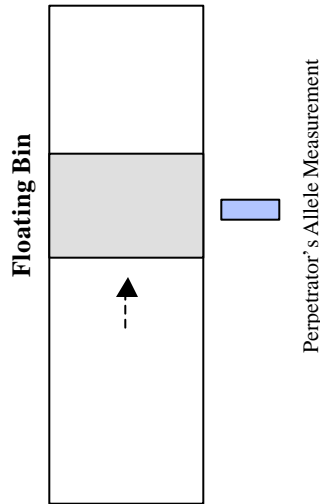
Fig. 20. *Floating Bin Slides into Place.*

The floating bin should be the same size as the match window because, like the match window, it must include all the allele measurements that are indistinguishable from and could be the same as the perpetrator's allele measurement. (Fig. 21.)
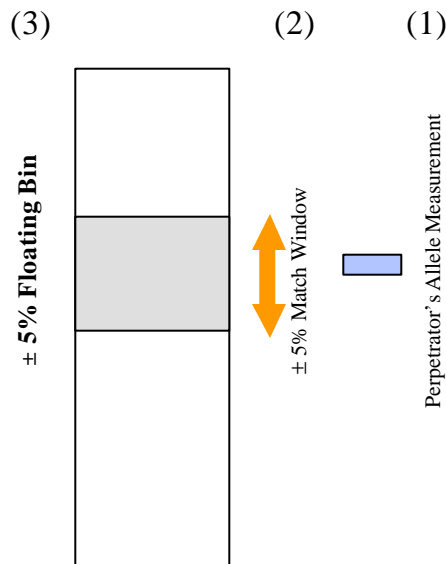
Fig. 21. *Match Window and Floating Bin.*
NOTE: The perpetrator's allele measurement (1) defines the center of the ±5% match window (2), which defines the ±5% floating bin (3).

Although the match window size dictates the minimum size of the floating bin, some laboratories choose to use a bin larger than the match window in order to make a more conservative estimate of allele frequency. In light of this practice, the match window does not technically define the floating bin. Thus, another window -- which we will call the "statistical window" -- determined by each laboratory actually defines bin size.[45]

---

[45] This window is called many things (in the briefs and elsewhere): the match window, the match window used to compare the evidence sample to the database, the window to compare with the bins in the database, the window used to go to the bins in the database, the window used to go to the frequency table, the window used to determine the frequency in the population, the match window to compare the evidence sample to the database, the bin assignment window, the bin frequency window, and the bin. Of course we cannot mandate usage, but we choose to use the generic term "statistical window" to avoid confusion and unwieldy descriptions such as these.

Although some may conclude the term "match window" is an appropriate name for this window, we note that the statistical window is sometimes a different size than the match window (e.g., some labs use a window larger than the match window in their use of the floating bin method; some labs, like the FBI in this case, use a window smaller than the match window in their use of the fixed bin method). If laboratories consistently used a statistical window the exact size of the match window, no term other than "match window" would be needed. Here, however, our use of "statistical window" is necessary to explain the differences between the two as applied by the FBI in this case (we recognize that where, in our discussion, the two are the same, the use of "statistical window" seems redundant).

Figure 22 shows an example of a statistical window larger than the match window.
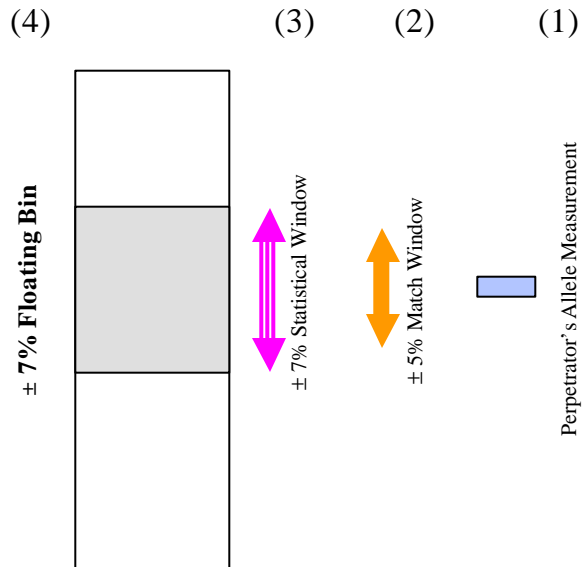
Fig. 22. *Statistical Window and Floating Bin.*
NOTE: The perpetrator's allele measurement (1) defines the ± 5% match window (2), which defines the minimum ± 5% size of the floating bin. Some laboratories, however, choose to use a larger statistical window (3) to define a larger floating bin (4).

In summary, the size and placement of the floating bin are determined as follows: the perpetrator's allele measurement determines the placement of the match window (± 5% around that measurement), the match window determines the minimum size of the statistical window (which some laboratories choose to enlarge), and the statistical window defines the floating bin.[46] Once the floating bin slides into place in line with the statistical window, it encompasses a group of allele measurements on the frequency table, each of which has a separate frequency. All the frequencies encompassed by the bin are added together to estimate the frequency of the perpetrator's allele -- or, more accurately, the frequency of all the allele measurements that could be the actual length of the perpetrator's allele. The following examples will illustrate.

---

[46] The various floating bin and fixed bin methods are named according to the size of statistical window used (e.g., the ± 5% floating bin method, the ± 5% fixed bin method, and the ± 2.5% fixed bin method).

a.   Height

In the height scenario, the ± 5% match window extends from 5 feet 4 inches to 6 feet 0 inches. These are the height measurements that match the 5-foot-8-inch measurement and define the minimum size of the statistical window. When a ± 5% statistical window is applied to the frequency table, the ± 5% floating bin slides into place. The bin, by definition, encompasses the frequencies of every height measurement matching Person 1's height measurement, shaded in figure 23. The frequencies within the floating bin are added together to account for all the measurements that match Person 1's height measurement of 5 feet 8 inches and *could actually be* Person 1's true height. Here, the total frequency of Person 1's height in the population is 0.83 (or 83%). (Fig. 23.)

Height	# of People	Freq. (#/100)
6'4"	1	.01
6'3"	0	0
6'2"	1	.01
6'1"	2	.02
6'0"	1	.01
5'11"	4	.04
5'10"	6	.06
5'9"	10	.10
5'8"	14	.14
5'7"	8	.08
5'6"	21	.21
5'5"	12	.12
5'4"	7	.07
5'3"	5	.05
5'2"	0	0
5'1"	3	.03
5'0"	1	.01
4'11"	0	0
4'10"	2	.02
4'9"	0	0
4'8"	1	.01
4'7"	0	0
4'6"	0	0
4'5"	1	.01

± 5% Floating Bin freq = .083

± 5% Statistical Window

± 5% Match Window

Others   **Person 1**

6'0"

5'8"

5'4"
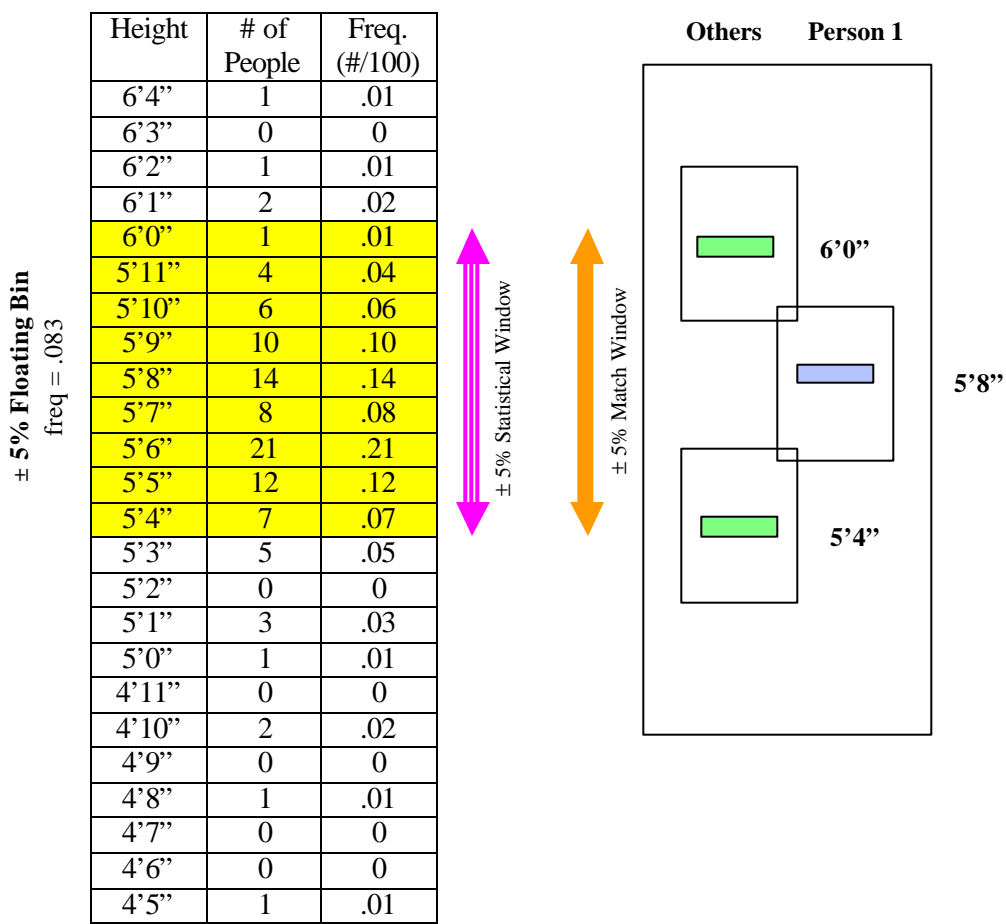
Fig. 23. *± 5% Floating Bin Method Applied to Database Height Frequency Table.*

In the RFLP example, the ± 5% match window extends from 950 bp to 1,050 bp. These are the allele measurements that match the perpetrator's 1,000 bp measurement. The ± 5% floating bin encompasses the frequencies of the matching alleles. The total frequency of the perpetrator's allele in the population is 0.079 (or 7.9%).[47] (Fig. 24.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

± 5% Floating Bin freq = .079

± 5% Statistical Window

± 5% Match Window

Others   Perp

1050

1000

950

Fig. 24. ± 5% Floating Bin Method Applied to Database Allele Frequency Table.

47	Realistically, actual tables are not created for use with the floating bin method. Instead, the range of the floating bin is entered into a computer that searches the database, then collects and adds together the allele frequencies within that range.

2. Fixed Bin Method

The second method of estimating allele frequency utilizes "fixed" bins. The fixed bin method attempts to approximate the floating bin method using a more convenient and accessible (no computer search is needed), but less accurate method. Whereas the floating bin method repositions the bin for each perpetrator's allele measurement, the fixed bin method uses prefabricated bins that are intended to mimic the effect of the floating bin. The frequency table does not have a single bin that slides into the exact position centered around the perpetrator's allele measurement; instead, the frequency table has several bins (up to 31) that are *already in position* -- they are predefined, preestablished, and fixed in place -- as though a statistical window has been applied to the frequency table repeatedly along its entire length. The fixed bins exist before a perpetrator's allele is measured, and their placement has nothing to do with that measurement.[48] (Fig. 25.)
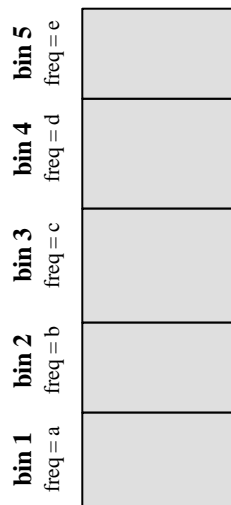
**Fig. 25.** *Fixed Bins on Database Frequency Table.*

---

[48]   The boundaries of the fixed bins are arbitrarily defined by the size markers on the autorad. Also, if a bin contains less than five alleles from the database, it is combined with a neighboring bin in order to avoid causing excessively rare frequencies.

Each fixed bin is intended to mimic a floating bin that would take a similar position on the table. The frequencies of the alleles encompassed by each fixed bin are added together to give a fixed bin frequency, just as they are for the floating bin. The fixed bin frequencies, however, are permanently assigned to the bins and are utilized for every allele tested at that locus in every case. Thus, while the floating bin method actually counts the matching alleles within the statistical window, the fixed bin method can only estimate the floating bin frequency by referring to counts already made that imitate the counts within the statistical window.

In the floating bin method, the perpetrator's allele measurement falls within the center of the floating bin (i.e., it defines the center), but in the fixed bin method, the perpetrator's allele measurement simply falls where it may among the preexisting fixed bins. (Fig. 26.)
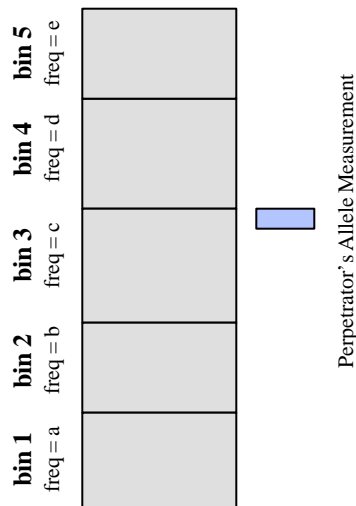
Fig. 26. *Fixed Bins Do Not Slide into Place.*
NOTE: The perpetrator's allele measurement falls where it may among the fixed bins.

In the floating bin method, the frequency of the bin centered around the perpetrator's allele measurement is assigned to that allele. In the fixed bin method, the frequency of one of the fixed bins -- expected to mimic the floating bin that would exactly surround the allele -- must be assigned to the allele. Of course, the fixed bins, which are arbitrarily positioned, are rarely in exactly the same position as the proper

floating bin would be, and therefore the fixed bin method can only estimate the correct frequency.

It may be tempting to assume that the frequency of whichever fixed bin the perpetrator's allele measurement falls into should be assigned to the allele (e.g., bin 3 in figure 26, *ante*), but that assumption fails to account for the *range* of matching allele measurements around the perpetrator's allele measurement. Thus, the *statistical window*, rather than just the allele, must be applied to the frequency table. When the statistical window (the size of which we discuss *post*) is applied to the frequency table, the window may fall within a single bin, in which case that bin's frequency is assigned to the allele, or the statistical window may overlap two or more bins, in which case the highest frequency of those overlapped bins is assigned to the allele. (Fig. 27.)
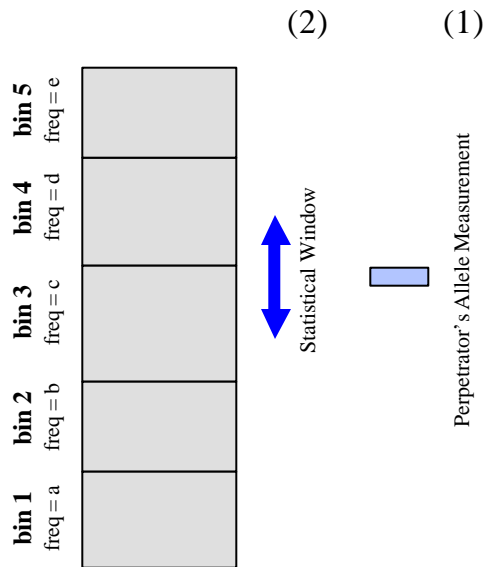
Fig. 27. *Statistical Window and Fixed Bins.*
NOTE: The perpetrator's allele measurement (1) defines the center of the statistical window (2), which may overlap more than one fixed bin.

a.   Statistical Windows

There is some controversy regarding the proper size and placement of the statistical window that should be applied to the fixed bin frequency table. (Indeed, this is an issue in this case.) We briefly discuss and diagram four types of statistical windows.

1.   ± 5% Statistical Window

The ± 5% statistical window is the same size and in the same position as the ± 5% match window. Figure 28 illustrates the unchanging nature of the ± 5% statistical window in three different situations (same perpetrator, three different defendants).
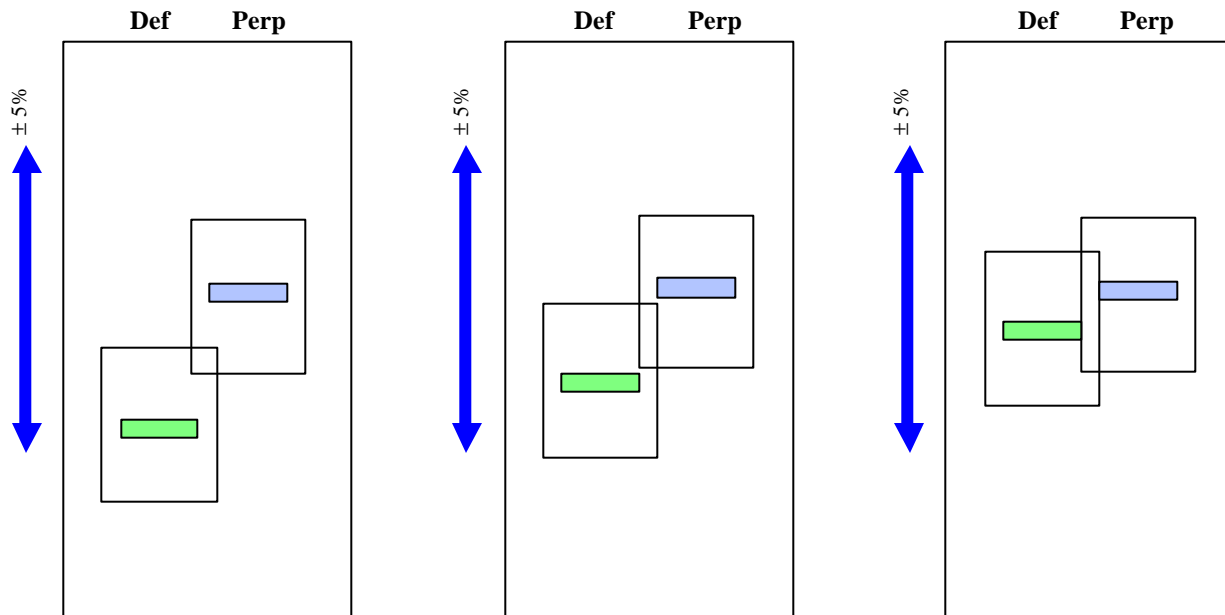
Fig. 28. *Three Examples of ± 5% Statistical Window.*
NOTE: The defendant's allele measurement has no effect on the ± 5% statistical window, which is centered on the perpetrator's allele measurement.

Like the match window, the ± 5% statistical window is centered on the perpetrator's allele measurement, without regard to the defendant's allele measurement.[49]

---

49   NRCII recommends use of this statistical window (the same one it recommends for use with the floating bin method) with the fixed bin method. (NRCII, *supra*, at p. 143 ["If the *match window* is entirely within a bin, the frequency used is that of the bin." (Italics added.)]; *id.* at p. 144 ["To approximate the floating-bin match probability, we recommend using the fixed bin with the largest frequency among those overlapped by the

A second method uses a statistical window created by the outline of the perpetrator's and defendant's overlapping uncertainty windows. Because the uncertainty windows are only $\pm 2.5\%$ wide, the overlap may be only about $\pm 2.5\%$ (half the size of the match window) or up to about the same size as the match window, depending on the closeness of the two bands. (Fig. 29.)
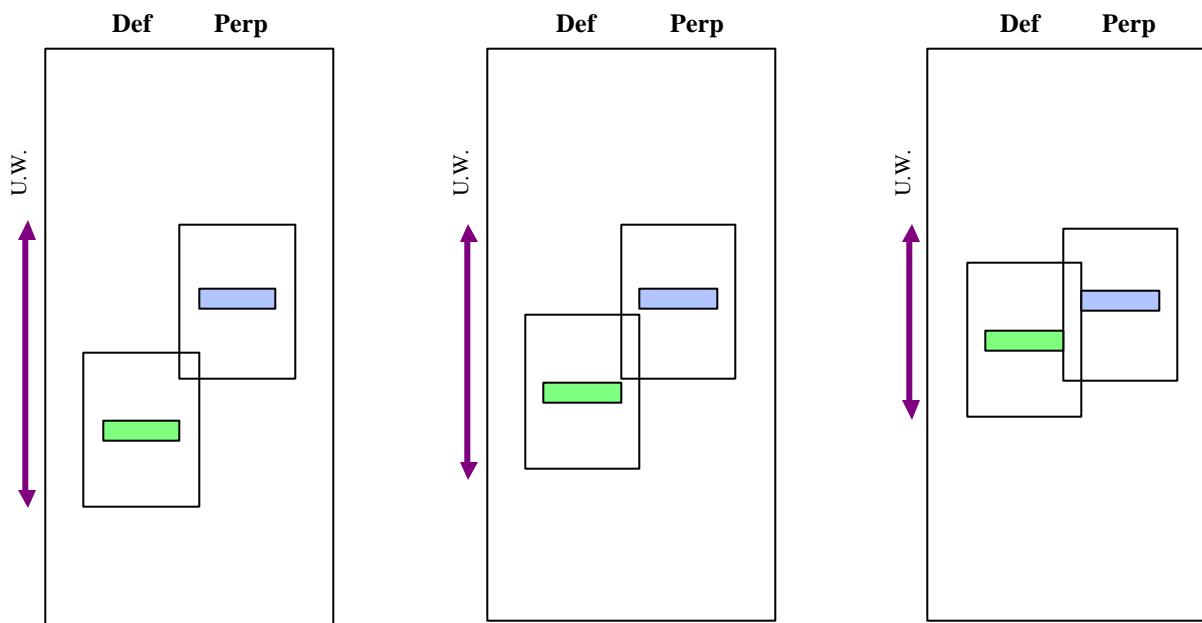
Fig. 29. *Three Examples of Overlapping-Uncertainty-Windows Statistical Window.*
NOTE: The statistical window based on the outline of the overlapping uncertainty windows varies in size depending on the distance between the perpetrator's and defendant's allele measurements. The closer the bands, the smaller the window.

This window is generally less conservative than the $\pm 5\%$ window. [50]

---

match window." (Italics added.)]; *id*. at p. 162 ["If fixed bins are employed, then the fixed bin that has the largest frequency among those overlapped by the *match window* should be used." (Italics added.)].) NRCII notes that when the $\pm 5\%$ statistical window is used, the fixed bin method provides results "very similar" to the floating bin method. (*Ibid*.)

[50] Although NRCII does not recommend this method, which it deems less conservative than the $\pm 5\%$ statistical window, NRCII nevertheless concludes it provides "adequate and usually conservative approximations to the correct floating-bin frequency." (NRCII, *supra*, at p. 144.)

Figure 30 compares these first two statistical windows in three situations:
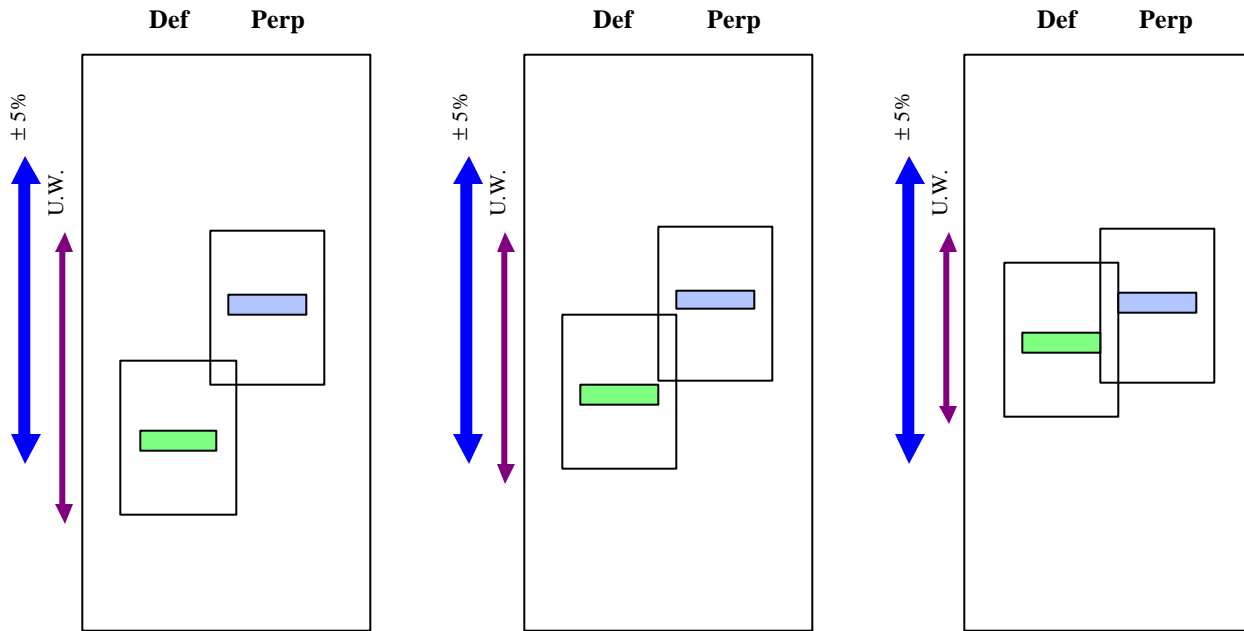
Fig. 30. *Comparison of Two Statistical Windows.*
NOTE: The heavy outside arrow (±5%) is a statistical window based on the ±5% match window. The finer inside arrow (U.W.) is a statistical window based on the outline of the two overlapping uncertainty windows. This window may be positioned differently than the ±5% statistical window, and may also be only half as large, depending on the closeness of the bands.

NRCII notes that this method is used by "[t]he FBI and many police agencies" (NRCII, *supra*, at p. 144), but this is *not* the method used by the FBI in this case.

3. ± 2.5% Statistical Window

A third method uses a ± 2.5% statistical window centered on the perpetrator's allele measurement. This method is the least conservative of the three because its statistical window is the smallest. This small window will overlap fewer bins than larger statistical windows. When fewer bins are overlapped, it is less likely that a bin with a higher frequency will be overlapped. (Fig. 31.)

Fig. 31. *Three Examples of ± 2.5% Statistical Window.*
NOTE: The defendant's allele measurement has no effect on the ± 2.5% statistical window, which is centered on the perpetrator's allele measurement.

Figure 32 compares these three statistical windows:
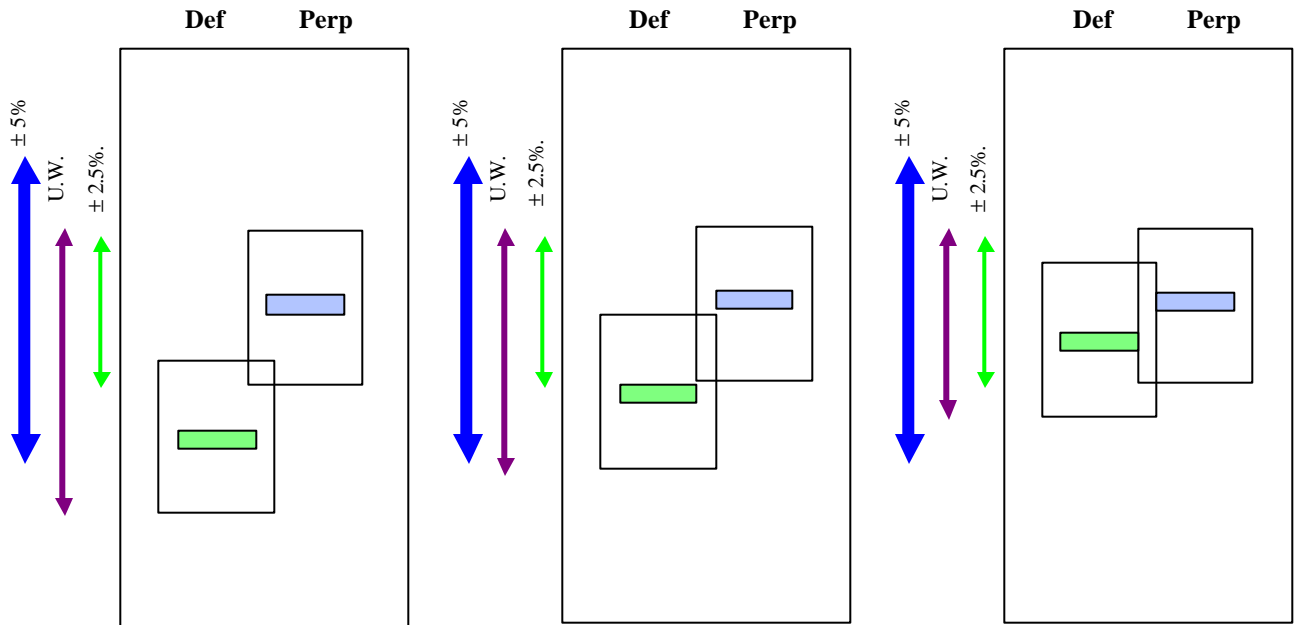
Fig. 32. *Comparison of Three Statistical Windows.*
NOTE: The heavy outside arrow (±5%) is a statistical window based on the ±5% match window. The finer middle arrow (U.W.) is a statistical window based on the outline of the two overlapping uncertainty windows. The finest inside arrow (±2.5%) is a ±2.5% statistical window.

4. ± 2.5% Average Statistical Window

A fourth type of statistical window is a ± 2.5% window centered not on the perpetrator's allele measurement, but on the average of the perpetrator's and defendant's allele measurements. This method is not mentioned by NRCII, but it may have been the method used by the FBI in this case and therefore we add it to our discussion.[51] The size of the window increases very slightly and its position shifts upward as the defendant's allele measurement increases.[52] (Fig. 33.)
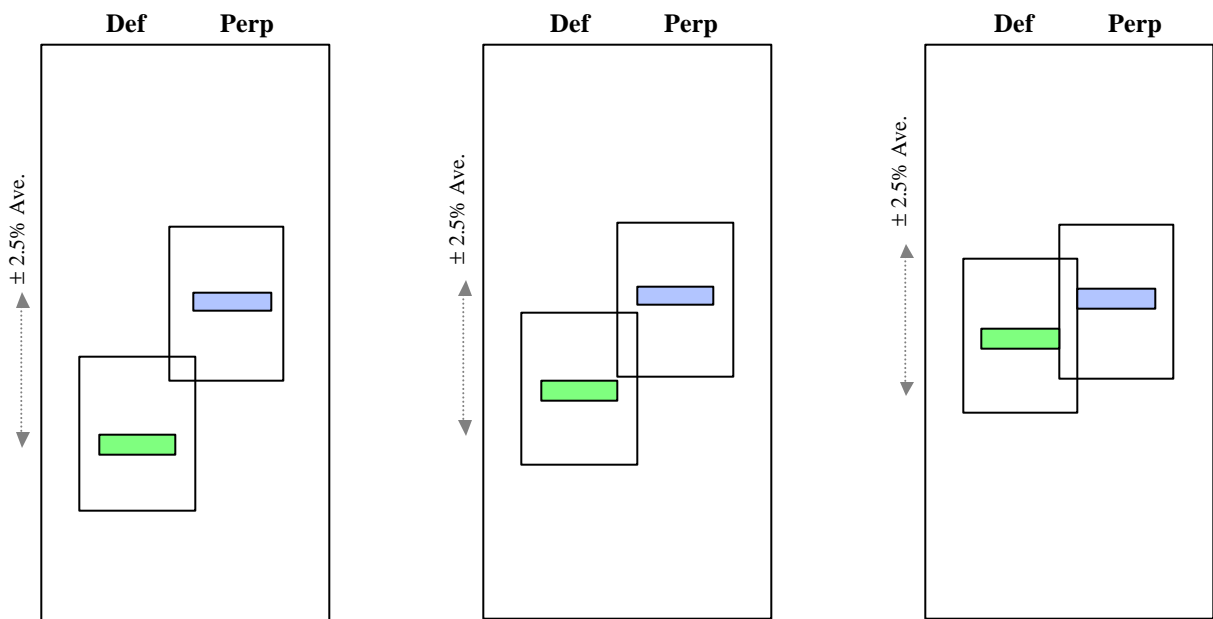
Fig. 33. *Three Examples of ± 2.5% Average Statistical Window.*
NOTE: The statistical window centered on the average of the perpetrator's and defendant's allele measurements does not vary much in size, but varies in position depending on the defendant's allele measurement.

---

51 There was conflicting testimony as to what type of statistical window the FBI used in this case. (See part VIII.D., *post*.)

52 For example, when the perpetrator's allele measurement is 1,000 bp and the defendant's allele measurement is 960 bp, this window ranges from about 956 bp to 1,005 bp (a range of about 49 bp). When the perpetrator's allele measurement is 1,000 bp and the defendant's allele measurement is 990 bp, this window shifts up and ranges from about 970 bp to 1,020 bp (a range of about 50 bp).

Figure 34 compares all four statistical windows:
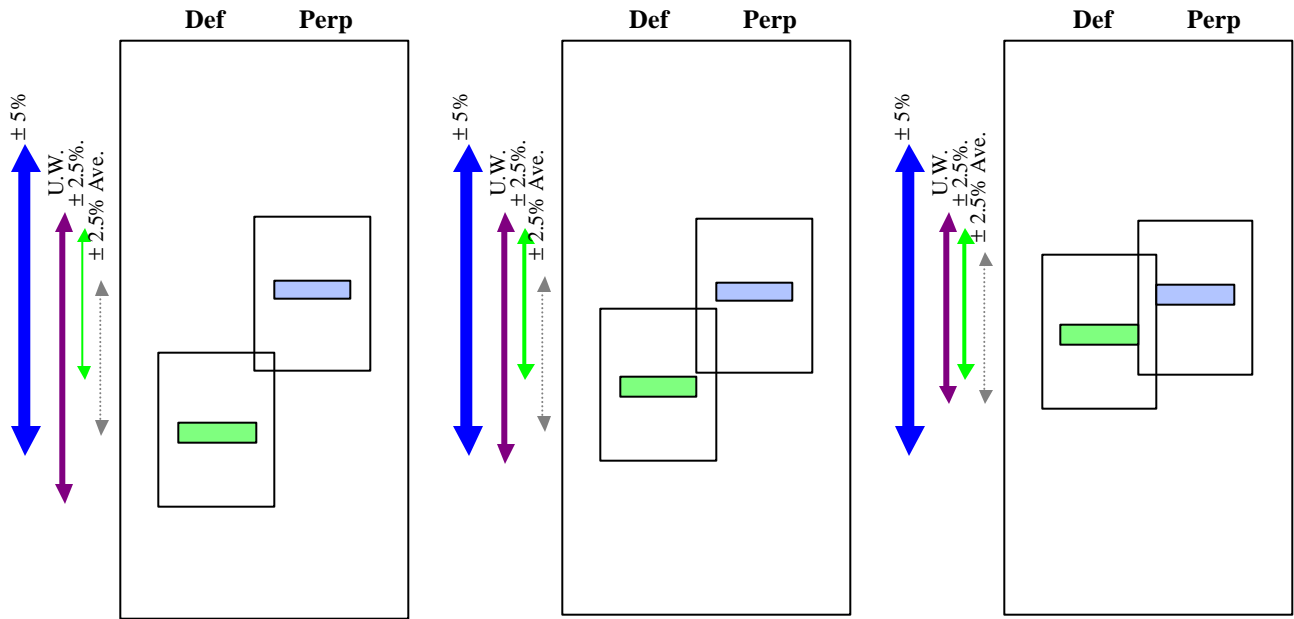
Fig. 34. *Comparison of Four Statistical Windows.*
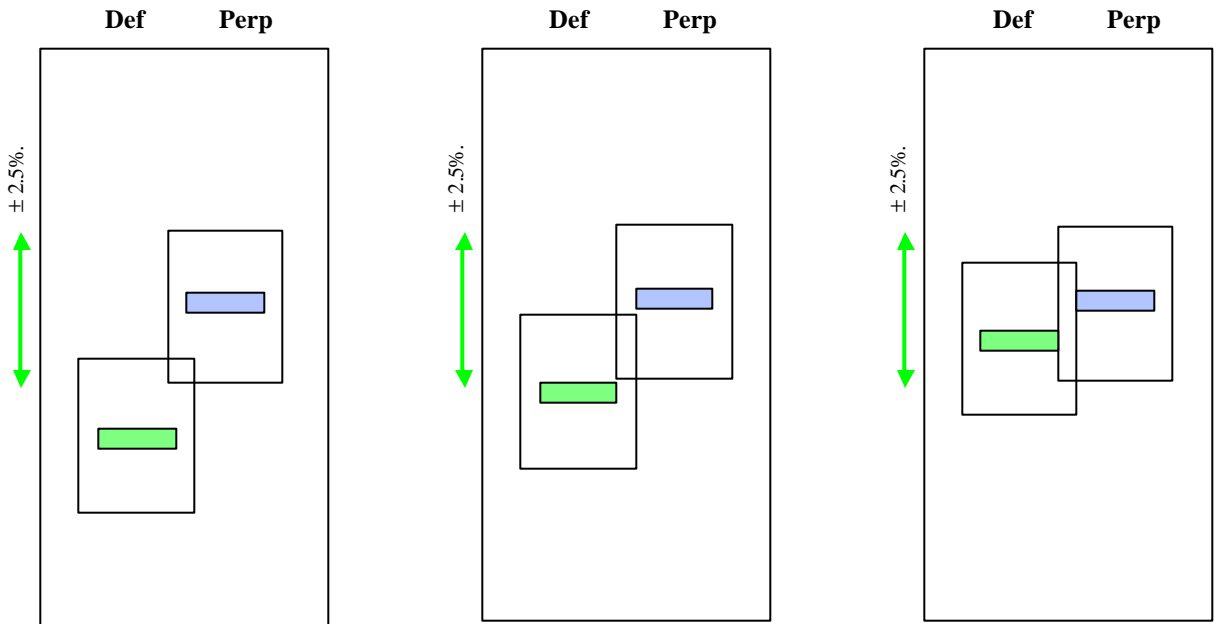NOTE: The heavy outside arrow (±5%) is a statistical window based on the ±5% match window.
The finer arrow (U.W.) is a statistical window based on the outline of the two overlapping uncertainty windows. The finest arrow (±2.5%) is a ±2.5% statistical window. The dashed inner arrow (±2.5% Ave.) is a ±2.5% statistical window centered on the average of the perpetrator's and defendant's allele measurements.

<center>*b.*   *Height*</center>

To demonstrate use of the fixed bin method here, we apply only the ± 5% statistical window. In the height scenario, the frequency table might be divided into three bins of roughly equal size. When the ± 5% statistical window is applied to the table, it overlaps two of the preestablished fixed bins on the table. Thus, the bin with the larger frequency is assigned to Person 1's height measurement. Here, the frequency of matching height measurements is 0.77 (or 77%) (Fig. 35.)

	Height	# of People	Freq. (#/100)
bin 3 freq = .15	6'4"	1	.01
	6'3"	0	0
	6'2"	1	.01
	6'1"	2	.02
	6'0"	1	.01
	5'11"	4	.04
	5'10"	6	.06
bin 2 *freq = .77*	5'9"	10	.10
	5'8"	14	.14
	5'7"	8	.08
	5'6"	21	.21
	5'5"	12	.12
	5'4"	7	.07
	5'3"	5	.05
	5'2"	0	0
bin 1 freq = .08	5'1"	3	.03
	5'0"	1	.01
	4'11"	0	0
	4'10"	2	.02
	4'9"	0	0
	4'8"	1	.01
	4'7"	0	0
	4'6"	0	0
	4'5"	1	.01

± 5% Statistical Window **freq = .77**

Others   Person 1   6'0"   5'8"   5'4"
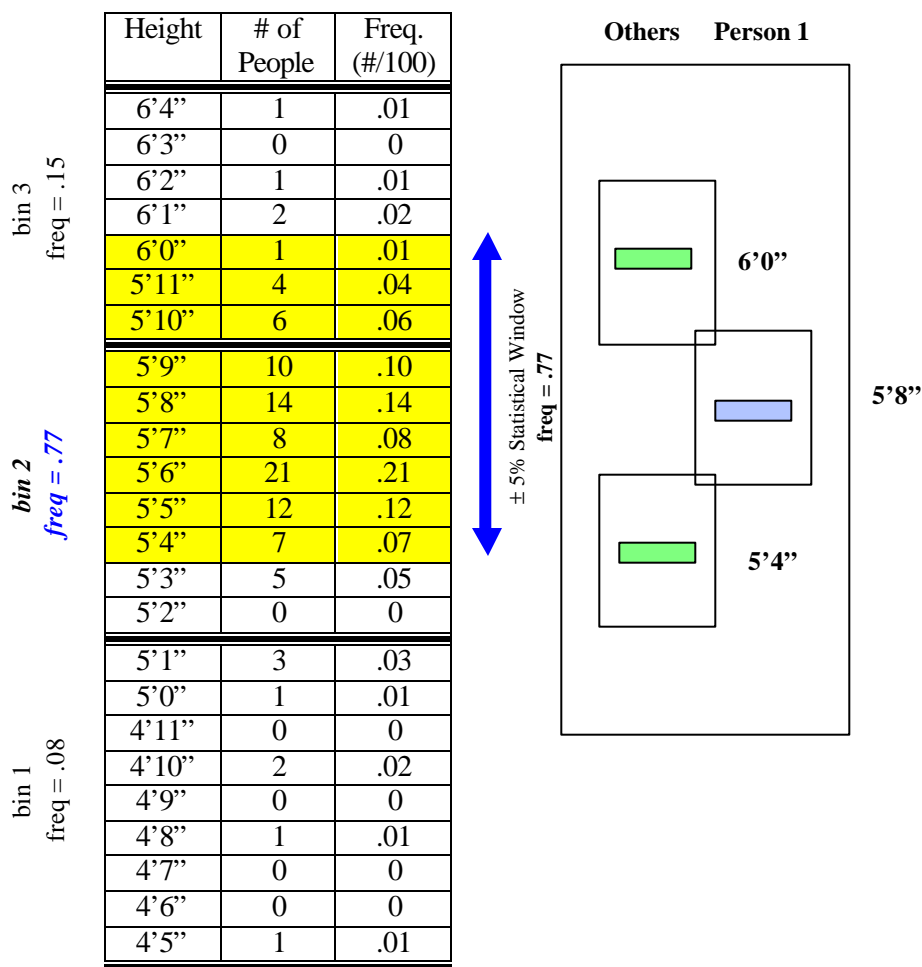
Fig. 35. *± 5% Fixed Bin Method Applied to Database Height Frequency Table.*
NOTE: The statistical window overlaps two fixed bins; the higher frequency (0.77) is assigned to the allele.

If, however, the statistical window falls entirely within a single bin, that bin's frequency is assigned to the perpetrator's allele.

c.   *RFLP*

In the RFLP example, the frequency table is divided into many bins, three of which are visible in our diagram. Now, the ± 5% statistical window overlaps three bins and the highest frequency, 0.085 (or 8.5%), is assigned to the perpetrator's allele. (Fig. 36.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
20002009	19	.019
1090-1099	6	.006
1080-1089	0	0
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

Fig. 36. *± 5% Fixed Bin Method Applied to Database Allele Frequency Table.*
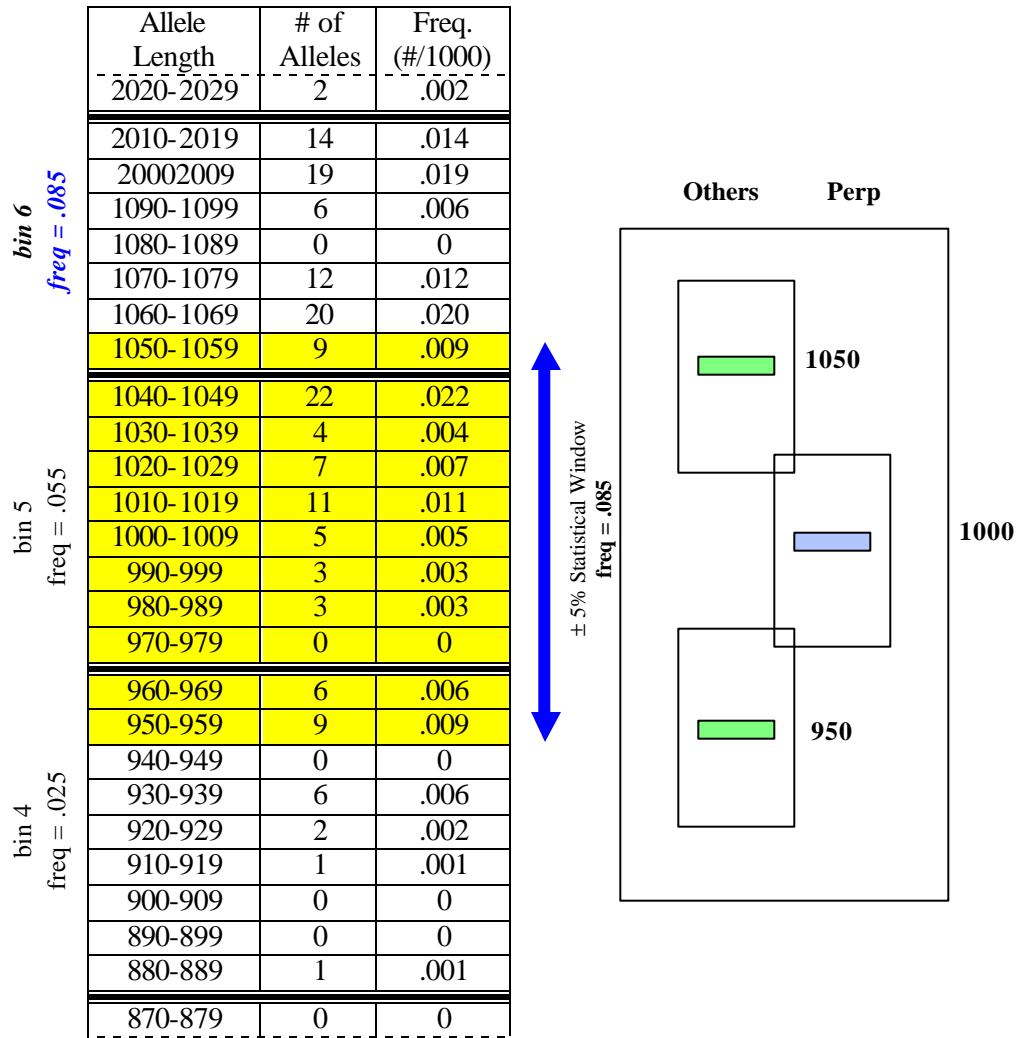NOTE: The statistical window overlaps three fixed bins; the highest frequency (0.085) is assigned to the allele.

75.

d.   Calculation of Genotype Frequency & Overall Profile Frequency[53]

The two allele frequencies at a locus are first calculated together to obtain a genotype frequency for that locus. Then, the frequency or probability of the perpetrator's overall genetic profile is calculated from all the genotype frequencies. The overall "numerical probability is generally calculated using the 'product rule,' which posits that the probability of several things occurring together is the product of their separate probabilities. [Citation.] For example, the probability of 'heads' coming up on three successive coin tosses is the probability of heads on the first toss (1 in 2), multiplied by the probability of heads on the second toss (1 in 2), multiplied by the probability of heads on the third toss (1 in 2), resulting in an overall probability of 1 in 8.[54] Similarly, if a set of paired alleles (a genotype) is known to occur in 1 in 3.47 people and another set of paired alleles is known to occur in 1 in 18.52 people, then the probability of *both* sets occurring in the *same* person is 1/3.47 multiplied by 1/18.52, or 1 in 64.26 people. When more alleles are examined, the probability of a multilocus profile can be exceedingly rare, even one in hundreds of billions, and therefore the profile is highly distinctive.[55]" (*People v Brown, supra,* 91 Cal.App.4th at p. 630.)

---

[53]   For further explanation of this step, see NRCII, *supra*, at page 45; and NRCI, *supra*, at pages 77-79.

[54]   "Probabilities are also often represented in decimal form."

[55]   "Obviously, there are situations in which the result of the product rule calculation exceeds the size of the particular population on earth. In that case, the result must be viewed in its alternative sense -- the numerical probability that a person randomly chosen from that population will possess the same genetic profile."

IV.   AUTORADIOGRAPHS

In Pizarro's case, three autorads were used to create the DNA profiles and to determine a match between the perpetrator's and defendant's profiles. Because our discussion refers extensively to this evidence, we include scans of the D1S7 (hereafter D1), D2S44 (hereafter D2), and D4S139 (hereafter D4) autorads.[56] Recall that each autorad is made from the same gel and hybridization membrane and thus the same underlying DNA fragments. The autorads look different because each memorializes a hybridization with a different probe that attached to and "lit up" different DNA fragments on the membrane.

Each autorad displays 12 vertical lanes. Four lanes contain size standards to which the unknown fragments can be compared and sized. The standards are run in several lanes on the gel to account for slight variations in electrical current in different regions of the gel. (Note that corresponding bands across the width of the autorads are not in perfect alignment.) One lane contains a control sample to ensure that there has been no obvious failure in the system. The remaining lanes contain the DNA samples specific to this case. In summary, the autorads display the following samples:

lanes 1, 5, 9 & 12	Size Stds:	size standards
lane 2	C:	control sample
lane 3	V:	victim's reference blood sample
lane 4	Def:	defendant's reference blood sample
lanes 6 & 7	V(ev):	victim's vaginal epithelium fraction of evidentiary sample[57]
lanes 10 & 11	Perp:	perpetrator's sperm fraction of evidentiary sample

---

[56]   The D17S79 (hereafter D17) autorad was inconclusive.

[57]   The evidentiary samples from two vaginal swabs were divided into sperm and vaginal cell DNA fractions, resulting in two sets of evidentiary DNA samples (4 lanes).
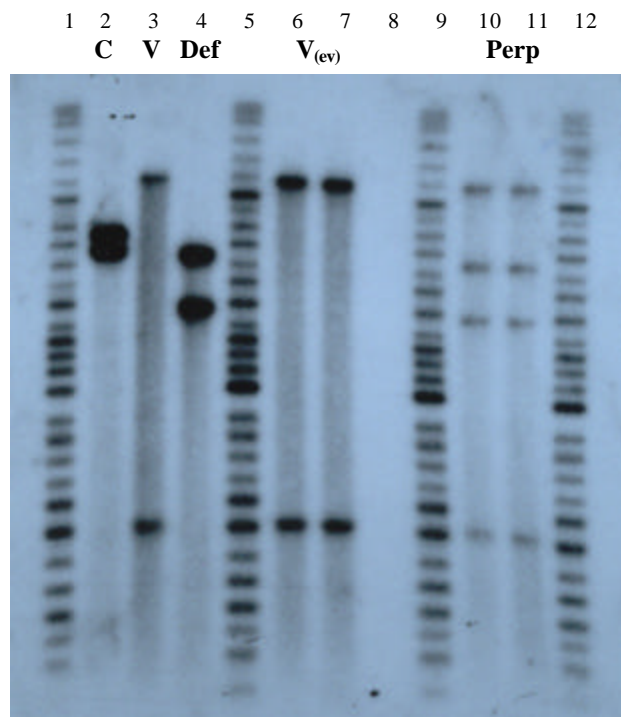
Fig. 37. *D1 Autorad.*
NOTE: C = control; V = victim; Def = defendant;
$V_{(ev)}$ = victim's (vaginal epithelium) fraction of
evidentiary sample; Perp = perpetrator's (sperm)
fraction of evidentiary sample. (Exhibit Y)
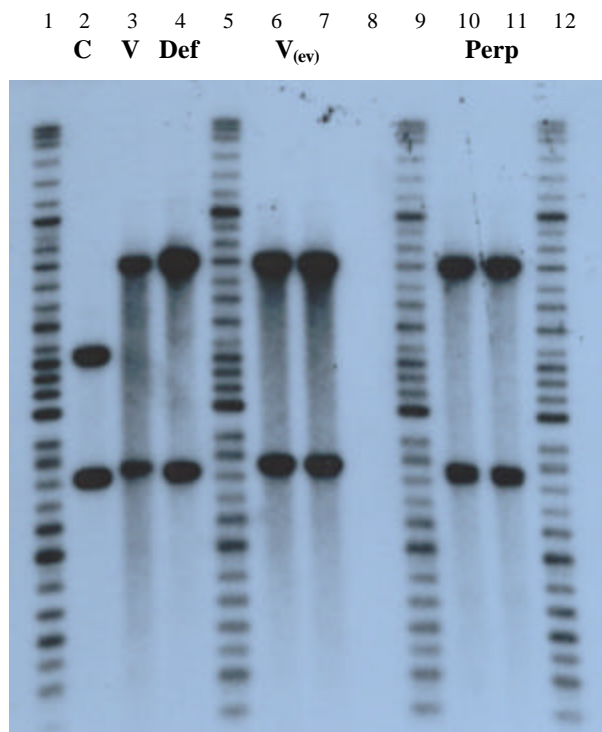
Fig. 38. *D2 Autorad.*
NOTE: C = control; V = victim; Def = defendant;
V(ev) = victim's (vaginal epithelium) fraction of
evidentiary sample; Perp = perpetrator's (sperm)
fraction of evidentiary sample. (Exhibit 2)

Fig. 39. *D4 Autorad.*
NOTE: C = control; V = victim; Def = defendant;
V(ev) = victim's (vaginal epithelium) fraction of
evidentiary sample; Perp = perpetrator's (sperm)
fraction of evidentiary sample. (Exhibit Am)

V.   WITNESSES

A.   *PROSECUTION WITNESSES*

1.   Sensabaugh

George Sensabaugh, Jr., was a professor in the School of Public Health at the University of California, Berkeley. He had been on the faculty at t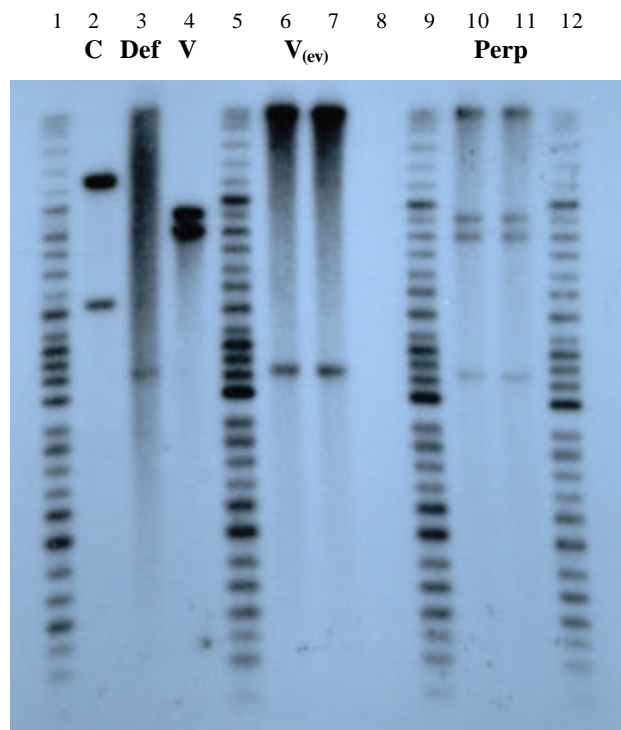hat institution since 1972. He taught courses in forensic science and infectious disease. In addition, his laboratory conducted research on certain forensic issues. The courses he taught included RFLP analysis, but his research emphasized the polymerase chain reaction (PCR). Prior to joining the faculty at Berkeley, he received a bachelor's degree from Princeton University in 1963 and a doctorate in criminology from Berkeley in 1969. He conducted post-doctoral research in chemistry and genetics at University of California at San Diego and the National Institute for Medical Research in London. He was on the editorial board of several forensic journals, and had authored over 130 publications, about half of which concerned DNA forensic issues. Sensabaugh was familiar with the procedures employed by the FBI and Cellmark laboratories. The court found Sensabaugh qualified as an expert in forensic DNA technology and human population genetics.

2.   Chakraborty

Ranajit Chakraborty was a professor of population genetics, biometry, and international health at the University of Texas Health Science Center at Houston. He taught courses in statistical genetics, epidemiology, population genetics, and population biology. His research focused on the application of population genetics to the study of diseases. He had conducted RFLP studies in a research setting since 1984. Before he joined the faculty at Houston, he received his bachelor's degree in statistics in 1967 and his master's degree in mathematical genetics in 1968, both from Indian Statistical Institute in Calcutta. In 1971, he received his doctorate in biostatistics and population genetics. He performed postdoctoral research in 1971 through 1973. In addition, he was a member of the editorial boards of American Journal of Human Genetics, Journal of

Physical Anthropologists, Human Biology, and others. He had published some 268 articles and over 100 book chapters and commentaries.

Chakraborty was familiar with the specific methodologies used by the FBI, Cellmark, and Lifecodes laboratories, and he had reviewed the FBI protocol. Chakraborty collaborated and worked closely with the FBI. He advised the FBI on population genetic issues and analyzed various databases. As a condition of his research, he had an arrangement with the FBI under which he had access to information regarding particular cases and could receive any data he might need. About 50 percent of his research funding was in the form of grants from the National Institute of Justice, which like the FBI was an arm of the Justice Department. Chakraborty had reviewed the FBI's response to the NRCI 1992 report at the request of Bruce Budowle, the head of the FBI's research laboratory, Quanitco.

The court found Chakraborty qualified as an expert in RFLP and population genetics.

3.   Adams

Dwight Adams was the FBI scientist who oversaw the DNA analysis in Pizarro's case. Although his technician performed the laboratory work,[58] Adams determined what samples to analyze, how much DNA to place on the gels, and whether samples needed to be re-analyzed. He followed cases to ensure they were analyzed to his satisfaction. Adams evaluated and made the final determinations on Pizarro's autorads and generated the final report.

Adams had been assigned to the FBI laboratory in 1987, and was present at the inception of its DNA analysis work. He was first assigned to the research unit that tested

---

[58]   As we mention *post*, Adams testified in 1990 that he had personally performed the laboratory work in Pizarro's case.

and validated the methods that were currently used in DNA casework. After completion of those studies in 1988, he was assigned to the DNA analysis unit that conducted actual casework. In 1993, he became chief of the DNA analysis unit of the FBI laboratory in Washington D.C., a post he held for 13 months. Shortly before the hearing, Adams had become a supervisory special agent for the FBI.

Adams held a bachelor's degree in biology from Central State University in Oklahoma, a master's degree in biology from Illinois State University, and a doctorate in biology from the University of Oklahoma. His education included courses in DNA analysis, in addition to laboratory work. He estimated he had performed RFLP in over 1,000 cases. He had authored about 12 publications on DNA analysis, three of which were book chapters. Adams had not, however, performed any type of DNA analysis until he began working for the FBI, and his education included very little emphasis on either population genetics or molecular biology.

During the time Adams was a member of the FBI's DNA unit, he and the other scientists there were required to undergo proficiency testing four times per year. In all the proficiency tests Adams had taken, he had never made an incorrect match or an incorrect nonmatch.

4.   Conneally

Patrick Michael Conneally was a professor of medical genetics at Indiana University Medical Center in Indianapolis. His major work involved genetic mapping of human diseases in the field of human population genetics. He taught courses in human population genetics and human genetics. Conneally received his bachelor's degree in science from Universal College in Dublin, Ireland, and his master's and doctoral degrees in genetics, human genetics, and statistics from the University of Wisconsin in Madison. He served on the editorial board of several journals.

Scientists in Conneally's research laboratory used both RFLP and PCR, although Conneally himself had never performed a complete RFLP procedure. Currently, his

laboratory's DNA analysis consisted of approximately 90 percent PCR and 10 percent RFLP. Conneally noted that he was not a molecular geneticist and had reviewed RFLP procedures only casually and not in great depth. He had, however, examined thousands of RFLP autorads, including hundreds in forensic cases. He did not consider himself an expert in the nuances of producing autorads; his expertise lay in the statistics of population genetics.

B.   DEFENSE WITNESSES

1.   Shields

Williams Shields was a professor of biology at the State University of New York in the College of Environmental Science and Forestry. He taught courses in evolutionary and systematic biology in genetics and conservation genetics. His research involved both behavioral ecology and population genetics, including the effects between population structure and evolution. In the previous four years, Shields had become involved in research in the forensic aspect of population genetics, which now occupied about 50 or 60 percent of his research. His work included examining and analyzing population databases used in forensics; he had examined 300 or 400 databases, including the FBI's. He also conducted molecular genetic research in his laboratory. He had reviewed thousands of RFLP autorads in forensic cases, and had also reviewed about 25 laboratory protocols.

Shields held a bachelor's degree in biology from Rutgers University, a master's degree from Ohio State University, and a doctorate in zoology from Ohio State University. He had taken 60 or 70 courses in various aspects of statistical analysis and probability theory.

Shields had published approximately 40 articles and one book, about half on the topic of population genetics. None of his articles was on the subject of human RFLP genetic variation issues. He worked as a reviewer for several journals and federal grant programs.

Shields considered himself an expert in molecular genetics and statistics, but not in molecular biology. The court ruled he was qualified as an expert.

2.   Zabell

Sandy Lew Zabell had been a professor of mathematics and statistics at Northwestern University since 1980. He received a bachelor's degree in mathematics from Columbia University, a master's degree in biochemistry and molecular biology and a doctorate in mathematics from Harvard University. Before moving to Northwestern University, he served as assistant professor of statistics at the University of Chicago, and visiting assistant professor at Rutgers University and University of California, Berkeley. Zabell had been asked to join the committee created to determine whether a second NRC report (NRCII) was needed. He personally had never performed RFLP or PCR.

The court found Zabell an expert in the field of statistics, specifically with regard to calculations and methodology.

3.   Bakken

Aimee Hayes Bakken was an associate professor of zoology at the University of Washington. In her research laboratory, she conducted research in developmental genetics, cell biology, and molecular biology. In addition to her university duties, she taught summer courses in molecular cytogenetics at Cold Spring Harbor Laboratory. She spent a six-month sabbatical at Edinburgh University with Edmond Southern, the originator of Southern blotting, a critical component of RFLP. She spent another six months conducting genetic engineering research in gene regulation at Fred Hutchinson Cancer Research Center. She also studied molecular transport through the nuclear membrane at the Max Plank Institute in Berlin.

Bakken had performed RFLP herself hundreds of times, beginning in about 1981, long before the FBI used RFLP in forensic applications. Her laboratory research, however, did not involve forensic work, but she had nevertheless reviewed approximately 2,000 autorads produced in forensic cases.

Bakken received her bachelor's degree in biology at the University of Chicago, after which she worked as a laboratory technician performing human chromosome analysis and in vitro fertilization research. She received her doctorate in developmental genetics at the University of Washington, then conducted post-doctoral research at Oak Ridge National Laboratory, where she took the first electron microscope photograph of human genes, and also worked at Yale University. She had been involved in DNA research since 1961. In 1973, she joined the faculty at University of Washington.

Bakken had worked as a reviewer for several journals and reviewed research grant applications for the National Science Foundation. Most of Bakken's 48 publications related to aspects of DNA. She was invited to appear before the NRC's update committee.

The court found Bakken an expert in the field of molecular biology.

4. Muller

Lawrence Muller was an associate professor of population genetics at the University of California at Irvine. Prior to that, he was an associate professor at Washington State University. In his laboratory, he conducted research on the genetics and physiology of aging, and the genetic stability and evolution of populations. Muller first became interested in the forensic application of population genetics in 1989. He had studied 30 or 40 different databases from various laboratories, including the FBI's. He did not perform RFLP in his laboratory, had never personally performed RFLP, and did not consider himself a forensic scientist. His area of expertise lay in the steps following production of an autorad -- analyzing the rules for declaring a match and the statistical implications of those rules on the final frequency estimate.

Muller held a bachelor's degree in science and chemistry and a master's degree in biology from Stanford University, and a doctorate in ecology from University of California at Davis. He conducted postdoctoral research on population genetics at Stanford University.

Muller had published 69 articles and book chapters, including two dealing specifically with the forensic application of DNA technology. He had reviewed several journals and had been invited to speak to various groups, including the NRC committee.

The court found Muller qualified as an expert in population genetics.

VI.   RELEVANT DATABASE

Defendant first contends correct scientific procedures were not followed and the requirements of the Evidence Code were not satisfied when the jury was informed that the DNA profile frequency applicable to Pizarro's case was the probability of finding a matching profile in the *Hispanic* population, although there was insufficient evidence that the perpetrator is Hispanic.

The People assert this contention must be rejected in light of the conservative nature of the Hispanic database and the fact that frequencies do not vary greatly by ethnicity. The People argue the error is harmless because the profile frequency from the Hispanic database was more common and thus more favorable to defendant than the profile frequencies calculated from other databases. Defendant maintains, however, that the error cannot be harmless because presentation of the Hispanic frequency itself -- regardless of the favorableness of the number -- and the manner in which the evidence was presented led the jury to believe that the perpetrator is Hispanic, even though no independent evidence justified the drawing of such an inference.

We conclude the trial court erred in determining there was sufficient foundational evidence that the perpetrator is Hispanic. Absent proof of this preliminary fact, the profile frequency based on the Hispanic database was neither relevant nor admissible.

A.   TRIAL TESTIMONY

At trial, there was evidence that the victim was last seen as she approached the area where defendant, who is half Hispanic, had been not long before. This was the extent of the evidence offered to establish that the perpetrator is Hispanic (or half Hispanic).

Adams, who conducted the scientific work in Pizarro's case in 1989, was the sole scientific witness at trial. He testified that "[t]he likelihood of finding another unrelated Hispanic individual" with a profile similar to the perpetrator's and defendant's profiles is approximately 1 in 250,000. His 1990 testimony follows:

"[PROSECUTOR:] What is your opinion as to the chances of another Hispanic male having the same DNA profile as Mr. Pizarro?

"[ADAMS:] *The likelihood of finding another unrelated Hispanic individual with a similar profile as Mr. Pizarro is one in approximately 250,000.*

"[PROSECUTOR:] *And this would also be the same statistic for the probability of a match of a DNA profile between the [perpetrator's DNA] obtained from the vaginal swab?*

"[ADAMS:] *That is correct.*

"[PROSECUTOR:] Same statistic?

"[ADAMS:] Yes.

"[PROSECUTOR:] And, again, this is only with Hispanic men?

"[ADAMS:] Hispanics, not broken down into gender. [¶] … [¶]

"[PROSECUTOR:] Dr. Adams, we have been talking about the chance for a match within the Hispanic community. Would the statistics for a match within the Caucasian community be different?

"[ADAMS:] Yes, generally there are going to be some differences in the population data from the different populations. So that's why we keep them separate. That's why we have a Caucasian and a Black and a Hispanic, American Indian population because there are differences. [¶] So if I were to compare one person in each of those different populations I would come up -- I'm sure I would come up with somewhat different results because in one population that pattern may be very rare, and another population that same pattern may be very common.

"[PROSECUTOR:] Have you done any of the calculations necessary to determine what the chances are of having matches of this particular DNA profile within the Caucasian community?

"[ADAMS:] Yes.

"[PROSECUTOR:] And what are those statistics?

"[ADAMS:] The statistics in those cases -- in that case comparing the same profile to the Caucasians is much greater. It would be one in 10,000,000.

"[PROSECUTOR:] But within the Hispanic group alone it is according to your testimony one in 250,000?

"[ADAMS:] Yes, ma'am.

"[PROSECUTOR:] *What about a situation where someone is half Hispanic and half Caucasian?*

"[ADAMS:] *Well, there is nothing we can do other than to compare them to the two populations and we would use only the smaller of the two in our report.* [Adams referred to the number with the smaller *denominator*.]

"[PROSECUTOR:] Why do you use only the smaller of the two?

"[ADAMS:] We attempt to be as conservative as possible. The smaller number is less detrimental to the defendant." (Italics added.)

B. *PIZARRO I OPINION*

In *Pizarro I*, to guide the trial court on remand, we explained that admission of evidence of the perpetrator's DNA profile frequency derived from the Hispanic database would require the trial court's determination of a preliminary fact -- that the perpetrator is Hispanic. Otherwise, the Hispanic frequency would not be relevant to show defendant is the perpetrator. We explained in *Pizarro I*:

"In *People v. Axell*, the unknown assailant left strands of hair at the crime scene.

"'July 28, 1988, Cellmark Diagnostics, a testing laboratory in Germantown, Maryland, received from the district attorney's investigator, whole bloodstains on cotton from the victim and appellant, and roots from 15 hairs recovered from the crime scene. The DNA was extracted from these materials, and Cellmark reported that the banding patterns obtained from the appellant's whole bloodstain matched the DNA banding patterns obtained from the 15 hair roots found at the scene of the murder. Subsequently, Cellmark reported that the frequency of that DNA banding pattern in the Hispanic population is approximately 1 in 6 billion. Appellant is part Hispanic. Simply put, Cellmark's analysis meant that the

chance that anyone else but appellant left the unknown hairs at the scene of the crime is 6 billion to 1.' ([*Axell, supra,*] 235 Cal.App.3d 836, 844 ….)

"This statement reveals the problem in the instant case. The selected racial or ethnic data base is predicated on the *suspect's* racial or ethnic background. However, the relevancy of the statistical probability depends on the *perpetrator* being the same racial or ethnic background as the suspect. In other words, examining the defendant's DNA banding pattern and concluding that it has an expected frequency of occurrence of, for example, 1 in 500,000 in a specific racial/ethnic data base would reflect the probability that the suspect committed the crime only if the perpetrator was within that same data base. It is clear that all population groups share common allele patterns according to the theory advanced by the FBI -- it is the frequency with which these patterns appear within different groups which will vary. Nothing in the record supports the conclusion that the banding patterns are race or ethnic specific so that a review of the banding pattern would conclusively establish that the person who left the sample was of a particular racial or ethnic background. Dr. Adams did not testify and, as we understand the evidence, could not testify that the perpetrator in the instant case was Hispanic based solely upon the allele pattern found in the evidence which was left at the crime scene by the perpetrator. What if the perpetrator was/were Black or non-Hispanic Caucasian, etc., and what is the relevancy of the estimated probabilities for these groups if we do not know the race or ethnic background of the perpetrator? It is a bootstrap argument to assume relevancy of a Black or Hispanic data base simply because the *suspect* falls within that racial or ethnic group. [¶] … [¶]

"Proffered evidence as utilized in section 403 'means evidence, the admissibility or inadmissibility of which is dependent upon the existence or nonexistence of a preliminary fact.' (Evid. Code, § 401.) Here the proffered evidence is the result of statistical analysis which utilizes ratios assigned to particular racial or ethnic databases. '"Relevant evidence" means evidence, including evidence relevant to the credibility of a witness or hearsay declarant, having any tendency in reason to prove or disprove any disputed fact that is of consequence to the determination of the action.' (Evid. Code, § 210.)

"The disputed fact generally is whether the suspect is also the perpetrator. Thus, the evidence is relevant if it tends to prove the suspect is the perpetrator. However, the preliminary fact upon which the relevancy of the proffered evidence depends is the racial/ethnic background of the *perpetrator*, not the suspect. If the only way you can conclude the perpetrator fits a racial/ethnic category is to assume the perpetrator was the same race/ethnic background as the suspect then the reasoning is circular,

90.

i.e.: proof of the racial/ethnic background of the perpetrator depends on the racial/ethnic background of the suspect from which we infer a statistical probability that the perpetrator is the suspect. Absent proof sufficient under Evidence Code section 403 to support the preliminary fact as to the racial/ethnic background of the perpetrator, we see no relevancy to a data base selected because of the racial/ethnic background of the suspect/defendant. The problems created by employing assumed relevancy of the data base are insidious. A jury hears an astronomical figure that not uncommonly depends for its relevance upon the very issue that they have to decide: is the defendant the perpetrator? The same Evidence Code section 403 problem does not appear, however, if the general population data base, which has been created without regard to race or ethnic background, is utilized.

"We must point out that the probative value of DNA matches using the general population data base may well be substantial. For example, the expected frequency of occurrence in the general population may be one in five thousand or even one in five million. This approach establishes a degree of probability that the suspect is the perpetrator, but it does so without *assuming* the suspect and the perpetrator belong to the same ethnic/racial background. Likewise, evidence sufficient under Evidence Code section 403 to support the preliminary fact as to the racial/ethnic background of the perpetrator alleviates this problem. We do not presume that evidence sufficient to support a preliminary factfinding in the instant case does or does not exist, our comments are designed to assist the trial court in assessing the relevancy of the proffered evidence." (*Pizarro I*, *supra,* 10 Cal.App.4th at pp. 92-95, fns. omitted.)

C.   KELLY HEARING

At the *Kelly* hearing on remand, Sensabaugh explained that the database population relevant for predicting allele frequency is the "[p]opulation of possible perpetrators [who] are possible sources of [the DNA] sample." He stated:

"This is the first case I have seen in which only the defendant's racial type is reported. That may or may not have been justified, depending upon the information that was provided to the FBI by the reporting agency."

No evidence beyond that presented at trial was presented at the hearing to establish that the perpetrator is Hispanic.

After the hearing, the trial court ruled that the DNA evidence was admissible. In its ruling, the court did not mention any finding on the preliminary fact question, but did conclude that the database used by the FBI was accepted in the scientific community. Nevertheless, the court's ruling "implies whatever finding of fact is prerequisite thereto." (Evid. Code, § 402, subd. (c); *People v. Williams* (1997) 16 Cal.4th 153, 196.) In other words, the court's admission of the evidence implies that the court determined the prosecution produced sufficient evidence that, if believed by the jury, would support a finding that the perpetrator is more likely than not Hispanic.

D.   ANALYSIS

As we explained in *Pizarro I*, sometimes the relevance and thus the admissibility of evidence depends on the existence of a preliminary fact. (Evid. Co de, §§ 403, subd. (a),[59] 350.) In such a case, the proponent of the evidence has the burden of producing evidence of the preliminary fact sufficient for a trier of fact to reasonably find by a preponderance of the evidence that the fact exists. (Evid. Code, § 403; *People v. Herrera* (2000) 83 Cal.App.4th 46, 61.) Until the preliminary fact is established, the evidence depending on it is neither relevant nor admissible. (*People v. Lucas* (1995) 12 Cal.4th 415, 466; *People v. Collins* (1975) 44 Cal.App.3d 617, 628 [evidence of threatening telephone call made to witness is not relevant until preliminary fact of caller's identity is established].)

The trial court should exclude the evidence "only if the 'showing of preliminary facts is too weak to support a favorable determination by the jury.' [Citations.] The decision whether the foundational evidence is sufficiently substantial is a matter within

---

59   Evidence Code section 403 provides in part: "(a) The proponent of the proffered evidence has the burden of producing evidence as to the existence of the preliminary fact, and the proffered evidence is inadmissible unless the court finds that there is evidence sufficient to sustain a finding of the existence of the preliminary fact, when: [¶] (1) The relevance of the proffered evidence depends on the existence of the preliminary fact …."

the court's discretion. [Citations.]" (*People v. Lucas, supra,* 12 Cal.4th at p. 466.) On review, we will not reverse the trial court's determination of the existence of a preliminary fact unless we find "the court exercised its discretion in an arbitrary, capricious or patently absurd manner that resulted in a manifest miscarriage of justice." (*People v. Jordan* (1986) 42 Cal.3d 308, 316; *People v. Ochoa* (2001) 26 Cal.4th 398, 437-438.)

Here, we cannot find evidence in the record sufficient to support a reasonable finding by the trial court that a jury could find by a preponderance of the evidence that the *perpetrator* is Hispanic. The only evidence suggesting such a finding was defendant's presence in the vicinity of the crime and his contact with the victim near the time of the crime. There was no evidence, for example, that only Hispanics could gain access to the vicinity or that only defendant was in the vicinity. We believe the fact that defendant was in the vicinity, and even spoke to the victim, *does not establish that the perpetrator is more likely than not Hispanic.* The Hispanic database frequency, presented as the figure applicable to defendant's case, was therefore irrelevant and inadmissible.

In light of both the inadequacy of the evidence and the manner in which the prosecution presented the evidence, it appears the trial court's finding that the perpetrator is more likely than not Hispanic was based on the assumption that *defendant is in fact the perpetrator*. This assumption, apparently also held by the FBI and the prosecution, is impermissible.[60] The record reveals that the FBI ascertained the perpetrator's otherwise unknown ethnicity by referring to *defendant's* ethnicity. For example, trial testimony

---

[60]   We are therefore deeply troubled by the prosecution's and the Attorney General's blurring of the lines between perpetrator and defendant. The *Kelly* hearing record abounds with such improper references, the prosecution's papers make similar violations, and the People's brief repeats them. Indeed, several issues in this case arise out of the assumption -- by both the FBI and the prosecution -- that defendant is the perpetrator.

regarding which database to choose when "someone is half Hispanic and half Caucasian" plainly referred to defendant.**61** In other words, the FBI used defendant's trait to describe the perpetrator.

Of course, the description of a perpetrator must be based on the *perpetrator's* traits, without regard to the defendant's or any other suspect's traits. A physical description or sketch of the perpetrator is intended to portray the perpetrator, in order to identify a suspect. The sketch artist first creates the sketch of the perpetrator, incorporating as many of the perpetrator's traits as possible, including his ethnicity if it is known. *Then* the defendant is held up to that description to determine whether he shares the perpetrator's traits. If, for example, the defendant happens to share the perpetrator's ethnicity, this will serve as further evidence against him; but, if the *perpetrator's* ethnicity is not known, the defendant's ethnicity cannot reasonably serve as evidence against him.

When, instead, the perpetrator's description is based on the *defendant's* traits, the following absurd scenario results: the sketch artist sits with the defendant, sketches him as the perpetrator, and then the prosecution introduces the sketch at trial as evidence that the defendant looks exactly like the perpetrator. If the sketch artist has no information regarding one of the perpetrator's traits -- his ethnicity, for example -- the artist does not refer to the *defendant's* Hispanic ethnicity to fill in the blank. Were the artist to do so, the prosecution's logic would follow this obviously faulty syllogism: the defendant is the perpetrator; the defendant is Hispanic; therefore, the perpetrator is Hispanic. The major premise is insupportable; the defendant's guilt is of course not a premise at all, but the

---

61   Adams explained that when "someone" is half Hispanic and half Caucasian, there is no half Hispanic and half Caucasian database to use; instead, the frequency is calculated using both databases, then the database producing the less detrimental frequency is used. Here, that was the Hispanic database.

ultimate conclusion sought by the prosecution. Reference to the defendant's ethnicity adds a trait to the perpetrator's description -- a fact not in evidence. Furthermore, the defendant necessarily shares that trait, to his prejudice, because it is *his* trait that has been added to the perpetrator's description. It is indisputable that the perpetrator must be described independently of the defendant, who plays absolutely no role at this stage.

In this case, the FBI used the Hispanic database because *defendant* is Hispanic. Consequently, the FBI's result estimates the frequency of the perpetrator's profile in *defendant's* ethnic population. The profile frequency, however, is intended to show the frequency of the perpetrator's profile in the *relevant* population, which, as Sensabaugh explained, is the population of *possible perpetrators* -- not the population of a particular defendant. Thus*, if the perpetrator's ethnicity is not known, an ethnic population is not relevant and cannot be used to estimate the frequency of the perpetrator's profile*. When, as here, the jury is informed that the relevant population is Hispanic, the jury draws the inference that the perpetrator is Hispanic. And when, as here, the jury is informed that the ethnic database was chosen based on the defendant's ethnicity, the jury draws the inference that the defendant is the perpetrator.

We note that commentators have agreed that reference to the defendant's ethnicity is an impermissible practice. NRCII states: "Usually, the subgroup to which the suspect belongs is irrelevant, since we want to calculate the probability of a match on the assumption that the suspect is innocent and the evidence DNA was left by someone else." (NRCII, *supra*, at p. 29.) "If the race of the person who left the evidence-sample DNA is known, the database for the person's race should be used; if the race is not known, calculations for all racial groups to which possible suspects belong should be made…." (*Id.* at p. 122 [Recommendation 4.1], 34.) Another commentator states: "To calculate a match proportion, laboratories need a reference population. The standard is to use the race of the suspect …. This makes no sense. A match proportion is calculated assuming the suspect is *innocent.* So the appropriate reference is the race of the criminal, assuming

the criminal is *not* the suspect." (Berry, *Statistical Issues in DNA Identification* in DNA On Trial: Genetic Identification and Criminal Justice (Billings edit., 1992) p. 106 (hereafter Billings.) The FBI's Worldwide Study explains:

> "The relative rarity of a DNA pattern in a suspect's ethnic subgroup, which might be of some academic interest, is not particularly relevant in the legal setting. To use the specific ethnic background of the suspect (which may be impossible to define) would presuppose that he or she is the true perpetrator. However, if the true perpetrator were known *a priori*, there would be no need for statistical estimates. Furthermore, if a particular subgroup were chosen as the reference database, for the majority of cases this would insinuate that a member of one subgroup is a more likely source of the crime scene evidence. Since the ethnicity of those people who are potential perpetrators rarely, if ever, is known, statistical estimates must be based on some sort of general population database. [¶] [T]he ethnic background of the suspect is not germane to selecting a reference database." (FBI Worldwide Study, Overview (1993) at p. 1.)

Another source states:

> "[T]he suspect is presumed innocent, so the suspect's claim of not contributing the [DNA sample found at the crime scene] is presumptively valid…. [¶] … The relative rareness of the DNA profile in the suspect's ethnic subgroup (or in any ethnic subgroup, for that matter) is not legally relevant …. It does not tell the jury anything about the likelihood that someone other than the suspect could have, in fact, left the sample at the crime scene. Instead, it only tells the jury the likelihood that someone in the suspect's ethnic subgroup could have left the crime scene sample. This has no bearing on the question of guilt or innocence in the typical criminal case. The relative rareness of the pattern i n some general population of potential perpetrators, on the other hand, does help the jury assess the likelihood that someone other than the defendant could have left the crime scene sample, and this has a direct bearing on the question of guilt or innocence." (Budowle, et al., *Reliability of Forensic DNA-typing Statistics* in Billings, *supra*, at pp. 81-82.)

And another explains:

> "In most cases … only a single suspect is tested, and without eye-witness or other reliable evidence, not even the race of the criminal is known.… [M]atching probabilities depend on the underlying allele and genotype frequencies (and therefore population), and if there is a considerable ethnic variability, the choice of the database used to evaluate a match is an

ethically significant action. One and the same sample DNA profile may be rare in one population, and therefore incriminate the suspect, but may be orders of magnitude more common in another. [¶] Morton[62] has rightly pointed out that the ethnic origin of the suspect is usually irrelevant … , and that the choice of the reference population should not be the expert's major concern. Since match probabilities are calculated under the assumption of *innocence*, the only argument for using allele frequencies from the suspect's population would be courtesy. There are good reasons to assume that under ethnic heterogeneity the suspect's profile is more frequent in his own population than in many (if not most) others." (Krawczak & Schmidtke, DNA Fingerprinting (1998) p. 80.)

The People's reply relies in great part on the *benefit* a defendant gains when his own ethnic population is used. The People contend the Hispanic frequency presented to the jury was conservative and beneficial to defendant in comparison to frequencies calculated from other ethnic populations.[63] This argument's flaw, however, is that it is not an *evidentiary* argument. It fails to recognize that only *relevant* evidence is admissible (Evid. Code, § 350), and that the proffered evidence (the Hispanic frequency derived from the Hispanic database) is relevant only if the *preliminary fact* (the perpetrator's Hispanic ethnicity) is proved by a preponderance of the evidence (Evid. Code, § 403). Here, there was no such proof and, as a result, the Hispanic frequency simply was not relevant. No amount of potential or actual numerical benefit to defendant could transform this irrelevant inadmissible evidence into relevant admissible evidence.[64]

---

[62]   Morton, N.E. (1993) Eur. J. Hum. Genet., 1, 172.

[63]   Also see *Axell*, in which the court explained that the Hispanic database had been correctly used since the defendant identified herself as Hispanic. (*People v. Axell, supra*, 235 Cal.App.3d at pp. 865-866.) The issue, however, is the ethnic identity of the perpetrator, not that of the defendant.

[64]   We are nevertheless in no position to engage in a prejudice analysis -- to weigh any numerical benefit against the inferential damage created by the presentation of the evidence, which revealed that both law enforcement and the prosecution had concluded the perpetrator and defendant are the same person -- because, due to the various errors committed, we do not know if defendant did in fact gain any numerical advantage (and, if so, its extent) from the erroneous use of the Hispanic database.

We conclude the trial court abused its discretion by impliedly finding sufficient evidence supporting the preliminary fact that the perpetrator is Hispanic and by failing to find use of the Hispanic database improper scientific procedure under *Kelly*'s third prong. The 1-in-250,000 figure derived from the Hispanic database population was irrelevant and inadmissible. Defendant's ethnicity was irrelevant and reference to it improper. Finally, the erroneous reliance on defendant's ethnicity promoted the inference that defendant is the perpetrator.

We comment briefly on the resolution of this problem. We find legally untenable the suggestion that several frequencies derived from various ethnic databases should be presented to the jury when there is insufficient evidence of the perpetrator's ethnicity. This suggestion illustrates the subtle, even unexpected, differences between the scientific and legal approaches to the same problem. Here, science logically promotes consideration of all possibilities, whereas law restricts consideration to possibilities it deems relevant. A conundrum such as this, bound to arise in scientific cases, can be detected and resolved only through the attentive and respectful contemplation of the two disciplines and the mindful evaluation of their separate and intersecting principles. Inevitably, some scientific principles, although correct in their scientific context, will not survive translation into legal application of relevancy principles.

Here, the legal problem that arises when several possible ethnic frequencies are presented is again one of preliminary fact -- now occurring multiply and simultaneously. In this case, just as there was insufficient evidence to justify use of a Hispanic database, *there was equally insufficient evidence to justify use of any other ethnic database because there was insufficient evidence of the perpetrator's ethnicity*, not merely insufficient evidence of the perpetrator's *Hispanic* ethnicity. Thus, the preliminary fact supporting use of *any* ethnic database was insufficiently proved. For this reason, if various ethnic frequencies are presented to the jury, *each* will have been admitted without adequate foundation and in violation of evidentiary requirements. Any ethnic frequency will be

98.

irrelevant and inadmissible. Furthermore, *the very mention of specific ethnicities encourages jurors to focus on ethnicity -- specifically the ethnicity of the defendant, the only suspect before them.* We therefore believe that when there is insufficient evidence to prove the preliminary fact of the perpetrator's ethnicity, a *single* frequency calculated from a general, multi-ethnic (i.e., non-ethnic) database should be presented as the profile frequency. Such a general database would be relevant because the perpetrator necessarily falls within the general population.

In addition, we believe cautious evaluation is appropriate because of the ambiguous nature of artificially defined ethnicities. The propriety of an ethnic database depends on the accuracy of both its creation and its utilization. These questions, among others, arise: Who determines that a sample person is Hispanic and should be placed in a Hispanic database? What are the criteria for doing so (e.g., the person's appearance, surname, self-description)?[65] Does the Hispanic database contain adequate and proportionate samples of all the various Hispanic populations to which the perpetrator, identified by an eyewitness as Hispanic, could belong? How accurate is the eyewitness's evaluation of the perpetrator's ethnicity (e.g., can an eyewitness mistake a person of Oriental, Native American, or African-American ethnicity for a person of Hispanic ethnicity)? Is the accuracy of an eyewitness's evaluation affected when the perpetrator is of mixed ethnicity? These uncertainties illustrate some of the problems involved in using an ethnic database and further support use of a general non-ethnic database.

---

[65] NRCII "recognize[s] that most populations are mixed, that the definitions are to some extent arbitrary, and that they are sometimes more linguistic (e.g. Hispanic) than biological. In fact, people often select their own classification." (NRCII, *supra*, at p. 57.)

VII.   MIXED DNA SAMPLE

A.   *INTRODUCTION & SCIENCE*

As we have explained, the initial step in genetic profiling is the determination of the lengths of the perpetrator's alleles at each locus. Normally, this is a straightforward procedure -- the scientist observes one or (the usual) two bands in the perpetrator's lane on the autorads and sizes the bands by comparing their locations to the locations of the size standards. Figure 40 is an example of a typical autorad.[66]
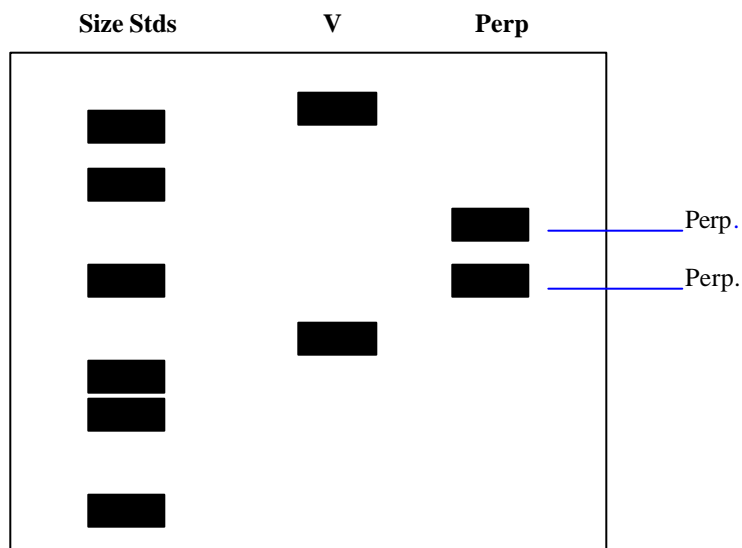
Fig. 40. *Normal Two-Band Perpetrator's Sample.*

In Pizarro's case, however, we are presented with a critical issue specific to cases in which discernment of the perpetrator's alleles is more complicated because *the perpetrator's DNA is mixed with (contaminated by) another person's DNA*. In these situations, it may be difficult if not impossible to locate the perpetrator's bands on the autorad. (NRCII, *supra*, at p. 129; NRCI, *supra*, at pp. 59, 66.)

Mixed DNA is a potential problem with postrape vaginal swab samples because they typically contain both perpetrator sperm cells and victim vaginal epithelial cells.

---

[66]   In all our examples, the victim is heterozygous, like the victim in this case.

(NRCII, *supra*, at p. 129; NRCI, *supra*, at pp. 65-66.) To separate the DNA from the two types of cells, scientists use a procedure called differential extraction, which relies on the different resistances of sperm and epithelial cell nuclei to breaking open.[67] (NRCI, *supra*, at pp. 65-66; Easteal, *supra*, at pp. 152-153; Butler, *supra*, at p. 32; Kirby, *supra*, at pp. 63-64; Robertson, *supra*, at pp. 54-55, 82-83.) Sometimes the procedure is not completely successful and some victim epithelial cell DNA may remain in the sperm fraction. When the autorads are produced, the scientist can usually see that the perpetrator's DNA contains more than the normal two bands and that one or two of them match the victim's bands. These findings reveal that the two types of DNA were not completely separated and that the DNA is mixed. (See NRCII, *supra*, at p. 129.)

The perpetrator/victim DNA mixture necessarily contains two alleles from the perpetrator and two alleles from the victim. Accordingly, autorads of mixtures generally reveal four separate and distinguishable bands, one for each of the four alleles in the mixture. The two victim's bands in the mixture can be discerned by comparing the mixed sample to the victim samples on the same autorad. The two bands in the mixture that match the victim's bands can logically be subtracted out of the mixture to leave the two remaining bands as the perpetrator's. (NRCII, *supra*, at p. 129 ["In many cases, one of the contributors -- for example, the victim -- is known, and the genetic profile of the unknown is readily inferred."].) Thus, in a four-band perpetrator/victim mixture, the mere locations of the four bands can provide adequate information for discerning the

---

[67]   "The differential extraction procedure involves preferentially breaking open the female epithelial cells with an incubation in a SDS/proteinase K mixture. Sperm nuclei are subsequently lysed by treatment with a SDS/proteinase K/dithiothreitol (DTT) mixture. The DTT breaks down the protein disulfide bridges that make up sperm nuclear membranes [citation]. Differential extraction works because sperm nuclei are impervious to digestion without DTT." (Butler, Forensic DNA Typing (2001) at p. 32 (hereafter Butler).)

perpetrator's alleles. Figure 41 shows two examples of two-band perpetrator's samples, four-band mixtures, and the subtraction out of the victim's bands from the mixture.
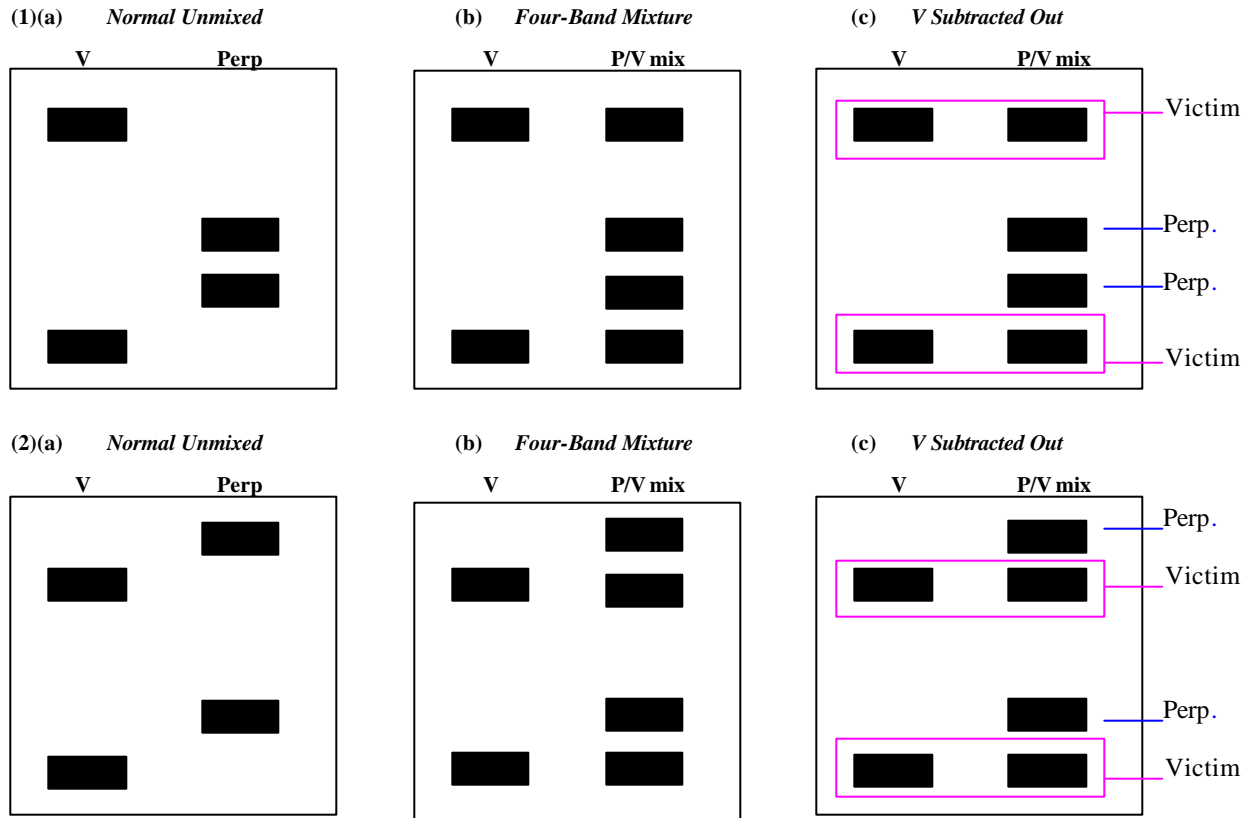
Fig. 41. *Two Examples of Four-Band Perpetrator/Victim Mixtures.*
NOTE: The victim's bands in a four-band mixture can be identified and subtracted out to leave the perpetrator's bands.
P/V mix = perpetrator/victim mixture fraction of evidentiary sample.

When the two victim's bands are subtracted out, the two remaining bands then represent the perpetrator's profile (genotype) at that locus. The two bands will later be compared to the defendant's bands, and, if a match is found, used in the statistical calculations to determine the overall perpetrator profile frequency.

A more complicated situation arises, however, when a mixture contains only two or three bands, rather than four. Because every person possesses two alleles at each locus, the presence of fewer than four bands in the mixture means one or more of the bands is probably *masked* by (superimposed on or coalesced with) another band. In these

situations, the victim's alleles cannot simply be subtracted out to reveal both of the perpetrator's alleles; the superimposed bands may conceal the perpetrator's genotype.[68] (NRCI, *supra*, at p. 66 [two-band mixture: "if the sperm fraction shows a genotype that matches that of the victim, one cannot conclude that this represents the genotype of the perpetrator, inasmuch as it could be due to residual vaginal epithelial cells."].) (Fig. 42.)

(1) *Two-Band Mixture*   **(2)** *Three-Band Mixture*
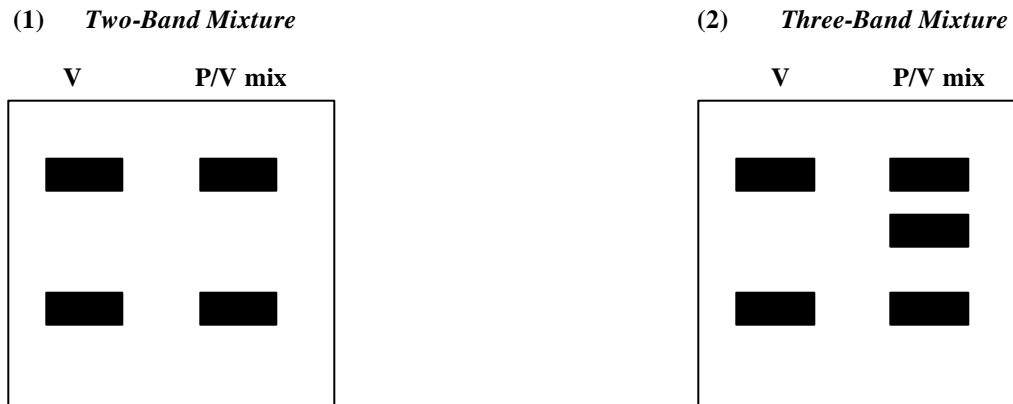
Fig. 42. *Indistinguishable Mixture Bands.*
NOTE: Four alleles are now contained in only two or three bands in the P/V mix.
The perpetrator's bands are no longer discernible by subtracting out the victim's bands.

In a two-band mixture, the perpetrator's masked profile may be one of three profiles or genotypes, as established by uncontroverted testimony, *post*. These three perpetrator profiles are shown schematically in figure 43: (1) heterozygous, sharing both bands with the victim (there are two alleles within each band), (2) homozygous for one allele, sharing one band (there are three alleles within one band and one allele within the other), or (3) homozygous for the other allele, sharing one band (there is one allele within one band and three within the other). (See NRCII, *supra*, at p. 162 [a two- or three-band mixture may mean that one of the contributors produced a single band.])

---

68   For the sake of simplicity, we ignore band intensity for the moment.

(1) Perpetrator is *heterozygous* (AB):
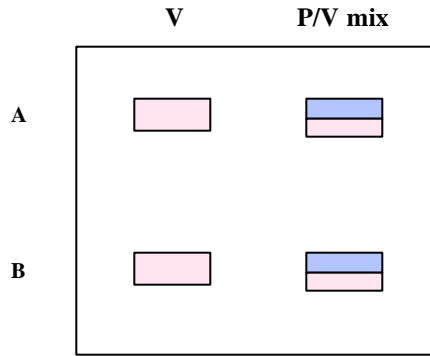
(2) Perpetrator is *homozygous* for top band (AA):
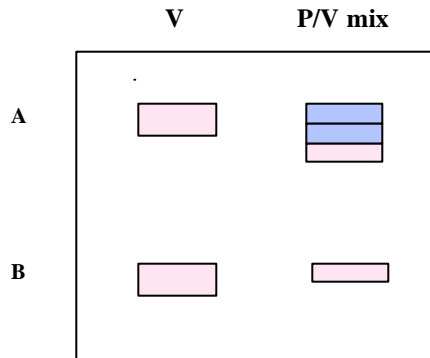
(3) Perpetrator is *homozygous* for bottom band (BB):
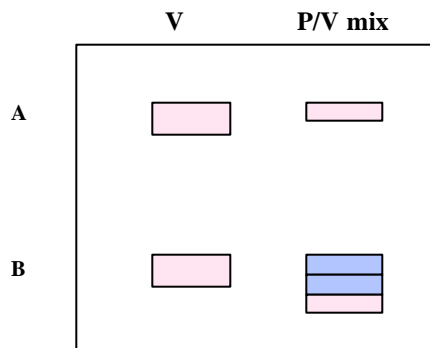
Fig. 43. *Three Possible Perpetrator Profiles within Two-Band Mixture.*
NOTE: ▭ = V's 1st allele; ▭ = V's 2d allele; ▭ = Perp's 1st allele; ▭ = Perp's 2d allele.
Each allele or group of alleles constitutes a single band. All bands should be presumed to be approximately the same size/intensity despite their schematic appearance.

In the present case, two of the three autorads (D1 and D4) contain four bands in the perpetrator's sample (as in fig. 41(1)(b) and (c), *ante*), two of which match the victim's bands, demonstrating that the perpetrator's and victim's DNA are mixed. (Figs. 44 & 45, lanes 10 & 11.) Recall that the DNA is therefore mixed on *all* autorads; although the mixture may be *revealed* by a single autorad, it *exists* identically on all of them. (NRCII, *supra*, at p. 162 ["It is also possible that there are only two bands, but other loci indicate that the stain is mixed."].) Autorads D1 and D4 are typical examples of a four-band mixture from which the victim's bands can be subtracted out to reveal the perpetrator's bands. (Figs. 44 & 45.)
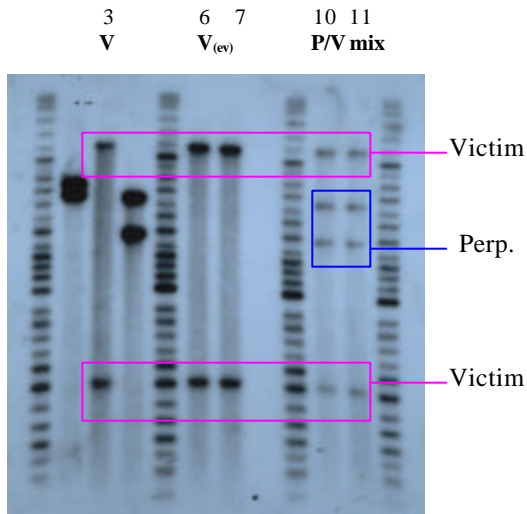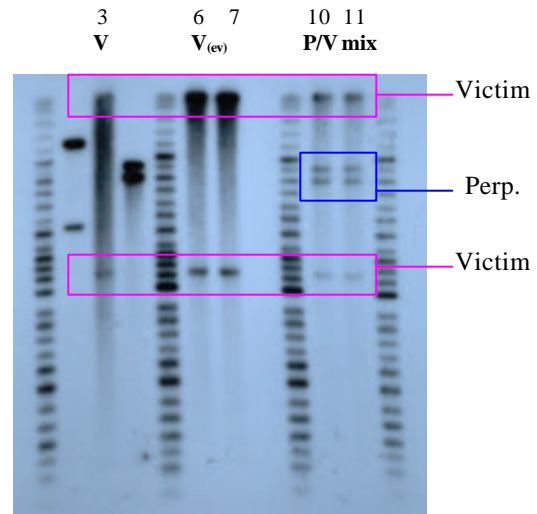
Fig. 44. *D1 Autorad.*   **Fig. 45.** *D4 Autorad.*

NOTE: Four-band mixture in lanes 10 and 11 of both autorads. The victim's bands in the mixture are discernible by comparison to the victim's known bands in lanes 3, 6 & 7. Once the victim's bands are subtracted out, the perpetrator's bands are also discernible.

The D2 autorad, however, presents the more complicated two-band mixture in which both bands are shared by the heterozygous victim, as in figure 42(1), *ante*. (Fig. 46.) The mixture still contains four alleles, but they now exist in some combination within only two bands.
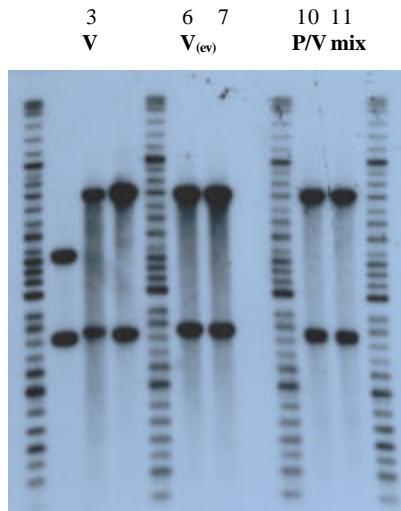
Fig. 46. *D2 Autorad.*
NOTE: Two-band mixture in lanes 10 and 11.
The victim's bands cannot be subtracted out.

The FBI scientists determined that the perpetrator is *heterozygous* at the D2 locus, as in figure 43(1), *ante*. In other words, the FBI concluded that the two-band perpetrator/victim mixture on the D2 autorad should be interpreted as representing two heterozygous individuals (AB and AB), in two sets of superimposed bands. The FBI then multiplied the frequency of this heterozygous genotype by the other two genotype frequencies (from the D1 and D4 autorads) to obtain the perpetrator's overall profile frequency.

At the *Kelly* hearing, the prosecution supported the FBI's conclusion that the perpetrator is heterozygous at the D2 locus with the theories that the uncertainty in the perpetrator's profile at that locus can be explained by reference to (1) defendant's profile and (2) relative band intensities. We summarize these two lines of reasoning as follows:

(1)   **Defendant's Profile** --

Reference to *defendant's* profile explains any ambiguity in the perpetrator's profile because:

defendant is heterozygous (AB), and his bands match the two bands in the perpetrator/victim mixture.

(2)   **Equivalent DNA Quantity** --

The two bands in the perpetrator/victim mixture each contain the *same amount of DNA* -- one contains two A alleles and the other contains two B alleles because:

(a)   The intensities of the two bands in the D2 mixture are *equal*; therefore the four alleles in the mixture must be divided equally in sets of two superimposed alleles (AA and BB) (as in fig. 43(1), *ante*);

(b)   The intensities of the bands in the two-band D2 mixture are *twice as strong* as the intensities of the bands in the four-band D1 and D4 mixtures; therefore the D2 bands must contain twice as much DNA as the D1 and D4 bands, which are known to contain one allele each.

We address these two propositions in turn, examining the evidence supporting and refuting each.

B.   REFERENCE TO DEFENDANT'S PROFILE

The People assert that reference to defendant's genetic profile provides guidance in the interpretation of the D2 autorad's perpetrator/victim mixture: because defendant is heterozygous at that locus, *the perpetrator should be assumed to be heterozygous also*.

1.   Prosecution Witnesses

a.   Sensabaugh

Sensabaugh explained on cross examination that "the most straightforward inference [from the two-band D2 mixture] is that in this case both individuals share indistinguishable typing at this particular locus [i.e., both individuals are heterozygous]."

107.

Defense counsel then asked Sensabaugh whether he was aware of the method in which all possible genotype frequencies in a mixture are added together.[69] Sensabaugh responded:

> "[SENSABAUGH:] That would be -- that is appropriate in some situations. If one has a four-band pattern and in making comparisons of the four-band pattern one cannot exclude the possibility of various combinations, then all the non-excluded frequencies of all the non-excluded combinations are put together."

He then explained that the National Research Council's recommendation (in NRCI) to add the frequencies for mixtures is "a bit naïve to anyone who has actual forensic practice," notwithstanding NRCI, on which Sensabaugh was a signatory. He agreed, however, that the two-band mixture on the D2 autorad could represent a mixture of the heterozygous victim and a homozygous perpetrator who shared one band with the victim (see fig. 43(2) & (3), *ante*), and that the frequency would be affected if this possibility were taken into account as NRCI recommends. Nevertheless, Sensabaugh stated:

> "[W]hen the presentation is as straightforward as this is[,] those numbers are not, in my experience, usually calculated. It is usually in more complex mixture cases that -- where there may be known and unknown individuals mixed together that one engages in that exercise."

On redirect, the prosecutor asked Sensabaugh if he knew of any fact in this case that would make it more likely that the victim and defendant would share the same alleles. Sensabaugh responded that it was his understanding that they were half siblings, and, in light of this fact, the bands were where Sensabaugh would expect to see them if the mixture contained the victim's and defendant's DNA. He thought the results were "interpretable."

---

[69] Because two-band mixtures present three possibilities, some authorities recommend that the statistical calculations account for all three genotypes by adding their frequencies. (NRCI, *supra*, at pp. 58-59.)

b.   *Chakraborty*

Chakraborty testified the D2 autorad reveals that the profiles of the victim and defendant are very similar, and, "as a consequence," the profiles of the victim and the perpetrator/victim mixture are also very similar.[70] Chakraborty stated that the half sibling relationship of defendant and victim might explain that occurrence. Chakraborty's calculations indicated the chance of two shared alleles is five times greater in half siblings than in unrelated persons. Chakraborty explained this calculation did not require any assumption about the source of the evidentiary sample because the D2 autorad shows defendant's profile matches the perpetrator's profile, and the victim's profile matches the victim's fraction of the evidentiary sample. It is very unusual to observe a defendant's profile so similar to the victim's profile, but "given the fact that they are half siblings these observations are expected to be observed."

On cross-examination, Chakraborty stated there is no way of telling whether the two bands in the D2 mixture came from the perpetrator or the victim. But the autorad did not exclude defendant as a possible perpetrator. Chakraborty did not disagree with the NRCI recommendation to add all possible combinations for mixed samples.

Chakraborty did not know of any laboratories that excluded autorads when the victim and defendant's profiles matched; he had heard of the concept, but did not understand its logic. He agreed that excluding the D2 autorad would change the frequency of the profile "substantially."

---

[70]   Chakraborty said: "And as a consequence the evidentiary samples, female fraction and male fraction DNA profiles were also very similar."

c.   *Adams*

Adams, who oversaw the FBI's DNA analysis in this case, testified he did not see sufficient reason to exclude the D2 autorad from the calculations,[71] although he agreed that excluding an autorad can make a significant difference in the resulting frequency. He stated it is impossible to determine the source of a band on an autorad. He was not aware of NRCI's recommendation to add all possible profile frequencies in the case of a mixture such as the one on the D2 autorad, and he did not believe NRCI did in fact suggest such an approach. But he explained that when the mixture contains only two bands, it is impossible to discern whether the mixture consists of two homozygous people; it is impossible to tell which bands are contributed by the victim and which by the perpetrator.

d.   *Conneally*

On direct examination, Conneally explained that the D2 autorad should not be excluded from the calculation because half siblings would be expected to share a band more often. He stated:

> "[CONNEALLY:] … The defendant and the victim shared a band in common there. And that's always a possibility to share a band. And, in fact, if the defendant were the perpetrator would he not be -- I understand that they were related, so this would not be unusual at all. Half siblings would be expected to share one band out of six. So, I do believe that there was no reason to -- there is no reason to exclude the results of D2S44."

On cross-examination, Conneally stated it is impossible to tell from a band on an autorad whether the band was contributed by the victim, the perpetrator, or both, and in what quantities.

---

[71]   His reasons for this conclusion fall under the second proposition, Equivalent DNA Quantity Based on Band Intensities, *post*.

2.   Defense Witnesses

a.   *Shields*

Shields testified that in a two-band mixture there is no way to determine whether the bands were contributed by the victim or by someone else. When the victim's fraction and the perpetrator's fraction of the evidentiary sample both contain the same bands (have the same profile), Shields believed the autorad should be excluded from the calculation. It is possible that the perpetrator is either homozygous or heterozygous. (See fig. 43, *ante*.) If the perpetrator is homozygous and defendant is heterozygous, defendant is actually excluded as a potential perpetrator (i.e., he is exonerated). Similarly, if the perpetrator is heterozygous and defendant is homozygous, he is again exonerated. There is no way to know what the mixture means. For these reasons, when there is even one shared band between the victim and defendant, the autorad should be excluded entirely.

b.   *Zabell*

Zabell explained that a mixture containing only two bands is a very different situation than a mixture containing four bands. When there are four bands, the victim's bands can be subtracted out, leaving the two bands that presumably belong to the perpetrator. But when the mixture contains only two bands, there are several possible perpetrator profiles represented by those two bands. First, the perpetrator could be homozygous for one band, or appear to be homozygous for that band because his two bands are so close together as to coalesce into one band on the autorad. (See fig. 43(2), *ante*.) Second, the perpetrator could be homozygous, or apparently homozygous, for the other band. (See fig. 43(3), *ante*.) Third, the perpetrator could be heterozygous for the same two bands as the victim. (See fig. 43(1), *ante*.) Fourth, the perpetrator could be heterozygous for one of the same bands as the victim, but his second band "ran off" the end of the gel and is not visible on the autorad. Fifth, the perpetrator could be heterozygous for the other band, but his second band ran off the end of the gel and is not

111.

visible on the autorad. Sixth, both of the perpetrator's bands, whether homozygous or

heterozygous, could have run off the gel and are not visible on the autorad. Ignoring the

unlikely cases of run-off, there are three possible profiles for the perpetrator represented

by the two-band mixture.

Zabell noted the FBI's calculation, however, took into account only the single

possibility that the perpetrator is heterozygous, sharing both bands with the heterozygous

victim. (See fig. 43(1), *ante*.) He explained that when calculating the match probability

-- the chance that a randomly chosen person will match the profile -- it is incorrect to

account for only one possibility as the FBI did in this case. If only the defendant's profile

is used to calculate the perpetrator's profile, an assumption is being made that the

defendant is the perpetrator. The proper procedure is to add up the frequencies for all the

possible explanations for the banding pattern to determine how frequently a match of any

possible kind could arise. Following this proper procedure significantly increases the

likelihood of a random match in this case. When the frequency is recalculated to take

into account the three possible profiles (add their frequencies together, but otherwise use

the FBI's method), using the updated H4 database, the frequency of the perpetrator's

overall profile becomes 1 out of 20,000, instead of 1 out of 894,000.

Zabell stated the existence of mixtures in forensic samples is not an uncommon or

new phenomenon. Scientists expect that taking into account the extra possibilities

presented by a mixture can cause the profile to become substantially more common.

NRCI plainly states that all possible genotype frequencies should be added together in the

case of mixtures. In the two-band mixture situation, some labs exclude the autorad from

the frequency calculation; others add up the possible frequencies. Zabell knew of "no

one who would say that when more than one profile could match a pattern you should not

add up the frequencies for the different profiles." He believed that, when more than one

profile could be declared to match the evidence sample's profile, there was "essentially

unanimity" among the scientific community that "those other frequencies must be taken

into account in the calculations." The "clear con[s]ensus" was that the calculations should be performed in this manner for the D2 autorad in this case.

Zabell explained that to disagree with this principle one would have to argue that, if defendant were homozygous, he would be excluded as a possible perpetrator. To avoid such a conclusion, the other two possible perpetrator profiles must be included in the calculation to account for all possible matches with the profile.

On cross-examination, Zabell stated that, although the likelihood of a three-locus match is "in general … a quite rare event[, w]e're in a special case here. And that's the single biggest concern I would have for the calculations. We do have a mixture and that obviously affects the frequency."

Later, the following colloquy took place:

> "[PROSECUTOR:] I understand that you have to assume the defendant is the perpetrator in order for them to calculate the significance as they did [in this case].

> "[ZABELL:] Well, strictly speaking what the significant calculation does is it doesn't refer to -- it doesn't refer to the defendant or suspect at all. They say[, ']suppose we choose someone at random, what's the chance that it would match the evidence profile[?'] So the calculations in certain instances does [*sic*] not refer at all to the suspect.

> "[PROSECUTOR:] There's nothing inconsistent between the defendant's profile and the questioned sample, is there?

> "[ZABELL:] That's right. When you use the statistical match rule, the defendant's profile [is] declared to match the [evidentiary] bands, yes.

> "[PROSECUTOR:] So you are not calling into question the FBI's call of a match in this case?

> "[ZABELL:] No."

In regard to the mixture on the D2 autorad, the following exchange occurred:

> "[ZABELL:] … I think the FBI is wrong. [¶] … [¶] If we have a [two-band] mixture then we have to do the calculation for all three potential profiles. The perpetrator is heterozygous for the top band A and the bottom band B, homozygous for band A or homozygous for band B. All those

three possibilities would be taken into account because that is precisely because of [the mixture revealed by the] D1 and D4 [autorads].

"[PROSECUTOR:] What about the relationship between Mr. Pizarro and the victim in this case? What effect does that have on your interpretation of this autorad?

"[ZABELL:] None because you remember the way the calculation is phrased. I mean I did see Dr. Chakraborty made some reference to that, which puzzled me the way the calculations go. You were saying here's the evidence sample, the evidence profile, and we have declared a match with the suspect.

"Now, the question is[, ']given the evidence profile, suppose we went out and picked an unrelated person.['] That's often investigated in the summary of the calculation. [']Suppose we chose someone who's unrelated, what would be the chance that we would get a matching profile?['] … The fact that the suspect is or isn't related is really irrelevant to that calculation.

"[PROSECUTOR:] But the fact is this suspect is related to the victim in this case, at least in hindsight [that fact] must affect the way you -- affect the autorad.

"[ZABELL:] Guess I don't see that, Counselor. I mean, it's true that because Mr. Pizarro is related to the victim that he would have a higher chance of matching up at any of the loci at one of the bands.

"[PROSECUTOR:] Which is what happened here.

"[ZABELL:] But, again, I mean, the calculation doesn't -- I mean, the calculation really does not refer to that. The calculation says, 'Here's the evidence sample.' The calculation in effect says, 'We don't have a suspect. Here's the evidence sample, suppose I chose someone at random, what's the chance they would match up?' Right. There is nothing in that sentence that refers to the suspect."

c.   *Bakken*

Bakken stated that in the case of the two-band D2 mixture there is no way to tell whether the perpetrator is heterozygous, homozygous for the top band, or homozygous for the bottom band. (See fig. 43, *ante*.)

d. Muller

Muller explained: "The genetic constitution of the perpetrator [in the D2 mixture] is somewhat ambiguous …. It's going to have DNA from the victim … and [DNA] from the perpetrator. But the perpetrator can have a variety of genetic constitutions …." The perpetrator's profile could be represented by just the top band, just the bottom band, or both bands. (See fig. 43, *ante*.)

> "All those alternative genetic states for the perpetrator produce an evidence sample that's consistent with a match. We can't distinguish whether the perpetrator is [the first, second, or third possibility]."

People who possess any of these three possible profiles cannot be excluded from the pool of possible perpetrators.

> "Now, it happens that Mr. Pizarro only has the [heterozygous] combination. Statistically, we have to take into account that if he [were homozygous for one band], we couldn't have excluded him. Had he [been homozygous for the other band], we couldn't have excluded him. So, all those combinations need to be taken into account because of the particular results in this case, which are that the DNA from the victim and suspect [*sic*] have not been completely separated."[72]

Muller stated that, if only one of the possible profiles were considered, the result would grossly underestimate the number of people in the population who might possess the perpetrator's profile. Incorporating all possible profiles into the calculation would make the profile "significantly more common." Further justification for including all possible profiles is the fact that, if the perpetrator were indeed homozygous for either one of the bands, Pizarro would be completely excluded as a possible perpetrator. Muller

---

[72] Defense witness Muller's mistaken use of "suspect" rather than "perpetrator" demonstrates the ease with which such an error can be made. Further proof of the potential for inadvertent misuse is found in this court's footnote 12 in the recent case of *People v. Brown, supra,* 91 Cal.App.4th at page 630. There, as Pizarro's appellate counsel noted at oral argument in this case, we also mistakenly used the term "suspect" for "perpetrator."

stated, "We have the possibility that the evidence may, in fact, be inconsistent with the conclusion that [the perpetrator's and defendant's profiles] match. So, the information from this particular locus [the D2 autorad] is not interpretable as a definitive match."

Muller noted that Chakraborty testified he agreed with NRCI's recommendation to add all possible profiles in a mixture. Muller interpreted Chakraborty's testimony as follows:

> "[MULLER:] Generally, what he said is, given that the suspect and victim have a genetic relationship, they can share alleles in common. The finding of the evidence of completely overlapping bands for the victim and the suspect is five times more likely in this case. Therefore, he's not surprised at all by seeing this pattern. To my mind, it means he thinks he understands the nature of that pattern.

> "[DEFENSE COUNSEL:] Another way of saying that is, he's assuming, because Mr. Pizarro, the defendant, has a two-banded pattern that it must be him in [the evidence sample]. He's only going to count the possibility that it's Mr. Pizarro or somebody else with two bands, right? I mean, he's using information about the suspect to make inferences about the pattern in the evidence, which is akin to assuming that the suspect's DNA must be in the evidence sample?

> "[MULLER:] But, of course, that's exactly why we do a DNA analysis at trials, to determine what extent that's a reasonable conclusion.

> "[DEFENSE COUNSEL:] And it's fair to say that when you're calculating the significance of a match, that's the exact opposite of what your [*sic*] supposed to do? You're trying to find out how many people other than your defendant could fit the pattern?

> "[MULLER:] Right. The presumption is if they think there's a match, if we were to choose people at random, what's the likelihood that people chosen at random would match in the fashion seen here? And the fashion of the match we've seen here, as I explained earlier for D2S44, is somewhat ambiguous, because there's several [*sic*] different genetic patterns the suspect could have and be declared a match here.

> "That level of ambiguity has to be taken into [account] [.] … Mr. Pizarro's particular genetic relationship to the victim [is] completely irrelevant for assessing that level of ambiguity, because, as we said earlier,

it presumes his DNA is in the evidence [sample], which, of course, is the whole focus of this study."

Muller also commented on Sensabaugh's testimony, noting Sensabaugh testified he was a signatory on the NRCI report that recommends adding frequencies in such a case and testified that frequencies should be added together when mixed samples contain a known and unknown source. Muller commented that this is "precisely the kind of situation we have here" -- a two-person mixture in which one person, the victim, is known, and the other, the perpetrator, is unknown.

3. Analysis

The defense witnesses explained extensively and unequivocally that the two-band mixture on the D2 autorad represents three possible profiles and that the perpetrator's true profile cannot be discerned from the autorad bands. The perpetrator could be heterozygous or homozygous. Shields recommended that autorads with mixtures such as this be entirely excluded from the statistical calculations, in part because two of the three possible perpetrator profiles would actually exclude defendant as a suspect. Zabell and Muller explained that the proper procedure in such a case is to take into account all three possible profiles by adding their frequencies, thereby increasing the commonness of the profile and the likelihood of a random match in the population.

The prosecution witnesses did not contradict the defense theory that the perpetrator could be homozygous at the D2 locus. Sensabaugh agreed that the two-band mixture on the D2 autorad could represent a homozygous perpetrator and that consideration of such a possibility would affect the frequency calculation. Chakraborty stated it is impossible to discern whether the two bands come from the victim or the perpetrator and he did not disagree with the recommendation to add all possible profiles for mixed samples. Adams testified it is impossible to determine the source of a band, and in a two-band mixture it is impossible to determine whether the mixture contains a homozygous individual and whether the bands come from the victim or the perpetrator.

Conneally also agreed it is impossible to discern who contributed a band in a mixture, and in what quantities.

Prosecution evidence that could be portrayed as contradictory on this specific issue was consistently and expressly based on the assumption that the DNA mixture contains *defendant's* DNA rather than the *perpetrator's* DNA -- and was therefore based on the assumption that *defendant is in fact the perpetrator*. Sensabaugh stated "the most straightforward inference" is that both people in the mixture are heterozygous. He explained that NRCI's recommendation to add all possibilities is naïve and that such an approach is usually limited to cases in which there is a mixture of "known and unknown individuals." As defense witness Muller noted, Sensabaugh's description precisely fits the situation in this case: the victim is known and the perpetrator is unknown. Sensabaugh deemed the results of autorad D2 "interpretable" because the autorad displays the results he would expect to see if the mixture contains the victim's and *defendant's* DNA. Chakraborty explained that the perpetrator and victim profiles are similar because the *defendant* and victim profiles are similar. Conneally said the results would not be unusual at all if *defendant* is the perpetrator.

As in the previous issue, the assumption that the defendant is the perpetrator is entirely improper. Calculation of a profile or match probability is based on the profile of the *perpetrator*, distinct and separate from the defendant or any other suspect. We return to the physical profile as an analogy. A sketch artist creates an artistic representation of the perpetrator from the eyewitness's description of the perpetrator's physical features. The eyewitness describes the perpetrator as having, for example, black hair, blue eyes, and 5-foot-8-inch stature.[73] The artist's sketch should portray the *perpetrator*, not the

---

[73] Using physical features and a composite sketch as an analogy requires the assumption that physical features, like genetic features, are immutable. Thus, the perpetrator could not have changed the color of his hair, and so on.

defendant or any suspect, and should be produced without reference to the appearance of any suspect. If a defendant happens to match the artist's sketch of the perpetrator, the match provides more evidence against him.

If the eyewitness cannot recall one of the perpetrator's features -- say, his hair color -- the sketch artist does not refer to the *defendant's* black hair color to provide the missing information. Doing so relies on this improper syllogism: the defendant is the perpetrator; the defendant has black hair; therefore, the perpetrator has black hair. As in our earlier discussion, the major premise cannot be the defendant's guilt. The eyewitness must describe the perpetrator independently of the defendant. The *perpetrator's* black hair color must first be established independently as a preliminary fact before the *defendant's* black hair color is either relevant or probative. The proper syllogism states: all possible perpetrators have black hair; the defendant has black hair; therefore, the defendant is a possible perpetrator.

If the eyewitness is uncertain about the perpetrator's hair color, but can narrow the color down to either black, brown, or blond, should each of the three possibilities be taken into account and presented to the jury? The logic supporting an affirmative answer states: all possible perpetrators have black, brown, or blond hair; the defendant has black hair; therefore, the defendant is a possible perpetrator. Although initially appealing, this logic improperly ignores the fact that if the perpetrator actually has brown or blond hair, the defendant simply *is not* the perpetrator. The correct logic requires a *choice* of these three possible syllogisms: (1) all possible perpetrators have *black* hair; the defendant has black hair; therefore, the defendant is a possible perpetrator; (2) all possible perpetrators have *brown* hair; the defendant has black hair; therefore, the defendant is not the perpetrator; (3) all possible perpetrators have *blond* hair; the defendant has black hair; therefore, the defendant is not the perpetrator.

It would defy the principles of evidence to allow the eyewitness to testify that the perpetrator has either black, brown, or blond hair when there is no way of establishing the

preliminary fact of which hair color the perpetrator actually possesses. This testimony is neither relevant nor probative, but it *is* potentially damning because it draws the defendant into the pool of possible perpetrators when in reality it more likely excludes him -- two of the three possibilities *exonerate* him. The eyewitness's testimony regarding the perpetrator's three possible hair colors is not admissible.

These principles apply equally to the genetic profile, in which a scientist creates a genetic representation of the perpetrator from what the DNA, the genetic eyewitness, describes about the perpetrator's genetic features. Each locus (autorad) can be thought of as describing a single physical feature in the sketch -- hair color, for example. In Pizarro's case, as in our sketch scenario, the description of one of the perpetrator's features -- the D2 locus -- is uncertain. The perpetrator could be one of three genotypes at this locus. Some witnesses suggested the uncertainty could be cured by assigning the defendant's feature to the perpetrator. Others suggested accounting for all three possibilities. Still others suggested discarding evidence of that feature altogether because of its uncertainty and potential to exonerate. As we have stated, we reject all theories but the last.

Although this issue is shrouded in scientific technicality, it is again one of preliminary fact. (Evid. Code, § 403.) Unless there was sufficient proof that the *perpetrator* is heterozygous at the D2 locus, *defendant's* heterozygous genotype at that locus was irrelevant. And if the perpetrator's genotype is not decipherable, a match between the two genotypes could not be declared.

First, *defendant's profile was not a proper reference for clarification of the perpetrator's profile* -- that is, *defendant's* genotype could not be used to prove the *perpetrator's* genotype, proof of which was required to render *defendant's* genotype

relevant in the first place.[74] Defendant's genotype (like his ethnicity in the previous issue) was irrelevant and inadmissible in the absence of sufficient proof of the perpetrator's genotype. Reference to defendant's genotype as an incriminating trait was error, and reliance on defendant's genotype was based on the improper assumption that *defendant is in fact the perpetrator*.[75]

We again note that various commentators agree that the perpetrator's genetic profile must be ascertained independently of the defendant's profile. DNA bands "must be identified separately and independently in [the perpetrator's and defendant's] samples. It is not permissible to decide which features of [a perpetrator's] sample to count and which to discount on the basis of a comparison with a [defendant's] sample, because this can bias one's interpretation." (NRCI, *supra*, at p. 53.) "In all cases, each lane must be

---

[74]   The People's brief goes so far as to remind us the perpetrator and defendant are the *same person*: "it is clear that there are two sets of two bands [in the D2 mixture]: one set for the victim, the other set for appellant (the perpetrator)." The brief concludes that because the defendant is heterozygous, there is no realistic probability of a homozygous perpetrator because "[n]either [the defendant] nor the evidence suggests this possibility."

[75]   We are aware NRCII has modified NRCI's position, now suggesting that adding together all possible profiles is "hard to justify, because it does not make use of some of the information available, namely, the genotype of the suspect." (NRCII, *supra*, at pp. 129-130.) NRCII recommends using a likelihood ratio which takes into account the defendant's profile because likelihood ratios are especially useful "provided that prior odds are available on the hypothesis that the two DNA profiles have the same source. (Prior odds are the odds that the two DNA samples came from the same person on the basis of [*evidence*] *other than the DNA*.)" (*Id.* at pp. 130-131, italics added.) Bayes's Theorem, invoked when the prior odds are multiplied by the likelihood ratio (*ibid.*), is used regularly in paternity cases, but rarely in criminal cases (*id*. at pp. 131-132, 200). As NRCII explains, "The main difficulty is probably an unwillingness of the courts to ask juries to assign odds on the basis of non-DNA evidence." (*Id.* at p. 132.)

We too see great difficulties with this approach; nevertheless, it was not taken in this case. Here, the FBI calculated the *random match probability*, not the likelihood ratio and prior odds; thus, as we have explained, there was no occasion for consideration of defendant's profile in the calculation.

121.

evaluated independently -- the presence of a band in one lane must not influence whether a questionable signal in another lane should be identified as a band." (OTA, *supra*, at p. 65.) Indeed, "[c]ommentators have noted a disturbing tendency for forensic analysts to resolve ambiguities in DNA patterns in a manner consistent with the expected result. The analyst may, for example, infer that a discrepancy between two DNA profiles on one autorad *must* be an artifact (rather than a true genetic difference) because there is a match on the other autorads or, worse yet, because other evidence in the case suggests the two profiles have a common source. Professor Eric Lander has condemned this kind of bootstrap interpretation in forensics because 'one runs the risk of discounting precisely those differences that would exonerate an innocent defendant.' An analyst who too readily dismisses discrepancies in a DNA test that do not fit with other evidence can mistakenly conclude that weak, equivocal evidence is quite powerful, and thereby mislead the trier of fact." (Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons From the "DNA War"* (1993) 84 J. Crim. Law & Criminol. 22, 53-54, fns. omitted (hereafter *Thompson*).)

Second, *because the perpetrator possesses only one genotype at a locus, only that one genotype was relevant*. The other two possible genotypes at the D2 locus were irrelevant. If the prosecution could not prove which genotype the perpetrator possesses at a certain locus, then there was no relevant evidence to admit from that locus. But, here, the most compelling reason for demanding proof of the perpetrator's genotype and for refusing to admit evidence of all three possible perpetrator genotypes is that the other two possible genotypes were more than irrelevant -- they potentially proved defendant's *innocence.* Thus, the evidence that was admitted to incriminate defendant actually had a greater chance of exonerating him. If the perpetrator is not heterozygous (i.e., if he is either homozygous for the top band or homozygous for the bottom band), then defendant does not match the perpetrator and he is *excluded* as a possible perpetrator. Only if the perpetrator is heterozygous does defendant match and become a possible perpetrator.

In sum, if the trial court admitted the D2 autorad without proof of the preliminary fact that the perpetrator is heterozygous, then it admitted irrelevant and highly damaging evidence against defendant, which in essence had a two-out-of-three chance of exonerating him. The D2 autorad evidence was used as a multiplier in the statistical calculation, the result of which was to make the perpetrator's profile rarer and defendant's possession of it more incriminating. Evidence from the D2 autorad was inadmissible unless the perpetrator's profile is discernable from the two-band mixture without reference to defendant's profile. We turn now to that matter.

C. EQUIVALENT DNA QUANTITY BASED ON BAND INTENSITIES

The admissibility of the D2 autorad evidence hinges on the People's remaining argument that the relative intensities of the bands on the autorads establish that the D2 mixture contains two heterozygous individuals (whom the People again improperly refer to as "appellant and the victim"). In other words, this evidence must prove the preliminary fact that the perpetrator is heterozygous at the D2 locus.

The People offer two related propositions to explain that the two bands each contain two alleles. The first states that because the two D2 bands appear to be the *same intensity* they each contain the *same amount* of DNA -- two alleles each (i.e., the four alleles are divided evenly). The second proposition states that because the D2 bands appear to be *twice the intensity* of the four-band mixture bands (which contain one allele each), they contain *twice the amount* of DNA -- two alleles each.

1. Prosecution Witnesses

a. Sensabaugh

Sensabaugh stated that heavier and broader bands are an indication of the quantity of DNA. When bands in a four-band mixture have different intensities, it may be possible to infer which two bands come from one person (i.e., the intensity of two of the four bands may match, and the intensity of the other two may match). But in the case of

the D2 two-band mixture, the bands give no clue which bands go together or whether they come from a male or female.

b.   Chakraborty

Chakraborty stated generally that band intensity is affected by DNA quantity.

c.   Adams

Adams did not believe there was sufficient reason to exclude the D2 autorad from the frequency calculation "based on the totality of the results." He explained, in reference to comparisons between the autorads, that relative band intensities can reveal information about DNA quantity. He explained that the other autorads clearly demonstrate there is a mixture of two people's DNA in the perpetrator's lanes because there are four bands of equal concentrations. Because the D2 mixture shows only two bands, "about double in strength" (compared to the bands in the four-band mixtures on the D1 and D4 autorads), he concluded two people's alleles are present, "but at the same locations." The D2 mixture bands appear twice as intense as the single-allele bands in the four-band mixtures on the D1 and D4 autorads and therefore they contain twice the DNA (two alleles each). (See fig. 47.)
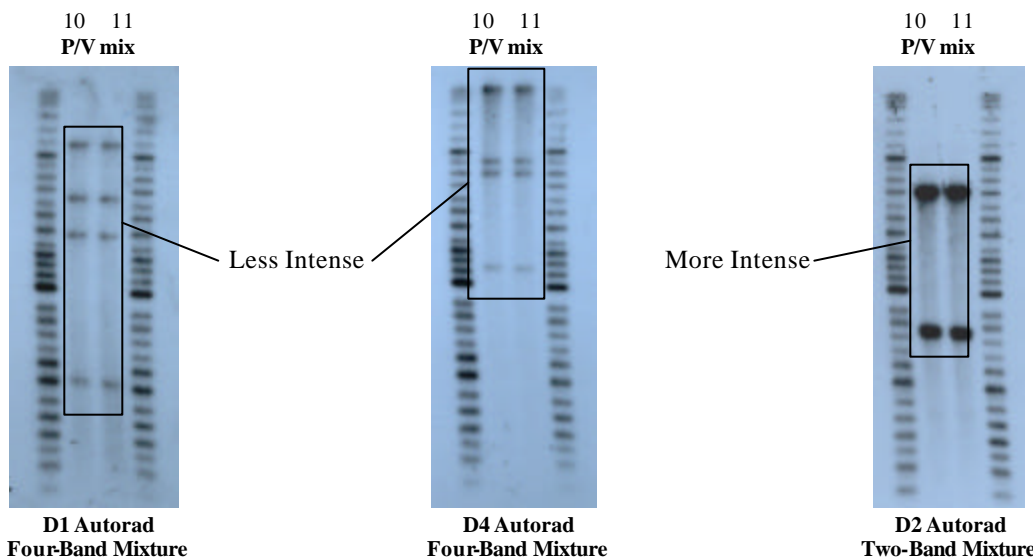
Fig. 47. *Band Intensity Comparison Between Perpetrator/Victim Mixture Bands On Different Autorads.*
NOTE: Adams testified the D2 bands are twice as intense as the D1 and D4 bands.

2. Defense Witnesses

a. *Zabell*

Zabell stated it is "very risky business" to make inferences regarding DNA quantity from band intensities, although a sharp difference in intensity may give a hint as to quantity.

b. *Bakken*

Bakken testified it is an invalid argument to say that both the victim and perpetrator are heterozygous (see fig. 43(1), *ante*) based on relative band intensities. Experience teaches that this prediction cannot be made. Bakken explained there are many studies that instruct against making an assessment of DNA quantity based on band intensity. The argument that the perpetrator cannot be homozygous (see fig. 43(2) & (3), *ante*) because the two D2 bands are of equal intensity is invalid and based on faulty reasoning. Bakken pointed to an example of the failure of this theory found on the D2 autorad itself. He noted that the two bands in the control lane are of differing intensities although it is known that each band contains the same amount of DNA (because the two alleles are inherited in equal proportions from the mother and father). (Fig. 48, lane 2.)
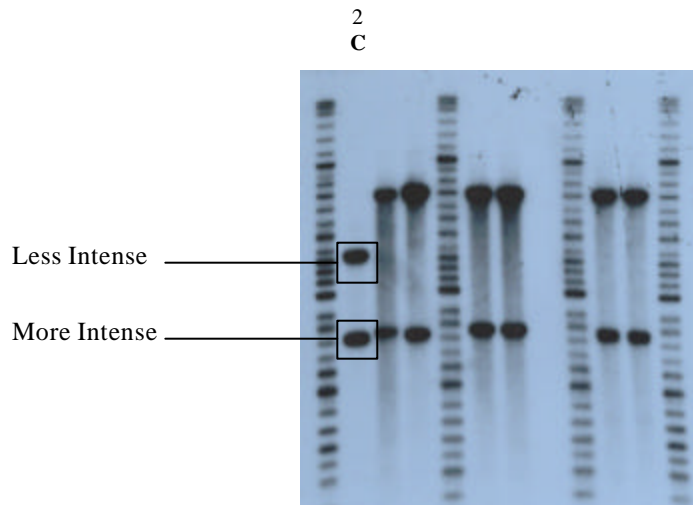
Fig. 48. *D2 Autorad.*
NOTE: Bakken testified to an intensity disparity between bands within lane 2.

125.

Another example, Bakken testified, can be seen in one of the four-band mixtures where there is a difference in the intensity between the two victim's bands, again known to contain the same amount of DNA. Bakken did not specify whether he was referring to the D1 or D4 autorad, but based on the distinctive pattern Bakken described it appears he was referring to the four-band mixture on the D4 autorad in which the top victim's band is significantly more intense than the bottom band. (Fig. 49, lanes 10 & 11.)
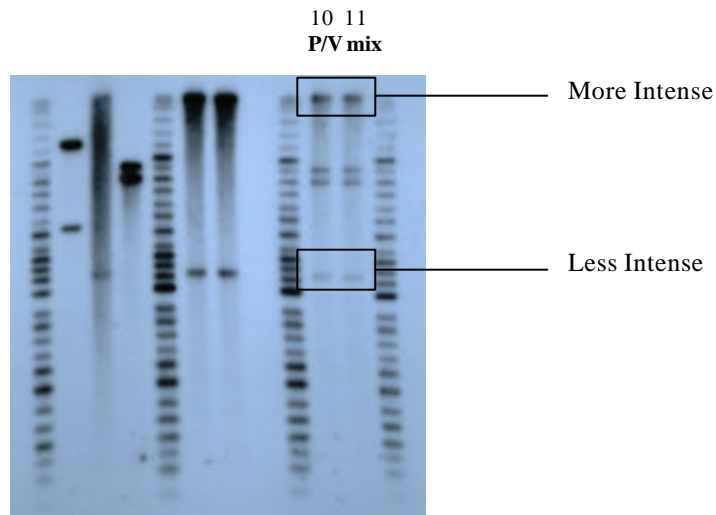
Fig. 49. *D4 Autorad.*
NOTE: Bakken testified to an intensity disparity between top band and its mate.

3.   Analysis

The FBI's D2 autorad results were admissible *only if, based on visually observable band intensities, the perpetrator's alleles are discernible from the two-band mixture.* More specifically, (1) band intensity must consistently and reliably correlate with DNA quantity and (2) visual examination must allow a reliable evaluation of superimposed or coalesced bands in a DNA mixture such that the perpetrator's alleles can be discerned from the other alleles. In our opinion, this procedure of band-intensity analysis to resolve masked bands in a superimposed mixed sample (hereafter sometimes band-intensity analysis) is subject to *Kelly* scrutiny. It must be, and has not yet been, established as a scientifically accepted procedure under *Kelly*'s first prong. And because there has been no assessment under the first prong establishing what the proper and

126.

accepted procedure actually is, the issue cannot be settled by a third-prong analysis to determine whether proper procedure was followed in this case. That is, we do not know what the proper procedure is and, until we do, we cannot decide whether it was in fact followed here. We will explain.

a.   *Applicability of Kelly*

Complicated scientific procedures must pass *Kelly* scrutiny before their results are submitted to the jury. Under *Kelly*'s first prong, the reliability of these procedures must be determined by the court, which asks: can this procedure *reliably be used* for this purpose? -- or more loosely, *should* this procedure be used for this purpose? Expert opinions responding to this question go to admissibility, *not* credibility. Such expert opinions include criticisms of the procedure as subjective, inconsistent, irreproducible, and so on. While *Kelly*'s first prong considers expert opinions regarding the procedure itself, including its theory, the third prong considers expert opinions regarding proper *use* of the procedure, including proper *interpretation* of its results. Both prongs are part of *Kelly*'s admissibility screening, not issues to be weighed by the jurors.

The *Kelly* test is required because sophisticated scientific procedures and their results are not only incomprehensible but irresistibly impressive to jurors. *Venegas* stressed that a procedure's complexity and incomprehensibility are key to the *Kelly* requirement, and that procedures "readily understandable by laypersons … need not be screened under *Kelly/Frye* before being admitted into evidence." (*People v. Venegas, supra*, 18 Cal.4th at p. 83.) In *Venegas*, the Attorney General argued that "the procedures for determining the statistical significance of a match are immune from the requirements of *Kelly/Frye*" (*People v. Venegas, supra,* 18 Cal.4th at p. 82) because the procedure "requires no more than well-established mathematical formulae such as those used to calculate the frequency of blood-group markers [citation]." (*Ibid.*) Disagreeing, the court explained that the statistical RFLP calculation is "much more complicated" than the blood marker calculation. (*Ibid.*)

127.

"*It is the very complexity of the issues* surrounding the propriety of the various recognized methods of computing RFLP probability frequencies *that draws them under the Kelly/Frye umbrella.* 'To ... leave it to jurors to assess the current scientific debate on statistical calculation as a matter of weight rather than admissibility, would stand *Kelly-Frye* on its head. We would be asking jurors to do what judges carefully avoid -- decide the substantive merits of competing scientific opinion as to the reliability of a novel method of scientific proof.... The result would be predictable. The jury would simply skip to the bottom line -- the only aspect of the process that is readily understood -- and look at the ultimate expression of match probability, without competently assessing the reliability of the process by which the laboratory got to the bottom line. This is an instance in which the method of scientific proof is so impenetrable that it would "' ... assume a posture of mystic infallibility in the eyes of a jury ....' [Citation.]" [Citation.]' [Citation.] The statistical calculation phase of RFLP analysis therefore requires *Kelly/Frye* screening of evidence on statistical probabilities of random matches at VNTR loci to assure that (1) the methodology used is generally accepted in the scientific community, and (2) the calculations in the particular case followed correct scientific procedures." (*People v. Venegas, supra*, 18 Cal.4th at pp. 83-84, italics added.)

Similarly, the propriety of band-intensity analysis is a complicated issue beyond the understanding of laypersons. It requires an understanding of genetic principles, knowledge and experience in molecular biology methods, particularly electrophoresis and autoradiography, and a trained eye for reading subtle variations on X-ray films. Lacking these, jurors are not equipped to competently consider opposing scientific opinions regarding whether the procedure is scientifically grounded, reliable, and generally accepted in the scientific community. Band-intensity analysis therefore requires independent *Kelly* scrutiny, as does the more straightforward already accepted autorad band analysis (see *People v. Venegas, supra*, 18 Cal.4th at pp. 76-79).

b.   *Kelly's First Prong*

Our analysis under *Kelly*'s first prong proceeds as follows:

1)   Has band-intensity analysis, *specifically*, already been deemed generally accepted by a published appellate opinion?

2)   If so, under *Venegas*, the trial court could properly rely upon that opinion as precedent to satisfy the first prong.

3)   If not, has another similar procedure -- which is not materially distinct from band-intensity analysis -- already been deemed generally accepted by a published appellate opinion?

4)   If so, under *Venegas*, the trial court could properly rely upon that opinion as precedent to satisfy the first prong.

5)   If not, band-intensity analysis has not been deemed generally accepted and the trial court was required to conduct a thorough hearing on that matter before admitting the D2 autorad evidence.

Accordingly, our first question is this: has band-intensity analysis already been deemed generally accepted? We look to the case law to see whether an opinion has set a precedent, assuming precedent can be so established, for the general acceptance of band-intensity analysis. Because we find no opinion addressing band-intensity analysis specifically, we look to see whether any opinions address similar procedures, and whether those procedures are materially distinct from band-intensity analysis. (*People v. Venegas, supra,* 18 Cal.4th at p. 53.)

c.   *Already Accepted Procedure*

As *Venegas* concluded, *Axell* and *Barney* have established the acceptance of the *basic* RFLP procedure. Thus, we must determine exactly what the basic RFLP procedure described in those cases entails, and whether band-intensity analysis is effectively the same procedure or, instead, a materially distinct procedure.

Venegas explained that *Axell* established general acceptance of the basic RFLP steps to: (1) "extract DNA from evidentiary samples"; (2) "generate autorad displays of bands indicating sizes of DNA fragments"; (3) "compare those bands with one another and declare a match"; and (4) "make statistical calculations of the frequencies of the matched bands in a population database." (*People v. Venegas, supra*, 18 Cal.4th at pp. 76-77.) Similarly, *Venegas* noted that *Barney* approved the general acceptance of

"the basic procedures applied to compare and match bands depicted on the autorads."

(*Id.* at p. 79.) *Venegas* concluded that:

> "for purposes of the trial of this case, the *Axell* and *Barney* opinions clearly
> established the general scientific acceptance, under Kelly's first prong, of
> the basic RFLP methodology utilized by the FBI in (1) producing autorads
> with bands reflecting the base-pair sizes of forensic samples at particular
> DNA locations, and (2) comparing the bands in order to determine whether
> the samples matched at those locations." (*Id.* at p. 79.)

Axell itself explained the relevant RFLP steps as follows:

> "… (6) autoradiography in which a film is developed on top of the
> nylon membrane, revealing the location of the DNA by bands on the X-ray
> film, called an autoradiogram or autorad.... [¶] The autorads must be
> interpreted and the bands produced by the migration of DNA in the gel in
> different lanes examined to ascertain if they match. [(7)] Essentially the
> bands on the autorad from the victim's, suspect's, and crime scene evidence
> samples are 'eyeballed' to see if they match within a certain measurement.
> [(8)] If a match is declared, the likelihood that a match is unique must be
> determined." (*People v. Axell, supra*, 235 Cal.App.3d at p. 846.)

In *Barney,* the court explained "[t]here are three discrete steps in DNA analysis as
performed by the FBI … and by Cellmark … : (1) *processing* of DNA from the suspect
and the crime scene to produce X-ray films which indicate the lengths of the polymorphic
fragments; (2) examination of the films to determine whether any sets of fragments
match; and (3) if there is a match, determination of the match's *statistical significance*."
(*People v. Barney, supra*, 8 Cal.App.4th at p. 806.) The court concluded *Axell* served as
precedent for the general acceptance of the DNA processing step and the matching step,
but not the statistical analysis step (*People v. Barney, supra*, 8 Cal.App.4th at p. 806),
which the court found was currently under debate in the scientific community (*id.* at pp.
819-821).

Barney summarized the substeps of the DNA processing step: (1) extraction,
(2) restriction, (3) electrophoresis, (4) Southern transfer and denaturing, (5) hybridization,
and (6) autoradiography. (*Id.* at pp. 806-807.) The court explained the last two substeps
as follows:

"5. Hybridization

"The last two substeps enable visualization of the lengths of the sample DNA fragments by producing X-ray films which show the distance the fragments traveled as a result of electrophoresis…. [¶] … [¶]

"6. Autoradiography [¶] … [¶]

"The location of a band on the X-ray film indicates the distance a fragment traveled as a result of electrophoresis, and hence the length of the fragment. The size-marker fragments also appear on the films, enabling measurement of the base-pair lengths of the sample fragments.

"… The bands are arrayed in varying positions, which indicate the distance the selected DNA fragments traveled during electrophoresis and hence the various lengths of the fragments." (*Id*. at pp. 807-808.)

From these three cases, we gather the following statements of accepted procedure

for discerning bands and identifying alleles from autorads:

- production of "autorad *displays* of bands" that "*indicat[e] sizes of DNA fragments*" (*People v. Venegas, supra*, 18 Cal.4th at pp. 76-77, italics added);

- production of autorads to which "*basic procedures* .. to compare and match *bands depicted on the autorads*" can be applied (*id*. at p. 79, italics added);

- production of "autorads with bands *reflecting the base-pair sizes*" of the DNA fragments (*ibid*., italics added);

- production of autorads "*revealing the location of the DNA by bands*" (*People v. Axell, supra*, 235 Cal.App.3d at p. 846, italics added);

- production of autorads that "*indicate the lengths of the [DNA] fragments*" (*People v. Barney, supra*, 8 Cal.App.4th at p. 806, italics added);

- production of autorads that "*show the distance the fragments traveled*" and thus "*enable visualization of the lengths* of the sample DNA fragments" (*id*. at p. 807, italics added);

- production of autorads "with bands arrayed in varying positions, which *indicated the distance* the selected DNA fragments traveled … and hence *the various lengths of the fragments*" (*id*. at p. 808, italics added).

131.

It is apparent that *Venegas*, *Axell*, and *Barney* address the typical cases in which the "basic" procedure is adequate -- the cases in which the autorads do indeed display and depict the perpetrator's bands, and indicate, reflect, and reveal the locations/sizes of the perpetrator's alleles. When the sample is not mixed, the perpetrator's one or two bands can readily be discerned because they are the only bands in the perpetrator's lane. Even when the sample is mixed, there are usually four bands from which the perpetrator's two bands can readily be discerned, as on the D1 and D4 autorads. In the typical cases, the locations of the perpetrator's bands are readily apparent and the sizes of the alleles can be determined from the size standards using "basic procedures" (*People v. Venegas, supra*, 18 Cal.4th at p. 79). Each band accounts for one allele and band locations reveal allele sizes. We believe these are the situations for which these three opinions serve as precedent for the general acceptance of discerning bands from autorads.

d.   *Material Scientific Distinction*

In our opinion, band-intensity analysis constitutes a materially distinct procedure for discerning the perpetrator's alleles from an autorad, not merely an immaterial variation on the accepted basic autorad analysis approved by *Venegas*, *Axell*, and *Barney*. As the Supreme Court's decisions have confirmed, materially distinct approaches to the same general purpose must independently pass *Kelly*'s first-prong scrutiny. In *Venegas*, the court deemed accepted the modified ceiling approach to determining the statistical significance of a match. (*People v. Venegas, supra*, 18 Cal.4th at pp. 84-90.) Then, in *Soto*, the court separately examined and deemed accepted the unmodified product rule approach. (*People v. Soto, supra,* 21 Cal.4th at pp. 518-519.) Both procedures are approaches to the same general purpose -- the statistical probability calculation -- but they address different theoretical concerns and can produce significantly different results.

The accepted autorad analysis addressed by *Venegas*, *Axell*, and *Barney* compares the locations of the perpetrator's *displayed* bands to the locations of the size standard bands to determine the sizes of the perpetrator's alleles. This procedure in fact involves

very little subjectivity or interpretation. As *Axell* and *Barney* determined, "'interpretation of bands on an autorad is fairly straightforward and involves a minimal amount of subjective analysis.' [Citation.]" (*People v. Barney, supra*, 8 Cal.App.4th at pp. 813-814.) On the other hand, *visual* resolution of a superimposed mixture using band-intensity analysis is not a straightforward, objective comparison of band locations to determine allele sizes. Unlike the accepted autorad interpretation procedure of *Venegas*, *Axell*, and *Barney*, band-intensity analysis addresses the anomalous situation in which the alleles in a mixture are *superimposed* into only two or three bands, all the bands are *not* displayed or depicted, and those that are do not by their presence indicate, reflect, or reveal the size of the perpetrator's alleles. The locations of the bands are entirely inadequate to permit determination of the perpetrator's alleles: there are too few bands to account for all four alleles, some of which are masked by others.[76] Band-intensity analysis is a subjective visual evaluation of subtle variations between bands to discern the alleles from a mixture that contains too few bands to yield readily discernible results. Furthermore, use of band-intensity analysis can significantly affect the resulting statistical calculation. We think *Venegas*, *Axell*, and *Barney* plainly do not speak to this methodology, and therefore do not encompass band-intensity analysis in the procedure they deem generally accepted. Band-intensity analysis of superimposed mixtures is a separate and distinct procedure for interpreting autorad bands and it must therefore independently "pass[] muster under the central first prong of the *Kelly* test." (*People v. Venegas, supra*, 18 Cal.4th at p. 81.) Because it has not, the evidence produced by that procedure -- the FBI's conclusion that the D2 autorad reveals a heterozygous perpetrator whose genotype matches defendant's -- was inadmissible.

---

[76]   This situation is very different than when the perpetrator's sample contains only *one* band but the sample is not mixed.

e.   *Lack of Evidentiary Foundation*

Although we express no opinion as to whether band-intensity analysis is in fact generally accepted by the scientific community,[77] we note that the evidence in this case appears to present instances in which band intensity does not correlate with DNA quantity. At the *Kelly* hearing, the prosecution presented testimony that relative band intensities can correlate with DNA quantity and that the intensities of the bands in the two-band D2 mixture appear to be approximately twice as strong as those in the four-band mixtures; thus the D2 bands must contain two alleles each. The defense, however, presented strong evidence that band intensity does not reliably and consistently correlate with DNA quantity. Bakken pointed to instances in this case in which two bands on the same autorad, expected to contain the same quantity of DNA, display significantly different intensities -- the control bands on the D2 autorad, and the victim's bands in the four-band mixture on the D4 autorad.[78] (See figs. 48 & 49, *ante*.) Figures 50 through 54 illustrate other possible inconsistencies.

---

[77]   We simply find an inadequate evidentiary showing here.

[78]   Again, we assume Bakken was referring to the D4 autorad.

(1)   the defendant's bands on the D2 autorad:
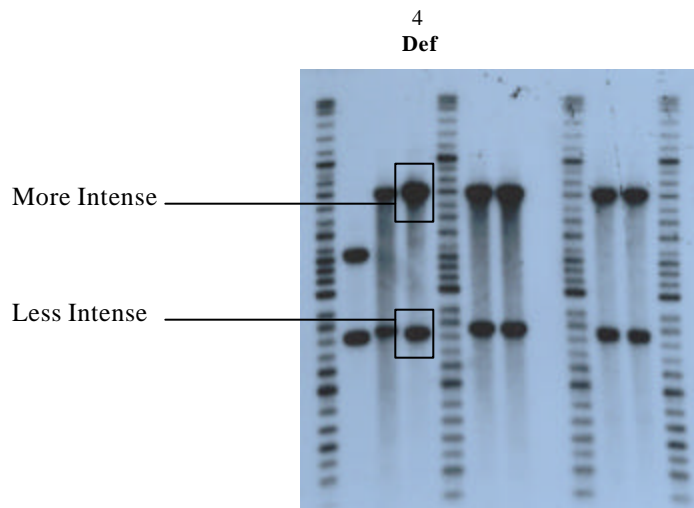
Fig. 50. *D2 Autorad.*
NOTE: Intensity disparity between bands within lane 4.

(2)   the control bands on the D4 autorad:
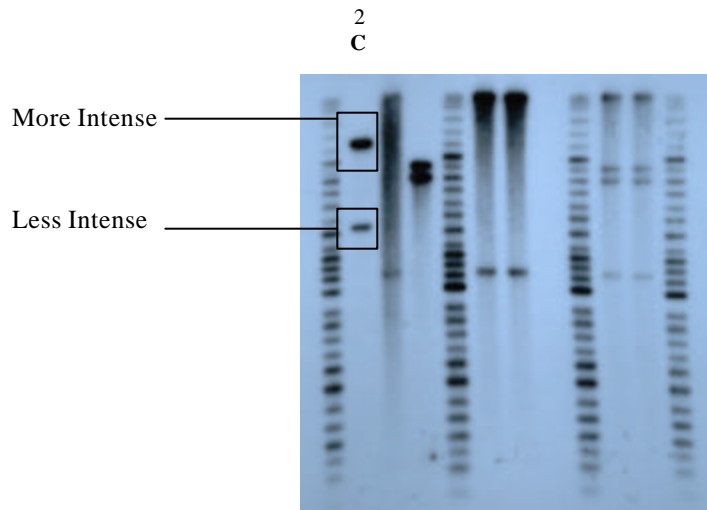
Fig. 51. *D4 Autorad.*
NOTE: Intensity disparity between bands within lane 2.

(3)   the victim's (evidentiary) bands on the D4 autorad:
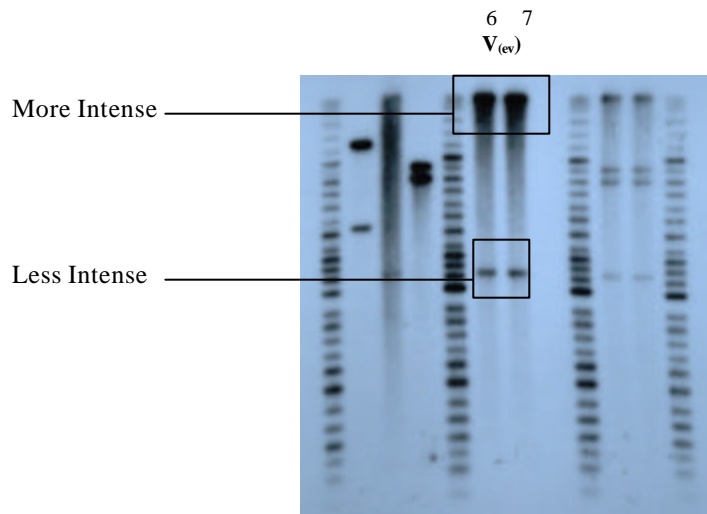
Fig. 52. *D4 Autorad.*
NOTE: Intensity disparity between bands within lanes 6 and 7.

(4)   the victim's bands on the D1 autorad:
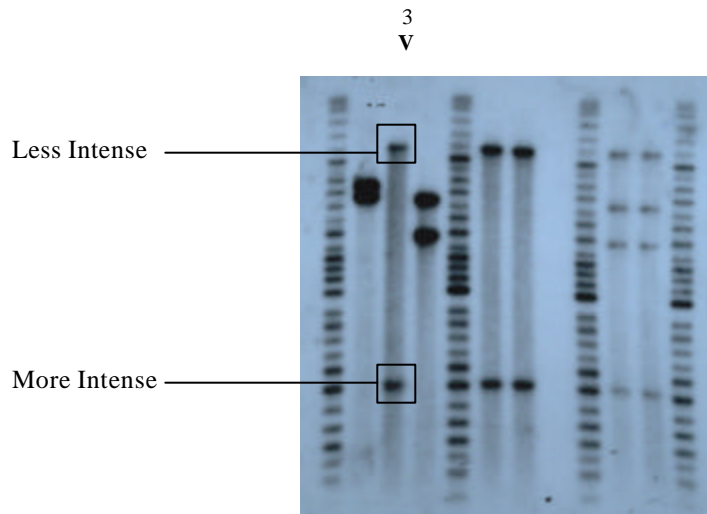
Fig. 53. *D1 Autorad.*
NOTE: Intensity disparity between bands within lane 3.

(5)   the defendant's bands on the D1 autorad:
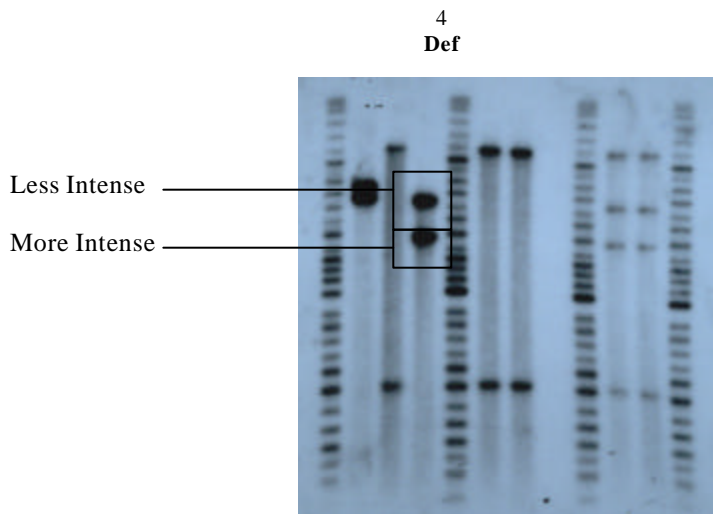
Fig. 54. *D1 Autorad.*
NOTE: Intensity disparity between bands within lane 4.

Furthermore, the comparison *between* autorads that prosecution witness Adams

testified showed a correlation between DNA quantity and band intensity (see fig. 47,

ante) does not necessarily find further support in the evidence. For example, the victim's

bands on the D2 autorad are far more intense than the victim's bands on the D1 and D4

autorads, yet the D2 victim's bands are not expected to contain twice as much DNA as

the D1 and D4 victim's bands. (Fig. 55.)
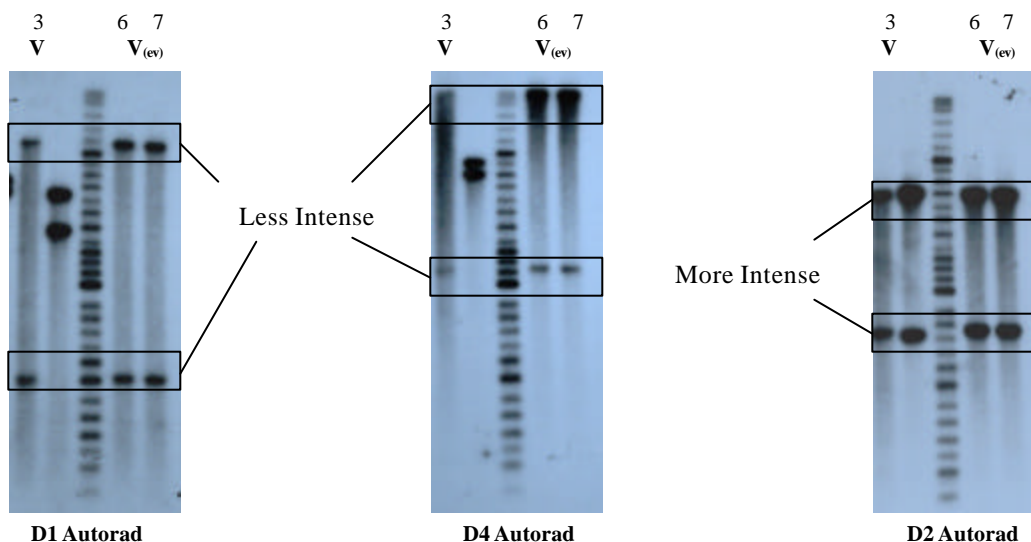
Fig. 55. *Band Intensity Comparison Between Victim's Bands on Different Autorads.*
NOTE: The D2 victim's bands are more intense than most of the corresponding victim's bands on the D1 and D4 autorads,
but there is no evidence the D2 bands contain twice as much DNA as the bands on the other autorads.

In addition, the stronger intensity of the D2 autorad bands may be due to the fact that the D2 probe was the first probe hybridized to the membrane. Sequential probing of the membrane gradually washes some of the DNA from the membrane, and thus later hybridizations may produce less intense results than earlier ones. It appears that the D2 probe was hybridized to the membrane first, followed by D17 (the autorads were inconclusive), then D1, and finally D4. Of course, there may also be other factors that influence band intensity differences between autorads. The point is that Adams's theory, which was based only on the comparison of the two perpetrator/victim lanes between the autorads, does not necessarily hold true for comparisons of the other five lanes.

We note that NRCI states: "Mixed samples can be very difficult to interpret, because the components can be present in different quantities and states of degradation. It is important to examine the results of multiple RFLPs, as a consistency check. Typically, it will be impossible to distinguish the individual genotypes of each contributor." (NRCI, *supra*, at p. 59.) "Mixed samples are a reality of the forensic world that must be accommodated in interpretation and reconstruction. As a rule, mixed samples must be interpreted with great caution…. Interpretations based on quantity can be particularly problematic -- e.g., if one saw two alleles of strong intensity and two of weak intensity, it would be improper to assign the first pair to one contributor and the second pair to a second contributor, unless it had been firmly established that the system was quantitatively faithful under the conditions used." (*Id.* at p. 66.) NRCII states: "In some cases, it might be possible to distinguish the genetic profiles of the contributors to a mixture from differences in intensities of bands in an RFLP pattern or dots in a dot-blot typing …." (NRCII, *supra*, at p. 129.) Modern Scientific Evidence: The Law and Science of Expert Testimony (2001) (hereafter Modern Scientific Evidence) states: "Studies in which DNA from different individuals is combined in differing proportions show that the intensity of the bands reflects the proportions of the mixture. Thus, if bands in a crime-scene sample have different intensities, it may be possible to assign

alleles to major and minor contributors. However, if the bands are present in roughly equal proportions, this allocation cannot be made, and the statistical interpretation of the observed results must include all possible combinations." (*Id.* at § 25-2.4.3, fn. 93.)

f. *Risk of Overlooking First-Prong Issues*

We pause to comment on an error likely to befall unwary courts. Although *Venegas* stressed the risk of mistaking third-prong issues for first-prong issues, we are also apprehensive of the converse problem -- mistaking first-prong issues for third-prong issues. First, courts may overlook the distinctness of a new procedure, believing it is merely an immaterial variation on an already accepted procedure (i.e., it is the same procedure). Second, courts may assume that a truly distinct procedure is merely one method for performing a more general, already accepted procedure (i.e., it is an implemental procedure). In both cases, the trial court, believing the procedure has already been deemed accepted, will erroneously perform only a third-prong analysis. Although both the first and third prong tests go to admissibility, as we have explained, the standards for admissibility are very different, and perhaps more importantly, the standards on review are very different.

1.   Same Procedure

Venegas determined that an already accepted procedure serves as precedent for the acceptance of a second procedure unless the defendant can prove the second procedure is materially distinct. For example, if the prosecution presents RFLP autorads produced by the FBI, *Axell* serves as precedent for the general acceptance of the RFLP procedure to produce those autorads unless the defendant shows that differences in the FBI's procedure make it materially distinct from the procedure approved in *Axell*. (*People v. Venegas, supra*, 18 Cal.4th at pp. 53-54.) There is, in effect, a presumption that the procedures are the same, and the defendant bears the burden of demonstrating that they are not.

If the defendant shows that the differences are significant enough to render the procedure materially distinct from the accepted procedure, the procedure must be analyzed under *Kelly*'s first prong. If, on the other hand, the defendant does not show that the differences make the procedure materially distinct from the already approved procedure, then the procedures are the same and the differences go to whether the proper procedure was followed in the particular case. (*People v. Venegas, supra*, 18 Cal.4th at p. 78.) If the differences amount to a failure to follow proper procedure, the evidence is inadmissible under *Kelly*'s third prong.

Obviously, there is a danger that courts may neglect or misunderstand which differences nudge a procedure into material distinctness. How different is distinct? In the continuum of what can be defined as differences in procedure, there inevitably comes a point at which the differences are dramatic enough to transform the procedure into a distinct procedure. A court that fails to recognize this transformation will conduct an inappropriate third-prong analysis where a first-prong analysis is proper.

2. Implemental Procedure

A second risk is that courts may construe a truly distinct procedure as merely one of a number of alternate methods for implementing or accomplishing a more general, already accepted procedure -- for instance, band intensity analysis as one method of accomplishing autorad analysis. When an accepted procedure is stated in broad terms as a general principle or step, courts may be tempted to assume that the acceptance of that procedure carries on its coat-tails all the methods of accomplishing it, and to assume that those methods comply with the general principle or step purely because they accomplish it. However, every conceivable procedure accomplishes a more general principle or step, and courts, liberated by this logic, could find that hundreds of highly sophisticated procedures are simply different methods of performing a single accepted procedure -- and again *Kelly*'s first prong would be handily eviscerated. Although these procedures would still be required to survive the scrutiny of *Kelly*'s third prong, every procedure could

satisfy the test under this perversion since every procedure could be said to comply with the general principle or step.

Courts therefore must be aware that acceptance of a general scientific principle or procedure does not automatically provide acceptance of every technical method for implementing that principle or procedure. *Venegas* supports the view that every distinct procedure, whether general or technical, must pass the first-prong test. There, the court examined the modified ceiling approach, which is one of several methods for accomplishing the general, already accepted step of calculating statistical probability. The court did not confer on the new procedure a passive surrogate acceptance from the accepted general procedure, but held the new procedure up to first-prong scrutiny. (*People v. Venegas, supra*, 18 Cal.4th at pp. 85-86.)

3. Appellate Error

Of course, if the appellate court also overlooks first-prong issues, it compounds the trial court's error. For example, the trial court, erroneously applying the third prong, may find the evidence admissible because it believes the procedure was in compliance with the accepted procedure. On review, the appellate court can rectify the mistake only if it recognizes that the first-prong test should have been applied originally. If, however, the appellate court labors under the same misconception as the trial court, the appellate court applies an abuse of discretion standard, and affirms the trial court's ruling if there is evidence to support it. Thus, the evidence slips by, its reliability unscrutinized. Only if the evidence fails to support the ruling will the appellate court reverse. Similarly, if the trial court correctly applies the first prong but incorrectly determines the procedure is accepted, the mistake goes uncorrected if the appellate court believes the issue is strictly a third-prong issue and reviews it as such.

Assume, for example, that we erroneously believed band-intensity analysis is a third-prong issue. We would review the trial court's finding that correct procedures were followed for abuse of discretion.[79] Under that test, we could reverse only if we concluded the finding was arbitrary, capricious, absurd, or outside the bounds of reason. Looking to the evidence, we would see the prosecution presented testimony that band-intensity analysis can and does reveal the perpetrator's alleles from a superimposed mixture. In opposition, we would find defense testimony that band-intensity analysis cannot be relied upon for this purpose.

Review under the third-prong discretionary standard could play out in two ways, depending on the state of the evidence. First, assume that, as in this case, there was evidence demonstrating to us the lack of correlation between band intensity and DNA quantity -- the defense testimony pointed out failings apparent from the autorads themselves. Even though the trial court was apparently unconvinced by the evidence, we would conclude the band-intensity analysis theory does not hold up because the autorads contain several instances of its failure, where approximately equal amounts of DNA do not produce approximately equal band intensities. We would conclude the defense presented evidence that the proper procedures were not followed in this case, and the prosecution's evidence in opposition to the defense testimony was insubstantial and founded on assumptions proven invalid by the evidence. For these reasons, we would hold the trial court's reliance on this evidence unreasonable and an abuse of discretion.

---

[79] In this case, we presume the trial court applied both first-prong and third-prong tests to band-intensity analysis. Although the court's ruling did not mention band-intensity analysis, the court found there was general acceptance of the FBI's procedure and held the evidence admissible. We presume all findings necessary to support the trial court's ruling. (*Denham v. Superior Court* (1970) 2 Cal.3d 557, 564 ["[a]ll intendments and presumptions are indulged to support [the judgment or order] on matters as to which the record is silent...."].)

Although in this scenario we, like the trial court, failed to recognize the issue as a first-prong issue, we would nevertheless (but only fortuitously) find the evidence inadmissible under the third prong.

Alternatively, assume that there was no such evidence; the *only* evidence on the issue was the two opposing expert views. Nothing in the evidence would suggest to us that the trial court was unreasonable in finding that band intensity could be relied upon to determine DNA quantity or the presence of particular alleles, and we would be compelled to find that the trial court reasonably relied on the prosecution testimony that band-intensity analysis allowed the discernment of the perpetrator's alleles from the mixture. The unreasonableness of that reliance would not be apparent to either the trial court or this court. We would have neither the suspicion nor the authority to find an abuse of discretion, and we would uphold the admissibility of the evidence.

This second scenario emphasizes why the first-prong analysis is so critical to the screening of scientific evidence. Without first-prong inspection, unreliable evidence can sneak into the trial and even survive appellate review. The first prong not only presents a more rigorous standard for admission in the trial court, but it also allows the reviewing court an opportunity to independently evaluate and ensure the reliability of the evidence. In the first scenario, as in the present case, the procedure's unreliability might have been apparent from the evidence. But in the many cases where it is not, the admission of such unreliable evidence would be affirmed by the appellate court, unaware of its unreliability and powerless to rectify its improper use.

In *State v. Harvey* (1997) 699 A.2d 596, the defendant challenged the reliability of a somewhat similar procedure, dot-intensity analysis, to analyze a mixed DNA sample on a dot blot (not an autorad). The majority concluded dot-intensity analysis was generally accepted. (*Id.* at pp. 624-629.) The dissenting judge articulated some of our concerns, as follows:

"The principal disagreement that I have with the majority concerns the general acceptance of dot-intensity testing. Dot-intensity analysis was the essential evidence relied upon by the State to demonstrate that defendant was in all likelihood the actual person whose blood contributed to the mixed sample found at the scene. The majority properly, if reluctantly, recognizes that dot-intensity testing, as a scientific method, must meet the standard of general acceptance even if DQ-Alpha and polymarker testing are themselves found to be generally accepted scientific tests. The majority, however, misconstrues the distinctive and distinguishing features of dot-intensity testing as a method of analyzing DNA, denigrates many of defendant's challenges to the testing as not going to the reliability of the procedure, but rather only to its weight, and then, on an embarrassingly deficient record, summarily concludes that the novel scientific procedure passes muster under our long-standing precedent. Dot-intensity analysis as used here -- a procedure never before used in any court case, successfully documented in any laboratory, or validated in any scientific study or published literature -- has not been shown to be an established and reliable procedure. Further, no foundation for dot-intensity analysis exists in the record, and the results obtained clearly show that such evidence is grossly unreliable. Finally, the analysis rests on a combination of assumptions that renders the evidence so unpersuasive and speculative that it is inadmissible under New Jersey Rule of Evidence 402." (*State v. Harvey, supra*, 699 A.2d at p. 658, Handler, J. diss. opn.)

"The polymarker and DQ-Alpha testing kits were designed solely to determine the presence or absence of certain alleles. Dot-intensity analysis, however, purports to determine more. It purports to quantify the alleles that are present and thereby to identify the specific alleles contributed by each donor to the DNA mixture. The majority only grudgingly rejects the State's argument that dot-intensity analysis is nothing new and that no independent basis for its admission need be established. Without discussion, it recognizes, without really appreciating, that that difference requires an independent foundation for admissibility. [Citation.] Notwithstanding its concession, the majority then erroneously devalues and mischaracterizes defendant's challenges to the evidence -- challenges to its competency -- as merely going to Cellmark's performance of the polymarker test ....[80] [Citation.] That conclusion derives from a distortion of defendant's claims and from a serious misunderstanding of the

---

80 Apparently, under the New Jersey court's three-prong test, the third prong goes to weight, not admissibility as it does in California.

144.

distinctive nature and purposes of dot-intensity analysis." (*Id.* at pp. 658-659.)

"The issue here is not whether the reverse dot-blots obtained on the polymarker strips can reveal the presence of alleles in the mixture -- they can. At issue is whether an interpretation made of those strips that goes beyond what results that the strips were designed to show -- the presence of alleles -- is generally accepted as scientific evidence. [Citation.] Thus, unlike 'an expert's ability to perceive an abnormality on an x-ray,' which concededly 'is a matter within the province of the jury,' [citation] here we must decide, by analogy, whether a doctor's interpretation of an x-ray can be admitted without restrictions when he testifies to a condition that the x-ray was not designed to reveal. Therefore, while a doctor's diagnosis of a broken bone from an x-ray may be admissible because it is based on a generally accepted interpretation of a generally accepted test, the doctor's diagnosis of cancer from that same x-ray ought not to be admitted unless and until the doctor can establish that such a diagnosis from an x-ray is generally accepted." (*Id.* at pp. 659-660, fn. omitted.)

"Not only do the results obtained here establish the gross unreliability of this evidence, but the entire practice of visualizing and weighing dot intensities to determine the makeup of a mixture is unavoidably subjective. A subjective test, especially one that is immune from later challenge, should not be admissible evidence in these circumstances. The standard for the admissibility of scientific evidence is designed to ensure that the testing procedure 'relies primarily upon objective factors for reaching a conclusion, with subjective factors playing only a minimal role in the analysis.' [Citation.]" (*Id.* at p. 670.)

" … A full hearing on the assumptions and the entire validity of the dot-intensity analysis should have been held. That hearing was necessary to explore the inconsistencies in both the State's experts' comments and in the actual results obtained. The uncritical admission of this evidence … without even remotely establishing its validity is an egregious wrong." (*Id.* at p. 672.)

D. CONCLUSION

Admission of the D2 autorad evidence required a preliminary fact determination of the *perpetrator's genotype* at the D2 locus to establish the relevance of the D2 autorad evidence. Reference to *defendant's* genotype to prove the perpetrator's genotype was improper; use of band-intensity analysis to prove the perpetrator's genotype required *Kelly* scrutiny of that method. Thus, the preliminary fact of the perpetrator's genotype at

the D2 locus was not sufficiently proved and, as a result, the D2 autorad evidence was irrelevant and inadmissible. We conclude that the trial court abused its discretion by finding sufficient evidence of the preliminary fact that the perpetrator is heterozygous at the D2 locus and by failing to find use of the D2 autorad results improper scientific procedure under *Kelly*.

The erroneous admission of the D2 autorad evidence had potentially grave consequences. The unscrutinized band-intensity analysis permitted the FBI to interpret the two-band D2 mixture and to conclude that the perpetrator's genotype is *heterozygous*, as opposed to either of the two homozygous possibilities (which would have proved defendant's innocence). That conclusion, in turn, had a critical effect on the evidence against defendant. The frequency of the heterozygous D2 genotype was multiplied by the frequencies of the D1 and D4 genotypes to obtain the perpetrator's overall profile frequency. Including the D2 frequency in this calculation made the overall profile frequency more rare in the population and made defendant's possession of it more incriminating. Because the D2 autorad evidence was not admissible, it should not have been included in the calculation and the overall profile frequency should have been based only on the D1 and D4 autorad evidence. The resulting profile frequency would have been more common and less incriminating to defendant.

A *Kelly* hearing on band-intensity analysis would have ensured against this outcome. If the trial court had found band-intensity analysis unaccepted or improperly performed, it would have excluded the D2 autorad evidence and the frequency of that locus would not have been multiplied into the overall profile frequency. If, on the other hand, the trial court had found band-intensity analysis accepted and properly performed, it would have admitted the D2 autorad evidence and the frequency would have been included in the overall profile frequency. Even in this situation, the *Kelly* hearing would have served another important but often overlooked purpose -- it would have defined and focused the scientific and legal issues for the attorneys and the trial court, affecting the

manner in which evidence would have been presented at trial. The thorough examination required for a *Kelly* hearing would have resulted in a greater understanding of these complicated issues and would have promoted challenges to the evidence. The trial court's *Kelly* ruling of admissibility would not have precluded the defense from challenging band-intensity analysis and the D2 autorad results before the jury at trial. Defense counsel would have presented experts to challenge the method and to explain to the jury that, if band-intensity analysis is in fact not reliable or was in fact not properly performed in this case and thus the D2 autorad cannot reliably establish the perpetrator's genotype as heterozygous, then two of the three possible interpretations of the D2 autorad would actually *exonerate* defendant. This information would have allowed the jurors to better weigh the value of the evidence. In this case, the jurors heard nothing regarding band-intensity analysis and the possible interpretations of the D2 autorad evidence; they were simply given the overall profile frequency.

It is true, of course, that the defense could have mounted such a challenge even though the *Kelly* hearing on band-intensity analysis did not occur; but, realistically, the attorneys and the court would likely have overlooked these esoteric issues, or at least their importance, in the absence of a hearing to expose and clarify them.

Although we have the prerogative to independently consider and render a decision on whether band-intensity analysis has gained general acceptance in the relevant scientific community, we decline to do so pending full and complete litigation of that issue, assisted by live expert witnesses, in the trial court. (*People v. Leahy, supra*, 8 Cal.4th at pp. 609-610; see also see also *Cramer v. Morrison* (1979) 88 Cal.App.3d 873, 888 [general acceptance of HLA paternity testing].) On retrial, the autorads may be re-examined by scientists at the FBI or another institution. The trial court must then conduct a thorough *Kelly* hearing, at which the prosecution must establish that the perpetrator's alleles can be discerned reliably from the perpetrator/victim mixture on the D2 autorad. If the method used to discern the perpetrator's alleles has not yet passed first-prong

scrutiny (like band-intensity analysis), the court must determine, based on expert testimony and scientific literature, the reliability and general scientific acceptance of that method under *Kelly*'s first prong. If the trial court deems the method generally accepted as a reliable method for discerning alleles from a superimposed mixture on an autorad, then the court will hear third-prong testimony regarding whether the mixture on the D2 autorad in this case was properly analyzed and interpreted according to that method. If the method used to discern the alleles in the mixture is not reliable and generally accepted, or if the testing in this case fails to follow proper procedure, then the D2 autorad evidence cannot be used to calculate the profile frequency -- which will then be based only on evidence from the other autorads.

VIII. STATISTICAL WINDOW

Defendant raises two contentions, both under *Kelly*'s third prong, regarding the FBI's statistical window. First, he claims the FBI failed to follow proper scientific procedure when it used a $\pm$ 2.5% window to determine allele frequencies from the database frequency table because this statistical window was smaller than the match window. Second, he argues the statistical window was erroneously centered on the average of the perpetrator's and defendant's alleles, rather than solely on the perpetrator's allele.

We find the trial court erred when it failed to rule, based on the evidence presented at the *Kelly* hearing, that these two procedures amounted to incorrect scientific procedure for calculating the statistical probability of the perpetrator's profile.

A.   *PROSECUTION TESTIMONY*

1.   **Sensabaugh**

Sensabaugh explained that, by repeated empirical testing, the FBI determined its RFLP measurement tolerance (i.e., uncertainty window) to be $\pm$ 2.5% for a total of 5%, so that two bands within this range are declared a match. Thus, the perpetrator's and defendant's alleles are found to match when their measurements are within 5% of each

other. When a match is declared, the FBI's fixed bin method begins by averaging the perpetrator's and defendant's allele measurements. Then a $\pm 2.5\%$ window is drawn around that averaged measurement. This $\pm 2.5\%$ statistical window is superimposed on the fixed bin database frequency table, and the frequency of the bin into which the window falls is assigned as the allele frequency.

Sensabaugh explained that the fixed bin method is designed to give conservative estimates that are favorable to the defendant. One conservative feature of the method is that the $\pm 2.5\%$ measurement tolerance is narrower than the fixed bin into which it falls. Thus, when the bin's frequency is assigned to the allele, the frequency includes extra alleles outside the $\pm 2.5\%$ statistical window but still within the bin. Sensabaugh did not think there was a controversy as to whether the fixed bin method is conservative.

The $\pm 2.5\%$ floating bin method yields less conservative (more rare and less favorable to the defendant) estimates than the fixed bin method. It simply counts the frequency of alleles that fall within $\pm 2.5\%$ of the allele measurement. Sensabaugh did not think there was a controversy regarding whether the floating bin should be $\pm 2.5\%$ or $\pm 5\%$. A laboratory can use either its match window or a larger window as the floating bin statistical window; it simply cannot use a statistical window smaller than its match window. This conclusion was prompted in part by *People v. Castro* (1989) 545 N.Y.S.2d 985, in which Lifecodes used a statistical window smaller than its match window, yielding an extraordinarily rare frequency.

Sensabaugh believed the FBI had never used floating bins.

The purpose of the fixed bin method is to determine how many people in the population would match a particular allele. For example, a 950 bp allele measurement would match a 1,000 bp allele measurement because they are within 5% of each other. By the same logic, alleles measuring up to 1,050 bp would also match the 1,000 bp measurement because they too are within 5% of each other. It is therefore nonsensical to count only the allele measurements falling within a $\pm 2.5\%$ window as matches to the

perpetrator's 1,000 bp allele; the alleles falling between 950 bp and 1,050 bp should all be counted. This is effectively a 10% window, not a 5% window. The FBI, however, uses a statistical window half the size of the 10% window. The FBI averages the two alleles to give 975 bp, then surrounds that with a ± 2.5% statistical window to include the alleles between 950 bp and 1,000 bp. Some people believe that a ± 5% statistical window rather than a ± 2.5% statistical window should be used. A ± 5% window will give a larger number (a less rare frequency that favors the defendant) because more alleles will fit in the window.

Analogizing to a height measuring system, Sensabaugh agreed that, if it were known that the perpetrator's height is between 5 feet 0 inches and 6 feet 0 inches, it would be inappropriate to count only the people who are between 5 feet 6 inches and 6 feet 0 inches when determining the percentage of the population that could be the perpetrator. If one were to count only the people in a small part of the total range, the resulting number would be wrong and would misrepresent the information. A larger statistical window could overlap more fixed bins. In fact, a ±5% window almost always overlaps two or even three fixed bins.

Sensabaugh noted that in this case the alleles were all within 2% of each other, and most were actually within 1% or less.

2.   Chakraborty

In Chakraborty's opinion, the FBI's method of calculating statistical probabilities is very conservative, resulting in overestimations of frequencies. He believed the majority of experts in his field agreed that the FBI's method of calculating statistical probability is conservative. Chakraborty had advised the FBI that its statistical protocol is too conservative and its profile frequencies should be much rarer. The genetic principles could be applied more strictly to show how uncommon a multi-locus profile is.

The FBI's fixed bin method is very conservative because (1) the fixed bins are too large, (2) bins with less than five alleles are collapsed and rebinned with a neighboring

bin, (3) the Hispanic databases are joined together into a composite database that utilizes the more conservative data from each, (4) when the $\pm 2.5\%$ statistical window overlaps two bins, the higher of the two bin frequencies is assigned to the allele, and (5) a 2p formula, rather than a p^2 formula, is applied to homozygote frequencies.

The FBI's fixed bin protocol takes the average of the perpetrator's and defendant's alleles. Then a statistical window centered on that average is applied to the frequency table. In Chakraborty's opinion, a $\pm 2.5\%$ window is the appropriate statistical window for determining frequency.

At the prosecutor's request, Chakraborty calculated various new frequencies (from the allele measurements in Pizarro's case) using a different, less conservative method than the FBI had used.[81] Chakraborty applied his method to the FBI databases[82] and the Orange County databases, which he believed would also be relevant to this case. In addition to comparing databases, his calculations compared the $\pm 5\%$ floating bin method with the $\pm 2.5\%$ fixed bin method. His intent in performing these calculations was to demonstrate the conservative nature of the FBI's 1-in-800,000 estimate (calculated by Adams using the updated H4 database).

The probabilities yielded by Chakraborty's calculations are summarized in figure 56. Using the FBI's databases and the $\pm 5\%$ floating bin method, the probabilities are 1 in 4.4 million for Caucasians, 1 in 2.5 million for Blacks, 1 in 2.6 million for Florida Hispanics, and 1 in 4 million for Texas Hispanics. Using the FBI's databases and the

---

[81] Chakraborty did not completely explain his calculation method. He stated that he did not rebin or use a composite Hispanic database, but that he did take the higher frequency when the statistical window overlapped two bins. He did not mention whether he used the 2p formula rather than the p^2 formula, but his discussion suggests he used the less conservative p^2 formula.

[82] Presumably, these are the new expanded databases, including the H4 Hispanic database.

±2.5% fixed bin method, the probabilities are 1 in 5.3 million for Caucasians, 1 in 3.2 million for Blacks, 1 in 3 million for Florida Hispanics, and 1 in 2.6 million for Texas Hispanics. Using the Orange County databases and the ±5% floating bin method, the probabilities are 1 in 7.8 million for Caucasians, 1 in 3 million for Blacks, 1 in 4.2 million for Hispanics, and 1 in 6.8 million for Orientals. Using the Orange County databases and the ±2.5% fixed bin method, the probabilities are 1 in 5 million for Caucasians, 1 in 6.1 million for Blacks, 1 in 6 million for Hispanics, and 1 in 4.3 million for Orientals.

	FBI Database ±5% FLOAT. Bin	FBI Database ±2.5% FIXED Bin	Orange Co. Database ±5% FLOAT. Bin	Orange Co. Database ±2.5% FIXED Bin
CAUCASIAN	1 in 4.4 mil.	1 in 5.3 mil.	1 in 7.8 mil.	1 in 5 mil.
BLACK	1 in 2.5 mil.	1 in 3.2 mil.	1 in 3 mil.	1 in 6.1 mil.
HISPANIC	FL - 1 in 2.6 mil. TX - 1 in 4 mil.	FL - 1 in 3 mil. TX - 1 in 2.6 mil.	1 in 4.2 mil.	1 in 6 mil.
ORIENTAL	--	--	1 in 6.8 mil.	1 in 4.3 mil.

Fig. 56. *Profile Frequencies Derived From Orange County and FBI Databases Calculated By Chakraborty's More Conservative Method.*

Chakraborty considered the original 1-in-250,000 figure presented to the jury (calculated from the H2 database) a very conservative estimate. In his opinion, the real frequency for Pizarro's case would be on the order of one in several million for any of the major populations. Chakraborty believed this would be the consensus of the "relevant people who understand the subject."

On cross-examination, Chakraborty agreed that a 1,050 bp allele measurement would match a 1,000 bp allele measurement because the two ±2.5% windows around them would overlap. For the same reason, a 950 bp allele would also match the 1,000 bp allele. The number of people in the population who would be declared a match to the 1,000 bp allele include all the people between 950 bp and 1,050 bp. Chakraborty agreed that "[w]hen you're trying to find a frequency, you are trying to find out how many people in our population would also match this band." He did not agree, however, that using a ±2.5% statistical window to determine frequency causes an underestimation of

152.

the true number of people in the population who match the allele. But he did agree that a ± 5% statistical window could overlap more fixed bins than a ± 2.5% window.

Chakraborty noted that, when the FBI uses the floating bin method, the statistical window is double the ± 2.5% "match window," for a total of ± 5%. However, he believed a ± 5% floating bin window was too conservative, and a ± 2.5% floating bin window was the correct size to use "because that's where all fragments would be matched with respect to each other."

3. Adams

Adams, who had supervised the work in Pizarro's case, explained that the FBI's RFLP fixed bin method utilizes population databases to determine how frequently "that particular profile was likely to be found in the population." Certain features make the FBI's fixed bin method conservative. First, each bin must contain at least five database alleles; otherwise it is combined with a neighboring bin. Second, a ± 2.5% statistical window is drawn around the band, and if that window overlaps two fixed bins, the higher of the two bin frequencies is used.

At Pizarro's trial in 1990, Adams had testified to a profile frequency of approximately 1 in 250,000, calculated from the H2 Hispanic database. The H2 database is a composite of a Florida Hispanic database and a Texas Hispanic database. The composite was formed by comparing the two databases and taking the higher of the two frequencies for each bin. Thus, if all the frequencies are added together for the alleles at a particular locus, they will exceed 100 percent. This, again, is a conservative feature. Since the trial, the Hispanic database had been expanded by several hundred people. This expanded Hispanic database is called H4.

The FBI uses a ± 2.5% window not only as the window to determine whether two alleles match, but also as the statistical window to determine the allele frequency from the database frequency table. Adams agreed that if the statistical window used to determine the allele frequency is larger than any of the fixed bins, the system is not

conservative at all. He stated that the FBI's match window had never been ± 5%, and its match window had never been larger than any of its fixed bins.

At the prosecutor's request, Adams had calculated some new frequencies that had not been presented to the jury. The new calculations utilized both the databases available in 1990 and those available in 1994. Using the old databases, the probabilities are 1 in 10 million for Caucasians, 1 in 3 million for Blacks, and 1 in 250,000 for Hispanics.[83] Using the new databases, the probabilities are 1 in 5.4 million for Caucasians, 1 in 3 million for Blacks, and 1 in 890,000 for Hispanics. (See fig. 57.)

	Old Database	New Database
CAUCASIAN	1 in 10 mil.	1 in 5.4 mil.
BLACK	1 in 3 mil.	1 in 3 mil.
HISPANIC	1 in 250,000	1 in 890,000

Fig. 57. *Frequency Calculations Using Six Different Databases Calculated By FBI's Method.*

On cross-examination, Adams agreed that the purpose of calculating allele frequencies is "to find the number of people in the population that match [the perpetrator's band] or could match it …." He agreed that, assuming an initial visual match, any allele measuring between 950 bp and 1,050 bp matches the perpetrator's 1,000 bp allele measurement. Two measurements could be 5% apart and still be declared a match, while in reality, if two alleles differ by only one base pair they are not actually identical. Adams agreed the range of people who would match the perpetrator's allele fall within a ± 5% window, although that would only occur when the two bands are the maximum of 5% apart.

In response to this testimony, the prosecutor elicited the following on redirect:

"[PROSECUTOR:] So when [defense] counsel talks about the difference between 1,050 base pairs down to 950 is that realistic that

---

[83]   The 1-in-250,000 and 1-in-10-million figures were presented to the jury.

something -- I mean are those the kind of differences that you routinely see? Is that the kind of variations you are talking about when you are dealing with this technology?

"[ADAMS:] No. The two bands that are that far apart would not even match. That would be greater than our matching criteria, plus or minus two and a half percent. Typically the variation between two samples coming from the same individual is going to be less than one percent difference.

"[PROSECUTOR:] How about samples coming from two different people. What kind of variations are you likely to see there?

"[ADAMS:] Great variation. So that -- they are not even close."

4. Conneally

Conneally was familiar with the FBI's fixed bin method. He explained that the RFLP system cannot precisely measure the size of alleles and thus groups of sizes are binned together in bins, the boundaries of which are roughly evenly spaced. Next, DNA samples are collected from the people in the database. These samples are tested using RFLP. The two alleles per person are sized from the autorads, categorized, and placed into the predetermined bins. Conneally believed that, when the FBI determines allele frequency from the fixed bins, it does not take the average of the perpetrator's and defendant's alleles, but instead places both alleles separately into bins. If the two alleles fall into different bins, the higher frequency is used. If either allele is within 2.5% of another bin, then the bin with the higher frequency is used.[84]

Some features of the fixed bin method are conservative. Any bins containing less than five alleles are collapsed into a neighboring bin in a conservative effort to avoid overly rare frequencies. When the $\pm 2.5\%$ statistical window overlaps two bins, the

---

[84] Thus, the statistical window is effectively drawn around the two overlapping uncertainty windows.

larger bin frequency is assigned to the allele.**85** Conneally believed the majority of knowledgeable scientists in the field would agree that the FBI's fixed bin method is conservative.

On cross-examination, Conneally explained that the fixed bin allele frequency step seeks to determine "how many other people might match the perpetrator," "how many people could be our perpetrator," "[h]ow many people could have that real band size." Conneally did not disagree that, if it is known that the perpetrator is between 5 feet 0 inches and 6 feet 0 inches tall, it would be inappropriate to count only the people between 5 feet 3 inches and 5 feet 9 inches (i.e., within the fixed bin); it would be desirable to count all the people between 5 feet 0 inches and 6 feet 0 inches. He agreed that the bins must be at least as large as "the tolerance that you can measure people within your match window" because "[i]f it was not as big as your match window you would get misleading numbers" and an inaccurately "low frequency …."

On redirect, the following occurred:

> "[PROSECUTOR:] I probably can't clear this up, but I think that maybe as I understand it -- and correct me if I'm wrong, in [defense] counsel's hypothetical where he is talking about you have a -- you're trying to figure out how many people would -- in the population would fall between say five-two and six feet, and then saying it would be inappropriate to only take into consideration measurements between five-three and five-nine. The problem is you're starting with the proposition that in that case that you got somebody who might be six feet.
>
> "[CONNEALLY:] Yes.
>
> "[PROSECUTOR:] Whereas in this system you know that your band falls in between one of those two bins; is that correct?

---

85   Conneally's testimony suggesting a $\pm 2.5\%$ statistical window is used conflicts with his testimony that the bin is chosen by determining whether *either allele* is within 2.5% of a higher bin, which describes use of a window drawn around overlapping uncertainty windows.

"[CONNEALLY:] That is correct."

Conneally noted that, when the FBI uses the floating bin method, it uses a ± 2.5% statistical window.

B.   DEFENSE TESTIMONY

1.   Shields

Shields believed that the ± 5% floating bin method is the most appropriate method for determining allele frequency. He estimated that in this case the ± 5% floating bin method would yield a profile frequency of between 1 in 30,000 and 1 in 50,000.

Shields was familiar with the FBI's fixed bin method. He stated that the FBI's actual match window is ± 5%, and that it is inappropriate to use a ± 2.5% statistical window to determine allele frequency. He explained that the question being asked is, "what band could ever be declared to come from an actual fragment that could produce the measurement" of the perpetrator's band? Since the FBI declares a match when two ± 2.5% windows overlap, then any band falling within ± 5% of a band measurement could be a "remeasurement of the band produced by someone" who actually possesses that fragment. Shields had seen cases in which the FBI had in fact declared a match between two alleles that were 5% apart.

If the perpetrator's band is measured to be 1,000 bp, any other band whose estimated size is between 950 bp and 1,050 bp would be declared to match the 1,000 bp band. This is a matching range of 1,000 bp ± 5%. Therefore, to determine which fixed bin frequency to assign to the 1,000 bp allele, a statistical window of 1,000 bp ± 5% is used. If the statistical window is only ± 2.5%, it obviously could overlap fewer bins than if the appropriate ± 5% window were used. This will affect the ultimate frequency.

Some of the FBI's fixed bins are smaller than ± 5%. If the bin is smaller than the match window, the result will be a misleadingly rare understatement of frequency that is not conservative at all.

2.   Zabell

Zabell explained that the FBI determined from experience that its measurements can vary by as much as 2.5% from the observed value. To account for that variance, the FBI constructs around every measurement a $\pm 2.5\%$ window. For a 1,000 bp allele measurement, the $\pm 2.5\%$ window extends from 975 bp to 1,025 bp. This window signifies that the actual allele is somewhere within this window. For a second 1,050 bp allele measurement, the $\pm 2.5\%$ window extends from about 1,026 bp to 1,075 bp. This would be a borderline case, but in principle, as long as these two windows overlap, a match can be declared. Any allele measurement from 1,000 bp up to 1,050 bp could be declared to match the 1,000 bp measurement even though they are as much as 5% apart. Similarly, any allele measurement from 1,000 bp down to 950 bp could be declared to match the 1,000 bp measurement. Thus, anything between 950 bp and 1,050 bp could be declared a match with the 1,000 bp band. In other words, the FBI's match window is $\pm 5\%$.

Zabell was familiar with the fixed and floating bin methods and the calculations they involve. In determining the significance of a match, the FBI attempts to find from its database what percentage of the population would have the banding pattern that would match the perpetrator's. In Zabell's opinion, the $\pm 5\%$ floating bin method is the natural answer to the question of how many people in the population could match an allele; the $\pm 5\%$ floating bin frequency is the appropriate frequency to calculate.

The FBI, however, uses the fixed bin method, which is intended to approximate the floating bin method. The FBI draws a $\pm 2.5\%$ window around both the perpetrator's and the defendant's corresponding allele measurements. Then the outline around those two windows together is used as the statistical window to assign a frequency from the database. If that window overlaps more than one bin, the higher frequency is assigned. (Zabell had heard reference to the FBI's use of the average of the two allele

measurements and a single ± 2.5% window around that average, but in all the FBI cases he had seen in the prior two years, the FBI had not used the average.)

The FBI's rationale for using the fixed bin method -- which fails to count the matching alleles falling outside the fixed bin -- is "presumably that it could be a wash, [that] it really doesn't matter." The problem with this rationale, Zabell explained, is there is no assurance that the alleles are evenly distributed within the bins and are not crowded up at the bin boundaries, which in fact is not an infrequent occurrence in the FBI database. For example, most of the allele mass may fall into a floating bin, but may not be accurately reflected when divided into the arbitrary fixed bins. There is no guarantee that a fixed bin similar in size to a floating bin will contain a similar number of alleles such that it will approximate the floating bin frequency (because the fixed bin will almost certainly not be in the same position as the floating bin). In some cases, there will definitely be "extra matching possibilities" that are excluded by the fixed bin method. Furthermore, it is possible that a statistical window might overlap even three bins, some of which are fairly narrow. One must know the corresponding floating bin frequency to know how seriously the fixed bin method underestimates the floating bin frequency.

Zabell noted that although Monson and Budowle's article[86] concluded that the fixed bin method is conservative, their figures demonstrate that in fact the ± 2.5% fixed bin method can often underestimate the correct ± 5% floating bin answer by as much as a factor of 10. The article found that the ± 2.5% fixed bin method was more conservative than the ± 2.5% floating bin method, but not more conservative than the ± 5% floating

---

[86]   Presumably, Zabell was referring to Monson & Budowle, *A Comparison of the Fixed Bin Method with the Floating Bin and Direct Count Methods: Effect of VNTR Profile Frequency Estimation and Reference Population* (1993) 38 J. Forensic Sciences 1037 (hereafter Monson & Budowle). This article is one of the many judicially noticed documents.

bin method, in which case the ± 2.5% fixed bin method could yield frequencies 10 times rarer.

3. Bakken

Bakken was familiar with the differences between the FBI's match window and the windows used by other laboratories. The FBI uses a ± 5% match window. Because the real perpetrator is unknown, defendants whose allele measurements are larger than the perpetrator's measurement and defendants whose measurements are smaller than the perpetrator's measurement could all match the perpetrator. This method accepts all suspects who are focused within a ± 5% window around the perpetrator's allele measurement.

4. Muller

Muller absolutely did not agree that the FBI's method of calculating profile frequencies was a conservative approach. He gave the following explanation:

> "[DEFENSE COUNSEL:] Which aspect of [the FBI's] application of the product rule, in your view, is not as conservative as they profess?

> "[MULLER:] Well, certainly one of the center points of that statement by the FBI has always been that the ingredients that go into a product rule are the frequencies of the individual bands that are computed by their fixed-bin system. Their claim has always been that their fixed bins are exceedingly large relative to their measurement and, therefore, provide an overestimate of how common matching bands are.

> "In fact, that turns out to be totally incorrect. The match criteri[on] of the FBI permits matches to be called between a[perpetrator] and [defendant] sample for [defendant's] bands that are up to five percent larger than the [perpetrator's band] or five percent smaller than the [perpetrator's band]. That is, there's a total interval of plus or minus five percent around the [perpetrator's] band in which matching bands can fall.

> "Logically, when we go to a data base, the appropriate thing to do is count up all the bands around your [perpetrator's] band and plus or minus five percent, because this describes the universe of matching bands. In fact, many of the FBI fixed bins go right to the mid-point of the bin, only a plus or minus three percent. So these fixed bins are way too small to incorporate all the matching bands. As a result, any calculations based on

160.

those small fixed bins will not be conservative. In fact, they will be just the opposite. They will overstate the rarity of the pattern."

On cross-examination, Muller again explained his opinion of the FBI's fixed bin method:

"[PROSECUTOR:] You talked about some of the ways in which you thought the FBI fixed-bin system was not conservative. Are there any ways that you've identified where it is conservative?

"[MULLER:] Well, my overall impression is, if a fixed-bin system consistently underestimates the true frequency of alleles, that is, gives a number which is rarer than [it] ought to be, then the final evaluation is that it's not conservative. My own impression is, although there are some bins which are sufficiently large and may even exceed the size[] they have to[, s]o many of the fixed bins are smaller than they ought to be. [¶] On average, I cannot consider the system, the whole system, to be conservative. Again, you might say some bins are too large. They don't have to be as large as they are. Surely, I would admit that. But on the whole, I would say that my overall assessment is the overall technique is not conservative.

"[PROSECUTOR:] Do you see any conservative aspects to the fixed-bin system? Can you identify any of those?

"[MULLER:] Again, the one I mentioned. There are a few bins which are larger than they have to be, given the match rule of the FBI. For that reason, their frequencies may be slightly greater than the FBI could otherwise get away with. They do a process of merging bins when they have fewer than five observations in them. [¶] Again, relative to not doing that process, you could say that's conservative. Of course, with respect to the calculations, that's almost an irrelevant adjustment because one would never report an allele frequency less than five or ten percent. [¶] … Really, the only thing I've come up with right now is the fact that there are a few bins which were larger than they would absolutely have to be. [¶] … [¶]

"[PROSECUTOR:] How about their treatment of bands or patterns which are apparent[ly] homozygous, but may not truly [be] homozygous? [This refers to the use of 2p rather than p^2 in the frequency calculation.]

"[MULLER:] I don't view that as conservative. I view that as a necessity. In fact, every forensic lab doing RFLP analysis does that. Basically, I don't think that there's a choice there. [¶] … [¶] [Labs have] recognized that the RFLP techniques cannot type certain alleles. That is,

161.

ones that are very small. There is basically no way to correct for that problem except to make a modification of how you treat single-banded patterns. [¶] So, yeah, people could ignore that fact and make an incredibly blatant error. And if you want to say the FBI's being conservative relative to an incredibly absurd position, yeah, they are. It's such common sense that, as I've said, every forensic lab does it. It's not something that people sort of -- some do, some don't. I mean, it's a recognized reality that the labs have all faced and tackled."

C.   UNDERSIZED STATISTICAL WINDOW

Defendant first contends the FBI's ± 2.5% statistical window used to determine each allele frequency was *too small*, resulting in an underestimation of allele frequency and an overestimation of allele rarity. He explains that the FBI's use of a ± 2.5% uncertainty window mandated use of a ± 5% statistical window for determination of allele frequencies.

The question before us is whether, based on the record evidence, the FBI properly implemented the fixed bin method in this case when it used a ± 2.5% statistical window rather than a ± 5% statistical window to determine allele frequencies. We conclude it did not.

1.   Purpose of Fixed Bin Method

At the *Kelly* hearing, there was uncontradicted evidence that *the fixed bin method, when properly performed, is intended to estimate the frequency of all the alleles in the population that match, or could be the same as, the perpetrator's allele*.[87] Unequivocal defense testimony to this effect was supported by every prosecution witness. Sensabaugh agreed that it would be inappropriate to count only the people who are in a portion of the match window when determining the percentage or fraction of the population that could be the perpetrator; doing so would misrepresent the facts and result in an incorrect

---

[87]   We note that, although the method was sometimes described as estimating how many *people* in the population match the perpetrator, at this stage the method actually estimates how many *alleles* in the population match the perpetrator's allele.

frequency. Chakraborty agreed that the allele frequency attempts to determine how many people in the population would also match the perpetrator's band. Adams also agreed that the purpose of calculating allele frequencies is to find the number of people in the population who match the perpetrator's allele.[88] Conneally stated that the allele frequency seeks to determine how many people might match the perpetrator; he agreed it would be desirable to count all the people who match the perpetrator, not just the people within a smaller range of that total matching range.[89]

---

[88]   We briefly address Adams's subsequent testimony on redirect: "[PROSECUTOR:] So when [defense] counsel talks about the difference between 1,050 base pairs down to 950 is that realistic that something -- I mean are those the kind of differences that you routinely see? Is that the kind of variations you are talking about when you are dealing with this technology? [¶] [ADAMS:] No. *The two bands that are that far apart would not even match.* That would be greater than our matching criteria, plus or minus two and a half percent. Typically the variation between two samples coming from the same individual is going to be less than one percent difference." (Italics added.)

    In our opinion, this testimony missed the point of defense counsel's previous line of questioning. In this response, Adams explained that a 950 bp allele and a 1,050 bp allele would not be declared a match because they are too far apart, which of course is true. But defense counsel's point was *not* that these two alleles, one from the perpetrator and one from the defendant, would match *each other*, but that all the alleles within the *range* from 950 bp to 1,050 bp would be declared a match to the perpetrator's *1,000 bp* allele because they are within the match window around the 1,000 bp measurement.

[89]   We note, again, the inapposite nature of the subsequent testimony on redirect: "[PROSECUTOR:] I probably can't clear this up, but I think that maybe as I understand it -- and correct me if I'm wrong, in [defense] counsel's hypothetical where he is talking about you have a -- you're trying to figure out how many people would -- in the population would fall between say five-two and six feet, and then saying it would be inappropriate to only take into consideration measurements between five-three and five-nine. *The problem is you're starting with the proposition that in that case that you got somebody who might be six feet.* [¶] [CONNEALLY:] Yes." (Italics added.)

    The proposition the prosecutor mentioned -- that a person might be 6 feet tall -- is indeed the proposition supported by the match window. The match window is intended to encompass *all* the people who could match the perpetrator (5'2" to 6'0" in the

163.

There was also uncontradicted evidence that *the FBI used a ± 5% match window --* that the alleles in the population matching the perpetrator's allele fall within a ± 5% range of the perpetrator's allele measurement. This ± 5% match window is a product of the FBI's measurement imprecision; the overlapping ± 2.5% uncertainty windows *above and below* the perpetrator's allele measurement define the ± 5% range of allele measurements that are all considered to match the perpetrator's allele measurement. As the defense witnesses explained, and as every prosecution witness agreed, all the allele measurements between 950 bp and 1,050 bp would be declared a match with the perpetrator's 1,000 bp allele measurement because they all fall within ± 5% of 1,000 bp.[90]

In sum, the evidence presented at the hearing established that *the fixed bin method, which estimates the number or frequency of alleles in the population that match the perpetrator's allele, must consider and attempt to count every allele measurement in the population that falls within the match window -- which was ± 5% in this case.*

---

prosecutor's hypothetical). This includes a person who is 6 feet tall. The match window does not refer to any particular defendant in the case.

[90] In light of this uncontradicted testimony, bald statements characterizing the FBI's "match window" as ± 2.5% are unconvincing. As one commentator summarized, "The FBI has described its match criterion as a '± 2.5% window.' [Citation.] This rather ambiguous characterization has left some commentators with the erroneous impression that the FBI requires the measured sizes of two bands to be within 2.5% of each other to be declared a match. In fact, the FBI protocol specifies that the ± 2.5% window be drawn around each band and that a match may be called if these windows overlap. [Citation.] Hence, bands that differ in measured size by up to 5% may be declared to match." (*Thompson, supra,* 84 J. Crim. Law & Criminol. at p. 41, fn. 86.)

2. Proper Implementation

To implement this requirement, *the statistical window applied to the database allele frequency table was required to also be ± 5% to encompass the entire range of matching allele measurements*; a smaller window necessarily fails to take into account the entire range of matching alleles -- which by definition are the same as the perpetrator's allele. Figure 58 shows that a ± 2.5% statistical window excludes alleles on the periphery of the match window, whereas a ± 5% statistical window includes those alleles.

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

Alleles Within ± 5% Match Window

Excluded Alleles

± 5% Statistical Window

± 2.5% Statistical Window

Excluded Alleles

Fig. 58. *± 2.5% Statistical Window Applied to Database Frequency Table.*
NOTE: The match window includes the shaded alleles, i.e., the allele measurements that match the perpetrator's allele measurement. The ± 2.5% statistical window is too small to span all the shaded alleles; those on the periphery are excluded.

The obligation to account for the full range of matching alleles is made evident by this scenario: A victim describes the perpetrator as having *medium brown* hair, but she admits seeing him in dim light. The police know hair color is difficult to discern accurately under these circumstances, and therefore they decide to search for suspects with hair color from *very light brown to extremely dark brown*. This is the range of color the police will accept as "brown" -- it is the range of color describing all potential perpetrators, despite the victim's description of the perpetrator's hair as medium brown. This range is the match window. Every color description within this range is the "same" because the differences between them were impossible for the victim to discern -- the colors are all brown as the police have defined it -- and any person whose hair color falls within the range of very light brown to extremely dark brown *could be the perpetrator*.

If the police later want to demonstrate how uncommon the perpetrator's brown hair is in the population, they must use this *same definition* of brown. The police must count all the people whose hair color is very light brown to extremely dark brown because that is how the police defined the people who are potential perpetrators -- and now they must account for them all. The police cannot at this point redefine brown as meaning only *light brown to medium brown*. Doing so unfairly refocuses and narrows the field to a more specific description that inaccurately rarifies the frequency of brown hair in the population. The ability to discern hair color has not improved; it is still based on the victim's uncertain observation. Therefore, once brown has been defined, that definition must be applied consistently.[91]

The logic is simple and undeniable: if a broad description is used to describe the perpetrator, then all the people who fit that broad description must be counted to

---

[91]   Note that it is irrelevant whether the defendant who is eventually apprehended by the police has light, medium, or dark brown hair. The defendant is irrelevant at this point.

determine how many people fit that broad description. It is inconceivable that people fitting a more specific description should be counted to determine how many people fit the broad description.[92]

We believe this principle is equally applicable to both the floating bin and fixed bin methods, which share the same goal. The difference lies in how each method proceeds from the application of this principle. We note the following evidence. Both the floating and fixed bin methods begin with the same ± 5% match window that results from the measurement imprecision and defines which allele measurements could be the same as the perpetrator's allele measurement. Both methods utilize the same database population (and thus the same hypothetical frequency table). Both attempt to determine the number of alleles in the population that fall within the match window by accounting for *all* the alleles within that range, such that the frequency estimate is not overly rare and excessively harmful to the defendant. The essential difference between the two methods is that the floating bin method actually *counts* the alleles falling within ± 5% of the perpetrator's allele measurement, while the fixed bin method *estimates* this count by referring to similar pre-counted groups of alleles. (Accord, NRCII, *supra*, at p. 7 ["To calculate the frequency of matching VNTR profiles, one must find the proportion of [alleles] that fall within a *match window* around each [allele] in the incriminating profile. *Floating bins* do this exactly, whereas *fixed bins* do this approximately."].)

---

[92]   This practice has been described as "'catching a match with a 10-foot-wide butterfly net, but then attempting to prove the difficulty of the feat by showing how hard it is to catch matches with a 6-inch-wide butterfly net.'" (Thompson, *supra,* 84 J. Crim. Law & Criminol. at p. 67.)

We also note that imposition of a narrower, more specific description of the perpetrator at the statistical stage flies in the face of RFLP's persistent recognition of and accommodation to its measurement imprecision. In light of this theme repeated throughout the system, use of a ± 2.5% statistical window with a ± 5% match window seems particularly incongruous.

In both the floating and fixed bin methods, the size of the statistical window can affect the allele frequency -- but in different ways. In the floating bin method, a larger statistical window obviously includes more alleles whose frequencies will be added to the bin's frequency. In the fixed bin method, the size of the statistical window can affect frequency because a larger statistical window will overlap more fixed bins than a smaller statistical window, as uncontradicted evidence established.[93] Because allele frequency in the fixed bin method is determined by choosing the highest frequency of the overlapped bins, the *number of bins overlapped* by the statistical window can affect the allele frequency. If a larger statistical window overlaps a bin not overlapped by a smaller window, and that bin has a significantly higher frequency than the bin(s) overlapped by the smaller window, then the larger window will yield a significantly higher frequency. As a result, the failure to consider all the bins overlapped by the entire ± 5% statistical window, rather than merely those overlapped by the ± 2.5% statistical window, can affect the allele frequency to the defendant's detriment. Moreover, if the effect occurs at more than one allele, the error in the ultimate profile frequency will be amplified.

The evidence therefore established that *proper implementation of the fixed bin method required use of a statistical window at least as large as the ± 5% match window.*

---

[93]   Sensabaugh stated that a ± 5% statistical window overlaps more fixed bins (almost always two or three) than a ± 2.5% statistical window. Chakraborty agreed that a ± 5% window could overlap more fixed bins than a ± 2.5% window. Shields stated that a ± 2.5% window obviously could overlap fewer bins than the correct ± 5% window. Zabell noted that a ± 5% window could overlap even three bins.

We will illustrate. In the hypothetical example below, the $\pm 2.5\%$ statistical window falls entirely within a single bin and therefore the 1,000 bp allele is assigned the frequency of that bin, which is 0.055.[94] (Fig. 59.)

	Allele Length	# of Alleles	Freq. (#/1000)
	2020-2029	2	.002
bin 6 freq = .085	2010-2019	14	.014
	2000-2009	19	.019
	1090-1099	6	.006
	1080-1089	5	.005
	1070-1079	12	.012
	1060-1069	20	.020
	1050-1059	9	.009
bin 5 freq = .055	1040-1049	22	.022
	1030-1039	4	.004
	1020-1029	7	.007
	1010-1019	11	.011
	1000-1009	5	.005
	990-999	3	.003
	980-989	3	.003
	970-979	0	0
bin 4 freq = .025	960-969	6	.006
	950-959	9	.009
	940-949	0	0
	930-939	6	.006
	920-929	2	.002
	910-919	1	.001
	900-909	0	0
	890-899	0	0
	880-889	1	.001
	870-879	0	0

Fig. 59. *± 2.5% Statistical Window Applied to Database Frequency Table.*

---

[94]   Of course, even a small window can overlap two bins if it happens to fall over their border.

A ± 5% statistical window, on the other hand, is more likely to overlap two or three bins, provi ding more bin frequencies from which to choose the highest frequency. In the same example, the ± 5% statistical window overlaps the same bin, plus two others. The 1,000 bp allele is assigned the highest frequency of the overlapped bins, which is 0.085. This frequency is 0.030 higher than the frequency obtained with the smaller ± 2.5% statistical window. (Fig. 60.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

bin 6 freq = .085

bin 5 freq = .055

bin 4 freq = .025

1000

± 5% Statistical Window freq = .085

Fig. 60. *± 5% Statistical Window Applied to Database Frequency Table.*

Figure 61 compares these two statistical windows, demonstrating their effects in this hypothetical case.

		Allele Length	# of Alleles	Freq. (#/1000)
		2020-2029	2	.002
bin 6 freq = .085		2010-2019	14	.014
		2000-2009	19	.019
		1090-1099	6	.006
		1080-1089	5	.005
		1070-1079	12	.012
		1060-1069	20	.020
		1050-1059	9	.009
bin 5 freq = .055		1040-1049	22	.022
		1030-1039	4	.004
		1020-1029	7	.007
		1010-1019	11	.011
		1000-1009	5	.005
		990-999	3	.003
		980-989	3	.003
		970-979	0	0
bin 4 freq = .025		960-969	6	.006
		950-959	9	.009
		940-949	0	0
		930-939	6	.006
		920-929	2	.002
		910-919	1	.001
		900-909	0	0
		890-899	0	0
		880-889	1	.001
		870-879	0	0

1000

± 5% Statistical Window **freq = .085**

± 2.5% Statistical Window **freq = .055**

Fig. 61. *Comparison of ± 2.5% and ± 5% Statistical Windows Applied to Database Frequency Table.*

3.   Approximating ± 5% Floating Bin Method

Further support for the impropriety of the ± 2.5% fixed bin method is its failure to *conservatively approximate the ± 5% floating bin method*. Defense witnesses explained that the ± 5% floating bin method is the most appropriate and logical method for determining allele frequency. The floating bin method counts all the alleles falling within the ± 5% match window and thus it cannot use a statistical window smaller than the match window. [95] The fixed bin method is intended to conservatively approximate the ± 5% floating bin method; however, it can instead seriously underestimate the floating bin frequency if it fails to account for matching alleles that fall within the floating bin, but outside the fixed bin (see, e.g., fig. 61, *ante*). Only if the floating bin and the fixed bin happen to contain a similar number of alleles -- and there is no guarantee this will occur -- will the fixed bin's frequency approximate the floating bin's frequency.

The evidence established that the ± 2.5% fixed bin method fails to conservatively approximate the ± 5% floating bin method. In fact, the ± 2.5% fixed bin method yields frequencies that are often significantly less conservative (more rare) than the ± 5% floating bin method. Although the prosecution witnesses claimed the ± 2.5% fixed bin method is conservative, the evidence does not support this claim. Both of the ± 2.5% fixed bin frequencies (1 in 250,000 and 1 in 890,000) calculated in this case are less conservative than the ± 5% floating bin frequency (between 1 in 30,000 and 1 in 50,000) estimated by Shields. Similarly, five of Chakraborty's eight calculations reveal that ± 2.5% fixed bin frequencies are less conservative than ± 5% floating bin frequencies

---

[95]   Only Chakraborty stated the floating bin method should use a ± 2.5% statistical window. He based his opinion on the belief that a ± 2.5% window includes all the matching alleles. This explanation, however, is entirely unsupported by the evidence, which irrefutably demonstrates that all the matching alleles fall within a ± 5% window, not a ± 2.5% window. Chakraborty himself agreed that the alleles matching a 1,000 bp allele fall between 950 bp and 1,050 bp -- a ± 5% match window.

(e.g., 1 in 5.3 million instead of 1 in 4.4 million; 1 in 6 million instead of 1 in 3 million). (See fig. 56, *ante*.) In addition, the Monson and Budowle article cited by some of the witnesses explains that in 19 percent to 43 percent of the individuals tested the ± 5% floating bin method was more conservative than the ± 2.5% fixed bin method by up to a factor of 10 (and by a factor of 19.7 in one case). (Monson & Budowle*, supra,* 38 J. Forensic Sciences at p. 1043.)

Although Chakraborty stated that a ± 2.5% window is the appropriate statistical window for use with the fixed bin method, his opinion was based on his express belief that the ± 2.5% fixed bin method was already conservative. This conclusion is simply not reasonable in light of the evidence, which demonstrates that the ± 2.5% fixed bin method is not a reliably conservative estimation of the ± 5% floating bin method.

a. Examples

We offer a few hypothetical examples to demonstrate how the ± 2.5% fixed bin method can fail to conservatively approximate the ± 5% floating bin method. We begin with the floating bin method, then proceed to the ± 2.5% and ± 5% fixed bin methods, comparing the allele frequencies they might produce.

1.   First Example

In the ± 5% floating bin method, the ± 5% statistical window around the 1,000 bp measurement encompasses the shaded frequencies below, which, when added together, amount to a frequency of 0.079. (Fig. 62.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

1000

± 5% Floating Bin Method
freq = .079

Fig. 62.   *± 5% Floating Bin Method Applied to Database Frequency Table.*

If, instead, the ± 2.5% fixed bin method is used to estimate allele frequency, a ± 2.5% statistical window is applied to the frequency table. Now the hypothetical frequency table is divided into fixed bins, but it is otherwise identical. The ± 2.5% statistical window falls entirely within fixed bin 5, and therefore the 1,000 bp allele is assigned the frequency of that bin, which is 0.055. (Fig. 63.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

bin 6 freq = .085

bin 5 freq = .055

bin 4 freq = .025

1000

± 2.5% Fixed Bin Method **freq = .055**

Fig. 63. *± 2.5% Fixed Bin Method Applied to Database Frequency Table.*

175.

Finally, if the ± 5% fixed bin method is used to estimate allele frequency, a ± 5% statistical window is applied to the frequency table. The ± 5% window overlaps not only fixed bin 5, but two other bins as well. The 1,000 bp allele is assigned the highest frequency of the overlapped bins, which is 0.085. (Fig. 64.)

Allele Length	# of Alleles	Freq. (#/1000)
2020-2029	2	.002
2010-2019	14	.014
2000-2009	19	.019
1090-1099	6	.006
1080-1089	5	.005
1070-1079	12	.012
1060-1069	20	.020
1050-1059	9	.009
1040-1049	22	.022
1030-1039	4	.004
1020-1029	7	.007
1010-1019	11	.011
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	6	.006
950-959	9	.009
940-949	0	0
930-939	6	.006
920-929	2	.002
910-919	1	.001
900-909	0	0
890-899	0	0
880-889	1	.001
870-879	0	0

bin 6 freq = .085

bin 5 freq = .055

bin 4 freq = .025

1000

± 5% Fixed Bin Method **freq = .085**

Fig. 64. *± 5% Fixed Bin Method Applied to Database Frequency Table.*

These three methods are compared in figure 65. In this example, the ± 2.5% fixed bin method *significantly underestimates* the ± 5% floating bin method (0.055 instead of 0.079). The ± 5% fixed bin method, however, *conservatively overestimates* the ± 5% floating bin method (0.085 instead of 0.079).

	Allele Length	# of Alleles	Freq. (#/1000)
	2020-2029	2	.002
	2010-2019	14	.014
	2000-2009	19	.019
	1090-1099	6	.006
bin 6, freq = .085	1080-1089	5	.005
	1070-1079	12	.012
	1060-1069	20	.020
	1050-1059	9	.009
	1040-1049	22	.022
	1030-1039	4	.004
	1020-1029	7	.007
bin 5, freq = .055	1010-1019	11	.011
	1000-1009	5	.005
	990-999	3	.003
	980-989	3	.003
	970-979	0	0
	960-969	6	.006
	950-959	9	.009
	940-949	0	0
	930-939	6	.006
bin 4, freq = .025	920-929	2	.002
	910-919	1	.001
	900-909	0	0
	890-899	0	0
	880-889	1	.001
	870-879	0	0

1000

± 5% Floating Bin Method **freq = .079**

± 5% Fixed Bin Method **freq = .085**

± 2.5% Fixed Bin Method **freq = .055**

Fig. 65. *Summary of All Three Methods.*

2. Second Example

We recognize there are instances in which the ± 5% fixed bin method does not yield a greater, more conservative frequency than the ± 2.5% fixed bin method. This possibility, however, does not affect the conclusion that proper utilization of the fixed bin method required use of a ± 5% statistical window. Nevertheless, with the intention of further clarification of these methods, we present an example of this outcome. (Note that the hypothetical allele frequency table contains completely different frequencies than the previous example because it applies to a different hypothetical locus.) The three methods are summarized in figure 66.

Allele Length	# of Alleles	Freq. (#/1000)
1080-1089	0	0
1070-1079	1	.001
1060-1069	7	.007
1050-1059	9	.009
1040-1049	7	.007
1030-1039	4	.004
1020-1029	10	.010
1010-1019	4	.004
1000-1009	5	.005
990-999	3	.003
980-989	3	.003
970-979	0	0
960-969	8	.008
950-959	11	.011
940-949	0	0
930-939	3	.003
920-929	1	.001
910-919	6	.006
900-909	2	.002
890-899	4	.004
880-889	1	.001
870-879	0	0
860-869	0	0

bin 5 freq = .050

bin 4 freq = .039

1000

±5% Floating Bin Method **freq = .064**

±5% Fixed Bin Method **freq = .050**

±2.5% Fixed Bin Method **freq = .050**

Fig. 66. *Summary of All Three Methods.*

D. IMPROPERLY CENTERED STATISTICAL WINDOW

Defendant also claims the FBI improperly drew the ± 2.5% statistical window around the *average* of the perpetrator's and defendant's allele measurements, rather than around only the perpetrator's allele measurement. Although Adams (who was most directly involved in the DNA analysis in this case) seemed to explain that the FBI used a ± 2.5% statistical window centered on the perpetrator's allele measurement (discussed *ante*), the evidence is not entirely clear as to which method the FBI used; some evidence suggested a ± 2.5 % window was centered on the average, and other evidence suggested a window was drawn around the outline of the two overlapping uncertainty windows. In light of this uncertainty, we discuss these two additional methods here. We conclude that use of either was improper scientific procedure within the meaning of *Kelly*'s third prong because both were improperly centered on something other than the perpetrator's allele measurement.

Adams explained that the FBI used a ± 2.5% statistical window, apparently centered on the perpetrator's allele measurement. Sensabaugh and Chakraborty both testified that the FBI centered its ± 2.5% statistical window on the average of the perpetrator's and defendant's allele measurements. Conneally believed the FBI did not use an average, but instead drew a statistical window around the outline of the two overlapping uncertainty windows of the perpetrator's and defendant's allele measurements. Zabell had heard of using the average, but had not seen it in the FBI's work during the previous two years; he also believed the FBI drew a window around the outline of the two overlapping uncertainty windows.

Use of either of these two additional types of statistical windows -- one centered on the average measurement or one drawn around the uncertainty windows -- was improper because *both required reference to defendant's allele measurement*. The defendant's allele measurement, however, is irrelevant to both the match window and the statistical window, which are calculated exclusively from the *perpetrator's* allele

179.

measurement. (Accord, NRCII, *supra*, at p. 144.)[96] The statistical step is intended to estimate the frequency of the *perpetrator's* allele in the population by estimating all the alleles in the population that match or could be the same as the perpetrator's allele. Again*,* reference to defendant required the erroneous assumption that *defendant is the perpetrator*. (Accord, Monson & Budowle, *supra,* 38 J. Forensic Sciences at p. 1037 ["[T]he appropriate hypothesis is to assume the [defendant] is *not* the contributor of the [DNA] sample and then to determine what portion of the population of potential perpetrators might be responsible for the sample. In other words, we assume the [defendant] is innocent …."].)

These two types of improper statistical windows are altered in both size and position, affecting which and how many fixed bins are overlapped. The size of the statistical window drawn around the *overlapping uncertainty windows* ranges from $\pm 5\%$ down to $\pm 2.5\%$, depending on the closeness of the perpetrator's and defendant's allele measurements; the position of this window is shifted toward the defendant's allele measurement. The size of the $\pm 2.5\%$ statistical window centered on the *average allele measurement* is about half the size of the $\pm 5\%$ window centered on the perpetrator's allele measurement; the position of this window is again shifted toward the defendant's allele measurement. Figure 67 compares the sizes and positions of these two improper statistical windows to those of the proper $\pm 5\%$ window centered on the perpetrator's allele measurement in three different situations.

---

[96] NRCII criticizes this method for incorporating the defendant's allele measurement into the determination of the statistical window, stating that the frequency should depend solely on the perpetrator's allele measurement and that the defendant's allele measurement and the uncertainty window around that measurement are irrelevant to this computation.
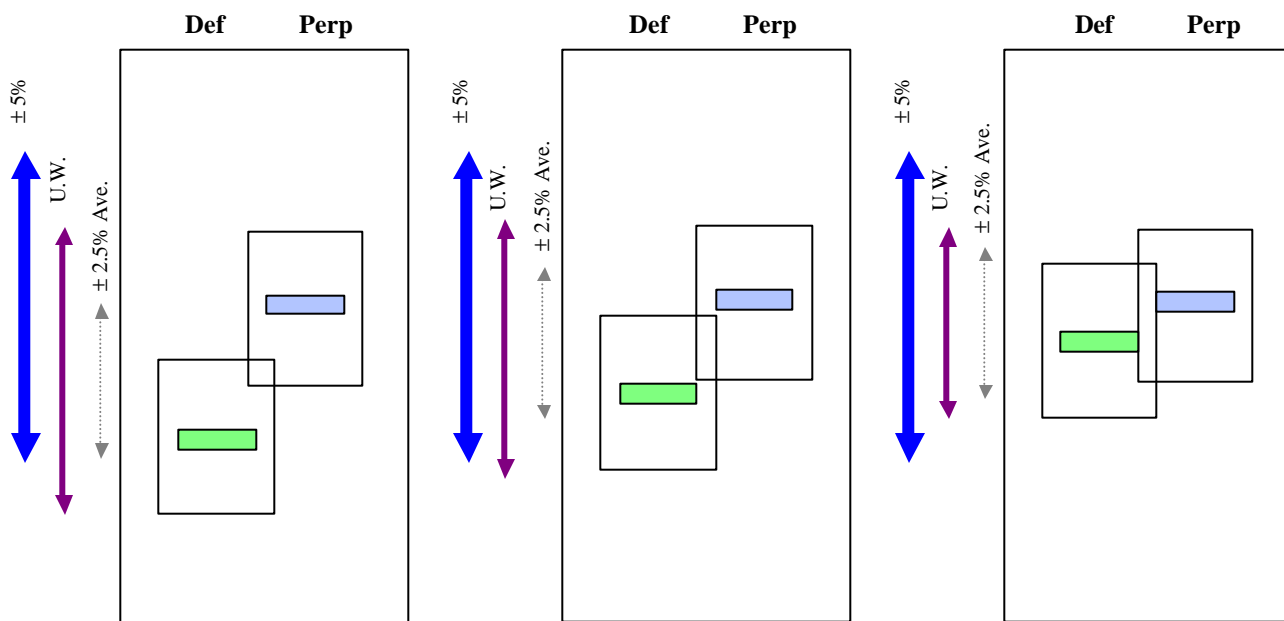
Fig. 67. *Comparison of Two Improper Statistical Windows with Proper Statistical Window.*
NOTE: The heavy outside arrow (±5%) is the proper statistical window based on the ±5% match window and centered on the perpetrator's allele measurement. The finer middle arrow (U.W.) is a statistical window based on the outline of the two overlapping uncertainty windows. The dashed inner arrow (±2.5% Ave.) is a statistical window centered on the average of the perpetrator's and defendant's allele measurements.

Use of either of these improper statistical windows, which can affect bin overlap and thus allele frequency, constituted improper scientific procedure.

E.   THE PEOPLE'S ARGUMENTS

We are compelled to comment on the People's responses to defendant's contentions regarding the statistical window because we believe they reflect a misunderstanding of the principles underlying this issue.

First, the People counter that NRCII deems the fixed bin method "'acceptable.'" Defendant, however, does not contend the fixed bin method is unacceptable or unaccepted; he contends it was improperly performed in this case. In addition, the People's citation to NRCII, when read in full, states: "*When our fixed-bin recommendation is followed*, the two methods lead to very similar results. Both methods are acceptable." (NRCII, *supra*, at p. 162, italics added.) The recommendation to which

this statement refers is use of a ± 5% statistical window. (*Id.* at p. 144.)[97] Thus, this reference is helpful only to defendant's case.

Second, the People assert that defendant's contentions are "effectively rebut[ted]" by testimony establishing the conservative nature of the FBI's fixed bin method, specifically, that the fixed bins are equal or larger than the match window. As we have noted, the evidence does not support the conclusion that the FBI's ± 2.5% fixed bin method is conservative. As for the evidence regarding the adequate size of the fixed bins, that testimony compared the bins to a match window incorrectly defined as ± 2.5%. If the fixed bins are instead compared to the correct ± 5% match window, many of the bins are smaller than the match window.

This situation, the evidence established, does not make for a conservative system. Sensabaugh testified that one reason the fixed bin method is conservative is that the ± 2.5% statistical window is smaller than the fixed bins. Chakraborty also stated that the method is conservative because the fixed bins are overly large. Adams explained that the fixed bin method is absolutely not conservative if the statistical window is larger than any of the bins. Conneally agreed that the fixed bins must be at least as large as the match window; otherwise, the result will be misleading and inaccurately low. Shields explained that some of the FBI's fixed bins are less than ± 5% wide; if the bin is smaller than the match window, the result will be a misleadingly rare underestimate of the frequency that is not conservative at all. Muller stated that many of the FBI's fixed bins are too small; many are only ± 3% wide. These bins are far too small to encompass all the matching

---

[97] "To approximate the floating-bin match probability, we recommend using the fixed bin with the largest frequency among those overlapped by the *match window*." (Italics added.) (*Id.* at p. 144.) NRCII defines the match window as the perpetrator's allele measurement ± 5% of that measurement. (*Id.* at p. 143.)

allele measurements and the resulting frequencies will not be conservative, but will overstate the rarity of the alleles.

Although defendant does not raise this issue, the evidence clearly suggests the fixed bins should have been approximately the size of the match window in order to estimate the frequencies of similarly sized *floating* bins, which are the size of the match window. If the fixed bins are significantly smaller than the match window, they will encompass fewer alleles (whose frequencies are added together as with floating bins) and thus will have lower frequencies. One of these frequencies will be assigned to the allele, and will likely underestimate the correct floating bin frequency.

Lastly, the People state defendant's contentions are rebutted by testimony that the allele measurements of the perpetrator and defendant in this case are in fact close together. Both Sensabaugh and Chakraborty testified that the bands in this case all fall within about 1% of each other. But the fact that the perpetrator's and defendant's alleles are close together is irrelevant to the size of both the match window and the statistical window. The closeness of the bands matters only to the matching step -- if the bands are within 5% of each other, they are declared a match. However, once a match is declared, the defendant's allele is irrelevant and reference to it improper. (Accord NRCII, *supra*, at p. 144 [match probability depends on perpetrator's allele measurement; thus defendant's allele measurement is irrelevant for computation of statistical window].)

F.   CONCLUSION

Based on the evidence presented at the *Kelly* hearing, we conclude that the FBI's use of a statistical window that was only ± 2.5% constituted improper scientific procedure in the calculation of the fixed bin allele frequencies in Pizarro's case. Furthermore, if the statistical window was centered on something other than the perpetrator's allele measurement, that too was improper procedure. The trial court therefore abused its discretion when it failed to rule that the FBI's scientific procedure was improper. "There was no substantial evidence upon which to base a contrary

conclusion, and therefore the trial court abused its discretion in not excluding the flawed statistical evidence. [Citations.]" (*People v. Venegas, supra*, 18 Cal.4th at p. 93.)

IX.   H2 HISPANIC DATABASE

Defendant argues that the FBI's H2 Hispanic database is defective because it contains over 20 duplicate samples and some ethnically misclassified samples. He also claims there are indications the database departs from Hardy-Weinberg equilibrium. Use of this H2 database, defendant explains, was improper procedure under the third *Kelly* prong. The People maintain that defendant waived this contention by failing to raise it in his posthearing brief,[98] and that the contention fails on its merits nonetheless.

Defendant challenges the propriety of the H2 Hispanic database in particular, but in light of our conclusion that use of *any* Hispanic database was error because there was insufficient proof that the perpetrator is Hispanic, this issue is moot. However, for the purposes of retrial, and to highlight another important issue of continuing interest, we address the merits of defendant's contention. (*Burch v. George* (1994) 7 Cal.4th 246, 253, fn. 4.) We conclude the trial court did not abuse its discretion by not ruling that the FBI's use of the H2 Hispanic database was improper scientific procedure due to any defect in the H2 database specifically.

A.   *PROSECUTION TESTIMONY*

1.   Sensabaugh

Sensabaugh testified that, since the time of trial, the FBI had recalculated the frequency in Pizarro's case using the updated H4 database. The original frequency presented to the jury is 1 in 250,000, but the recalculated frequency, using the H4 database, is 1 in 890,000. Sensabaugh had not examined the two databases and did not

---

[98]   For the sake of economy and to avoid review on the issue of ineffective assistance of counsel, we address the issues the People claim are waived.

know their differences. He presumed the H4 database is simply bigger. He speculated that as a database increases in size, the assignment of bands to bins better reflects the population. In a smaller database, such as H2, more of the bins may contain less than five alleles, and thus more bins may be collapsed and combined into larger bins. In a larger database, such as H4, a bin that may have contained only two or three alleles in the smaller database, may now contain five, six, or seven alleles and therefore the bin remains a (smaller) bin in its own right. This sampling phenomenon may affect the final frequency computation, even by several fold. The idea of making databases more complete is generally to make the sample more representative of the population, but there is a point at which increasing the sample size does not affect the population further.

2. Chakraborty

Chakraborty explained that, in developing databases, laboratories share and collect DNA samples. When a laboratory collects data from different sources to compile a database, it is that laboratory's duty to find the shared samples, or duplicates, in the data and remove them. Duplicates are common due to this scientific practice of sharing samples. When the duplicates are discovered in the collected samples, they are removed. The FBI incorporated some samples into its database without realizing the samples were duplicates; when it realized the database contained duplicates, it removed them.

The FBI's initial purpose did not require removal of duplicates. Various studies have shown that allele frequencies are not consistently changed when a small fraction of duplicates are present. There is no systematic change either up or down in the frequencies. Any error would not be meaningful. The FBI discontinued its use of the H2 database and began using the H4 database in January, 1992. The H4 database contains more samples than the H2 database. The FBI updated other databases as well.

Chakraborty believed the 1-in-250,000 frequency given to the jury is a very conservative number, even in light of the database expansion.

At defense counsel's request, Chakraborty calculated the frequencies for a *hypothetical* three-locus profile using the FBI's ± 2.5% fixed bin method with both the H2 and H4 databases.[99] He determined that the profile using the H2 database is 1 in 518,468, whereas, using the H4 database, it is 1 in 195,486. (Fig. 68.)

	± 2.5% FIXED Bin
H2 DATABASE	1 in 518,468
H4 DATABASE	1 in 195,486

Fig. 68. *Chakraborty's H2 and H4 Database Hypothetical Calculations.*

Chakraborty agreed that in this hypothetical the H4 database yielded a frequency more than twice as common as the H2 frequency. He explained that he would expect the frequency to be more common in a database containing more people, although this does not always occur. He believed that a five-fold difference between these numbers is not biologically meaningful or forensically significant.

3. Adams

Adams stated that in 1990 he had testified before the jury that the profile frequency in this case is 1 in 250,000. This figure was calculated using the H2 database, the first Hispanic database used by the FBI. Since that time, the FBI databases had been expanded. The H2 database was expanded by several hundred individuals. The expanded H4 database was the FBI's most current Hispanic database.

At the prosecutor's request, Adams calculated the frequency in Pizarro's case using the H4 database. His result was 1 in 890,000, a frequency more rare and less beneficial to defendant than the original 1-in-250,000 H2 frequency. (Fig. 69.)

---

[99]   The resulting frequencies are not directly comparable to the frequencies calculated by Adams in this case because they are based on hypothetical numbers, not the numbers generated in this case.

	± 2.5% FIXED Bin
H2 DATABASE	1 in 250,000
H4 DATABASE	1 in 890,000

Fig. 69. *Adams's H2 and H4 Database Calculations.*

4. Conneally

Conneally was not familiar with the differences between the H2 and H4 databases, but he thought H4 is a larger sample. Conneally was aware that the FBI had cleansed some of its databases after it realized they contained duplicate samples. He explained that he did not know whether, prior to that time, the FBI had used a database containing duplicates in its casework. The FBI might have, but it would not concern him if it had. He explained that 4 duplicates in a database of 500 would not concern him, but that 10 duplicates in a database of only 30 would really concern him. There is a point at which the presence of duplicates would be important.

B. DEFENSE TESTIMONY

1. Shields

Shields explained that most of the early FBI databases contained duplicate samples. Two samples from the same person were treated as though they were from two separate individuals. The FBI discovered that the multi-locus profiles of some of the samples were identical to the multi-locus profiles of other samples. The FBI reported this to the laboratories that had donated the samples, and found that a number of individuals had indeed been added twice under separate identification numbers. The FBI ran an internal matching program on the database samples. When the FBI found individuals who matched at four or more loci, it had the samples tested at additional loci, and removed them if they matched. Approximately 20 duplicates were removed. In addition, 17 percent of those duplicates were found to have been placed within more than one ethnic database.

The H2 database used in this case existed prior to the time the FBI corrected these problems. The FBI later removed the duplicates and misclassified samples, and may

have added a few individuals, to create the H4 database. The H2 database was no longer in use.

Shields also stated that there were indications that the H2 database is out of Hardy-Weinberg equilibrium, but that the FBI and others had suggested the indications had to do with technical problems in distinguishing between true homozygotes and apparent homozygotes. He believed the FBI's explanation did not entirely answer the question.

C.   ANALYSIS

The trial court did not err in finding that the prosecution made the necessary foundational showing that the H2 database is not defective and that the FBI's use of the database was not improper on that basis. Although the later-developed H4 database is larger and had 20 or so duplicates removed from it, there was evidence that the differences are not significant and any resulting error would not be meaningful. The evidence suggesting the H2 database is out of equilibrium was insubstantial. Most significantly, Adam's calculations established the H2 database is more favorable to Pizarro by approximately 3.5-fold (even though Chakraborty's calculations suggested the H4 database is more favorable to a hypothetical defendant). The trial court reasonably determined that the FBI's use of the H2 database was not improper due to any defect.

X.   LABORATORY ERROR

Defendant argues that the possibility of laboratory error should have been presented to the jury with the profile frequency and that the failure to do so amounted to *Kelly* error. Further, he contends that, in the absence of an acknowledgment that laboratory errors occur and a statement of their frequency, the evidence was prejudicial and misleading under Evidence Code section 352. The People repeat that the contention has been waived for failure to raise it in the post-hearing brief. They maintain that there was nevertheless no error, and if error did occur it was harmless.

We believe the issue of whether laboratory error rates should have been presented to the jury in addition to the profile frequency is not one that goes to the very integrity or

reliability of the DNA results. Furthermore, the defense was not barred from challenging the profile frequencies and presenting evidence of laboratory error at trial. We decide here that the trial court did not abuse its discretion by finding that presentation of the profile frequency without a separate laboratory error rate was proper.[100] No substantial evidence suggested otherwise. Whether calculation and presentation of laboratory error rates could have improved the FBI's scientific procedure went to the weight of the evidence, not its admissibility.[101] (*People v. Brown, supra*, 91 Cal.App.4th at p. 654.)

XI.   CONFIDENCE INTERVAL

Defendant lastly asserts, again under *Kelly*'s third prong, that the failure to present a confidence interval with the profile frequency was incorrect scientific procedure. A

---

[100]   In *People v. Reeves* (2001) 91 Cal.App.4th 14, the court held that the question of whether laboratory error rates should be *incorporated* into the calculation of random match probabilities is a first-prong *Kelly* issue. (*Id.* at pp. 43-44.) The court then determined that it is generally accepted by the scientific community to calculate such probabilities without modification to account for laboratory error rates, as NRCII recommended. (*Id.* at p. 46.)

[101]   NRCII explains that "[t]he question to be decided is not the general error rate for a laboratory or laboratories over time but rather whether the laboratory doing DNA testing in this particular case made a critical error. This risk of error in any particular case depends on many variables (such as number of samples, redundancy in testing, and analyst proficiency), and there is no simple equation to translate these variables into the probability that a reported match is spurious." (NRCII, *supra*, at pp. 85-86.) NRCII concludes that "a calculation that combines error rates with match probabilities is inappropriate. The risk of error is properly considered case by case, taking into account the record of the laboratory performing the tests, the extent of redundancy, and the overall quality of the results. However, there is no need to debate differing estimates of false-match error rates when the question of a possible false match can be put to direct test …." (*Id*. at p. 87.) NRCII recommends duplicate independent testing to confirm the original results: "[N]o amount of care and proficiency-testing can eliminate the possibility of error. However, duplicate tests, performed as independently as possible, can reduce the risk of error enormously. The best protection that an innocent suspect has against an error that could lead to a false conviction is the opportunity for an independent retest." (*Id.* at p. 88.)

confidence interval is a *range* or window around the *profile frequency* that is expected to include the true value a certain percentage of the time. (NRCII, *supra*, at p. 146; Easteal, *supra*, at pp. 99-100.) It is intended to account for sampling error -- the fact that a profile frequency might be different if another database were used. (NRCII, *supra*, at p. 146.) Defendant contends that the jury should have been given this range in addition to the profile frequency because it provides depth and meaning to the profile frequency. In addition, he claims the failure caused the frequency to be more prejudicial than probative and to mislead and confuse the jury. (Evid. Code, § 352.) The People argue the frequency was conservative even without a confidence interval. They contend the confidence interval in this case would have been forensically insignificant and thus any error was harmless.

As with the previous issue, we find this issue does not underlie the integrity of the DNA results, but instead goes to the weight of the evidence. No evidence suggested presentation of a confidence interval was required, although it may in fact be advisable.[102] The trial court acted within its discretion when it found that presentation of the profile frequency without a confidence interval was proper scientific procedure. Consequently, whether the FBI's proper procedure could have been made more accurate by the presentation of a confidence interval was an issue for the jury to weigh. (*People v.*

---

[102] NRCII describes the use of confidence intervals as "helpful" (NRCII, *supra*, at p. 146) and "desirable" (*id.* at p. 148), but explains that "confidence limits address only part of the uncertainty. For a more realistic estimate, [NRCII] examined empirical data from the comparison of different subpopulations and of subpopulations with the whole. The empirical studies show that the differences between the frequencies of the individual profiles estimated by the product rule from different adequate subpopulation databases (at least several hundred persons) are within a factor of about 10 of each other, and that provides a guide to the uncertainty of the determination for a single profile." (*Id.* at p. 160.) Thus, NRCII suggests that a ± 10-fold range is a reasonable estimate of the uncertainty arising from database variation. (*Id.* at pp. 148-156.)

Brown, supra, 91 Cal.App.4th at p. 654.) The defense was not precluded from challenging the accuracy of the profile frequency or from presenting the jury with evidence of confidence intervals.

XII.   CONCLUSION

Highly technical issues, such as those in this case, require close and careful scrutiny. The technicality often disguises fairly straightforward evidentiary issues that, when revealed in a more familiar form, can be resolved readily. Here, one such revelation is a persistent and insidious tendency to assume the defendant's guilt -- the perpetrator's description is created from the traits of the *defendant*, who is incriminated because he now fits the description of the perpetrator. The logical and evidentiary infractions in such an exercise are stunning in scope and consequence, even if not immediately apparent.

We summarize our conclusions and recommendations:

(1) The profile frequency derived from the Hispanic database was admitted without adequate foundation. The trial court abused its discretion both in finding sufficient evidence of the perpetrator's Hispanic ethnicity and in failing to find this scientific procedure improper. In addition, the jury was encouraged to draw improper inferences when it was informed that the Hispanic database was chosen because defendant is Hispanic.

(2) The D2 autorad data were also admitted without adequate foundation. The perpetrator's genotype could not be discerned from the mixed perpetrator/victim sample on the D2 autorad, except possibly by band-intensity analysis, which has not been subjected to *Kelly* first-prong scrutiny. Reference to defendant's genotype to establish the perpetrator's genotype was error. The trial court abused its discretion both in finding sufficient evidence of the perpetrator's genotype and in failing to find this scientific procedure improper.

On retrial, the D2 autorad and the genotype frequencies estimated from it will be admissible only if the prosecution presents adequate foundational evidence to establish that the perpetrator's genotype may be discerned reliably from the D2 autorad. If band-intensity analysis is used to discern the perpetrator's genotype, the prosecution must show that the procedure is generally accepted by the scientific community as a reliable procedure for that purpose. Then, the prosecution must demonstrate that the FBI (or other laboratory) followed that procedure in this case. If adequate foundational evidence is not presented regarding the D2 autorad, data from that autorad cannot be included in the overall profile frequency calculation.

(3) The evidence established that the FBI's fixed bin statistical window was required to be at least $\pm 5\%$ because the FBI's match window was $\pm 5\%$. A smaller window overlaps fewer fixed bins and can significantly underestimate allele frequency. The trial court abused its discretion by not ruling that the FBI's use of an undersized statistical window was improper scientific procedure.

(4) The evidence established that the FBI's statistical window may have been centered on the average of the perpetrator's and defendant's alleles or drawn around the outline of the perpetrator's and defendant's overlapping uncertainty windows. The trial court abused its discretion by failing to rule that either of these methods constituted improper scientific procedure.

(5) The issue of whether the H2 Hispanic database should not have been used because it is defective is moot; however, the trial court did not abuse its discretion in determining that use of the database was not improper due to any defect.

(6) The trial court did not abuse its discretion by not ruling that the failure to present the possibility of laboratory error in addition to the profile frequency was improper scientific procedure.

(7) The trial court did not abuse its discretion by not ruling that the failure to present a confidence interval in addition to the profile frequency was improper scientific procedure.

The People contend that the improper admission of the DNA evidence in this case was harmless because of the "overwhelming non-DNA evidence of guilt." However, we expressly stated in *Pizarro I* that, although "the prosecution presented a strong circumstantial case" against defendant, "the DNA identification evidence clearly 'sealed [his] fate.' Although the jury might have had a reasonable doubt regarding [defendant's] guilt absent the DNA evidence, it is difficult to imagine how the jury could have reached other than a guilty verdict when presented with the evidence that the likelihood of finding someone else with a DNA profile in the non-Hispanic Caucasian population was 1 in 10 million and 1 in 250,000 in the Hispanic population. Therefore, it cannot be established that the admission of the evidence constituted harmless error." (*Pizarro I, supra,* 10 Cal.App.4th at p. 90.) We decline to reconsider this finding here.

Upon retrial, we hope this opinion clarifies the issues, which admittedly can be extraordinarily complex and daunting. The DNA evidence in this case is represented by the three autorads (four if the D17 autorad is found to be readable). These autorads may now be re-analyzed by the FBI or another appropriate institution. The new profile

frequency determined from those autorads, if supported by sufficient foundational evidence, may be presented at trial.

DISPOSITION

The judgment is reversed.

_____

Ardaiz, P.J.

WE CONCUR:

_____

Vartabedian, J.

_____

Harris, J.