IN THE COURT OF APPEALS OF THE
STATE OF OREGON

STATE OF OREGON,
Plaintiff-Respondent,
v.
ODELL TONY ADAMS,
Defendant-Appellant.
Multnomah County Circuit Court
18CR73436; A177071

Jerry B. Hodson, Judge.

Argued and submitted October 11, 2023.

Neil F. Byl, Deputy Public Defender, argued the cause for appellant. Also on the briefs was Ernest G. Lannet, Chief Defender, Criminal Appellate Section, Office of Public Defense Services.

Peenesh Shah, Assistant Attorney General, argued the cause for respondent. Also on the brief were Ellen F. Rosenblum, Attorney General, and Benjamin Gutman, Solicitor General.

Janis C. Puracal filed the brief *amicus curiae* for Forensic Justice Project.

Before Aoyagi, Presiding Judge, Egan, Judge, and Joyce, Judge.*

EGAN, J.

Reversed and remanded.

---

* Egan, Judge *vice* Jacquot, Judge.

EGAN, J.

Defendant appeals convictions for unlawful use of a weapon with a firearm and second-degree criminal mischief, raising two assignments of error. First, he argues that the trial court erred by admitting as scientific evidence the testimony and report of an expert who, using the Association of Firearm and Toolmark Examiners method (the AFTE method), "identified" cartridge cases found at the scene of a shooting as having been fired from a gun found in defendant's home. Defendant contends that the state failed to establish that the AFTE method is scientifically valid and thus that evidence based on it is admissible under *State v. O'Key*, 321 Or 285, 899 P2d 663 (1995), and *State v. Brown*, 297 Or 404, 687 P2d 751 (1984). Second, he argues that the trial court erred in denying his motion to controvert and suppress evidence obtained pursuant to warrants. As explained below, we agree with defendant that the state failed to show that the AFTE method is scientifically valid. However, we disagree that the court erred in denying defendant's motion to controvert and suppress. We reverse and remand for further proceedings.

I. PROCEDURAL HISTORY

The charges against defendant arose out of a shooting at the Speakeasy Lounge in 2018. Directly after the shooting, police recovered several .40 caliber shell cases from the ground and found bullet holes in two cars parked in the nightclub's lot. The police eventually applied for a warrant to search a residence where officers believed defendant was residing. During the course of the search, officers seized, among other things, a Taurus handgun. In preparing the state's case against defendant, forensic examiner Todd used the AFTE method to analyze the Taurus and the cartridge cases found at the scene.[1] Todd concluded that the Taurus seized from defendant's residence had fired the

[1] As we will discuss below, before defendant's trial in state court, he was tried on a federal charge of felon in possession of a firearm, 18 USC § 922(g)(1). In that case, the government sought to present evidence from a different AFTE examiner who had analyzed the Taurus and the cartridge cases, but the district court excluded that evidence, concluding that the AFTE methodology is "quasi-scientific" and depends on subjective determinations of the examiner. *United States v. Adams*, 444 F Supp 3d 1248 (D Or 2020). Defendant was acquitted of the federal charge.

10 cartridge cases found at the crime scene. After Todd concluded her analysis, another AFTE examiner, Alessio, examined the Taurus and cartridge cases and agreed with Todd's conclusion.

Defendant moved to controvert the warrant and moved to exclude, among other things, evidence from Todd and Alessio. He argued that, under *Brown* and *O'Key*, the state could not meet its burden to show that the AFTE method, on which Todd's and Alessio's expert opinions were based, was scientifically valid. After holding a hearing pursuant to OEC 104, at which it heard testimony from Todd, Alessio, and defendant's expert, the trial court denied the motion to exclude the AFTE identification evidence. The court also denied defendant's motion to controvert the warrant. At trial, the state presented testimony and a report from Todd and testimony from Alessio, as well as evidence discovered during execution of the warrant. The jury found defendant guilty of two counts of unlawful use of a weapon with a firearm and two counts of second-degree criminal mischief; it acquitted him of attempted second-degree murder and attempted first-degree assault.

Defendant appeals, assigning error to the denial of the two motions. We begin by considering the motion to exclude the evidence based on the AFTE method.

II. SCIENTIFIC VALIDITY OF THE AFTE METHOD

A. *Standards for Admission of Scientific Evidence*

On appeal, the parties agree, as they did below, that Todd's and Alessio's evidence is scientific evidence, that is, that it "draws its convincing force from some principle of science, mathematics and the like." *Brown*, 297 Or at 407. At trial, Todd testified that she had been employed by the Oregon State Police Crime Lab as a forensic scientist; she had used "the AFTE methodology and tool mark analysis" to analyze the Taurus and the cartridge cases that were found at the scene; and her analysis "identified" the Taurus as having fired the cartridge cases. Alessio testified similarly. The state also presented Todd's "Analytical Report," which stated, "Based on microscopic comparisons, [the Taurus] was **IDENTIFIED** as having fired [the cartridge cases found at the scene]." (Boldface

and capitalization in original; internal footnote omitted.) In a footnote after "**IDENTIFIED**," the report stated, in a smaller font, the AFTE definition of identification: "Identification: Agreement of all discernible class characteristics and sufficient agreement of a combination of individual characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool."

Evidence that jurors perceive to be based on science is particularly persuasive. *O'Key*, 321 Or at 291. As a result, before allowing scientific evidence to be presented to the jury, the court, performing a gatekeeping function, must "ensure that the persuasive appeal is legitimate." *Id.* Specifically, "admissibility of scientific evidence requires a showing that it is based on scientifically valid principles." *Id.* at 301-02.[2]

The proponent of the scientific evidence must make that showing at an OEC 104 hearing by a preponderance of the evidence.[3] *Id.* at 307 n 29. On appeal, we review for legal error the trial court's determination that the proponent made the necessary showing. *Id.* at 320 ("'Notwithstanding the usual deference to trial court discretion, we as an appellate court retain our role to determine the admissibility of scientific evidence under the Oregon Evidence Code.'" (Quoting *Brown*, 297 Or at 442; internal footnote omitted.)); *id.* at 320 n 45 ("[a]lthough this court typically is deferential to a trial court's findings of preliminary facts under OEC 104(1), good

[2] We have explained that it is possible for expert testimony based on a "practical" expert's training and experience—that is not presented as and will not be perceived as scientific—to be relevant and helpful to the jury and, thus, admissible as nonscientific expert evidence: "[E]xpert testimony that is based on an expert's training and experience is not scientific evidence if it is, and will be recognized by the trier of fact as, merely the 'practical observations and assessments' of a person who has had 'unusual specialized experiences.'" *State v. Beltran-Chavez*, 286 Or App 590, 605, 400 P3d 927 (2017) (quoting William Strong, *Language and Logic in Expert Testimony: Limiting Expert Testimony by Restrictions of Function, Reliability, and Form*, 71 Or L Rev 349, 369 (1992) (cited in *O'Key*, 321 Or at 291, 292)). This case concerns only scientific expert evidence, not nonscientific expert evidence based on training and experience.

[3] We decline the state's invitation to rule that scientific evidence based on the AFTE method is presumptively valid, because it does not present "a clear case, a case for judicial notice, or a case of *prima facie* legislative recognition." *O'Key*, 321 Or at 293 (internal footnote omitted).

reasons exist to modify this approach in the context of scientific evidence"; explaining those reasons (internal citation omitted)).

In *Brown* and *O'Key*, the Supreme Court set out a variety of factors that may assist courts in determining whether the proponent of scientific evidence has met their burden of showing that its underlying principles are scientifically valid. *Brown* was decided before the United States Supreme Court adopted a similar approach in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579, 113 S Ct 2786, 125 L Ed 2d 469 (1993). In *O'Key*, the Oregon Supreme Court expressed agreement with the United States Supreme Court's approach in *Daubert* and adopted its factors—as well as retaining the factors it had articulated in *Brown*—under Oregon law, noting that Oregon courts should treat *Daubert* as "instructive."[4] 321 Or at 306-07.

The court applies the factors to guide its ultimate determination regarding scientific validity: "[T]he 'overreaching subject' of this multifactor, 'flexible' inquiry 'is the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission.'"[5] *Id.* at 305 (quoting *Daubert*, 509 US at 594-95).

Although the *O'Key* court used the term "validity" to describe the overarching inquiry, it recognized that it includes consideration of both the ability of the method to measure what it purports to measure (validity) and the

---

[4] The factors, as articulated in *O'Key*, include the factors articulated in *Brown*—the technique's general acceptance in the field; the expert's qualifications and stature; the use which has been made of the technique; the potential rate of error; the existence of specialized literature; the novelty of the invention; and the extent to which the technique relies on the subjective interpretation of the expert—and the factors articulated in *Daubert*—whether the theory or technique in question can be (and has been) tested; whether the theory or technique has been subject to peer review and publication; the known or potential rate of error; and the degree of acceptance in the relevant scientific community. *O'Key*, 321 Or at 299, 303-04.

[5] Although the court in *O'Key* recognized that the court's determination of scientific validity of a particular methodology might be appropriate for judicial notice in future cases, 321 Or at 293 n 8, it also recognized that understandings of what is scientifically valid may change over time and on different records: "[N]o particular reason of logic or good sense exists to immunize particular areas or principles simply on the basis of longevity or the fact that their introduction antedated imposition of the new standard. Supposedly valid science has not infrequently been unmasked." *Id.* at 293 n 9 (internal quotation marks omitted).

trustworthiness of its conclusions (reliability): "Whereas validity describes how well the scientific method reasons to its conclusions, reliability describes the ability of the scientific method to produce consistent results when replicated." *Id.* at 301 n 19. Evidence that is based on a method that is both scientifically valid—it is "capable of measuring what [it] purport[s] to measure"—and reliable—it "produce[s] consistent results when replicated"—may be relevant under OEC 401 and sufficiently helpful to the jury to be admissible under OEC 702. In that case, "it will be excluded only if its probative value is substantially outweighed by one or more of the countervailing factors set forth in OEC 403." *Id.* at 299. In applying all of those evidentiary rules, "the court must identify and evaluate the probative value of the proffered scientific evidence, consider how that evidence might impair rather than help the trier of fact, and decide whether truthfinding is better served by admission or exclusion." *Id.* (internal footnote omitted).

B. *Analysis*

As we will explain, in this case, the state did not meet its burden to show that the AFTE method is scientifically valid, that is, that it is capable of measuring what it purports to measure and is able to produce consistent results when replicated. That is so because the method does not actually *measure* the degree of correspondence between shell cases or bullets; rather, the practitioner's decision on whether the degree of correspondence indicates a match ultimately depends entirely on subjective, unarticulated standards and criteria arrived at through the training and individualized experience of the practitioner. For a similar reason, the state did not show that the method is replicable and therefore reliable: The method does not produce consistent results when replicated because it cannot be replicated. Multiple practitioners may analyze the same items and reach the same result, but each practitioner reaches that result based on application of their own subjective and unarticulated standards, not application of the same standards.

AFTE identification evidence is presented to jurors as relevant to and probative of whether a particular bullet or cartridge case was fired from a particular firearm. In the

abstract, evidence of that type may be capable of helping jurors; as we will explain, many practitioners of the AFTE method succeed in matching bullets and cartridge case to the firearms from which they were fired more often than would be expected based on chance. However, when presented as scientific evidence, AFTE identification evidence—an "identification" purportedly derived from application of forensic science—impairs, rather than helps, the truthfinding process because it presents as scientific a conclusion that, in reality, is a subjective judgment of the examiner based only on the examiner's training and experience and not on any objective standards or criteria. AFTE identification evidence is thus presented to jurors cloaked with the "aura of reliability" of science even though it is not actually derived through science. *O'Key*, 321 Or at 302 n 20. We conclude that the AFTE method is not scientifically valid under *O'Key*.

1. *The AFTE theory*

We begin by describing the AFTE theory. The initial premise is that unique toolmarks are left inside every firearm during the manufacturing process. Alessio testified that the tools used to manufacture firearms leave marks on the internal parts of those firearms. As those tools cut the metal of the firearms they are manufacturing, the tools wear and change, and, as a result, the marks that they leave inside the firearms "will change from one firearm to the next."

The next premise of the AFTE theory is that bullets and cartridge cases that have been fired from a given firearm have microscopic features created by the toolmarks left inside the firearm, as well as other attributes of the firearm, for example, environmental exposure after manufacture. Because the manufacturing tools and other features create marks that are unique to each firearm, the theory goes, the microscopic features on bullets and cartridge cases fired from a given firearm will also be unique. The AFTE theory refers to firearms as "tools" and the microscopic features left on bullets or cartridge cases by the toolmarks inside the firearm as "toolmarks."[6]

[6] In other words, somewhat confusingly, there are two sets of tools at issue. First are the tools used to manufacture firearms, which leave marks inside the firearms. Second, when examining bullets or cartridge cases, examiners look at "toolmarks"

The final premise of the theory is, as Alessio explained, that the individual marks left on each firearm and, correspondingly, the resulting marks on each bullet or cartridge case fired from each firearm, are different "enough that a trained examiner can differentiate between" a bullet or cartridge case fired from one firearm and a bullet or cartridge case fired by the next firearm manufactured by the same tool—and, by extension, a bullet or cartridge case fired by any other firearm.

The AFTE has defined the theory of identification as follows:

> "1. The theory of identification as it pertains to the comparison of toolmarks enables opinions of common origin to be made when the unique surface of two toolmarks are in 'sufficient agreement.'
>
> "2. This 'sufficient agreement' is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compare to the corresponding features in the second set of surface contours. Agreement is significant when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that 'sufficient agreement' exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.
>
> "3. Currently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner's training and experience."

---

on the bullets or cartridge cases, which are the marks left by the "tool"—*i.e.* firearm—that fired the item. The "toolmarks" on the fired item are made, in part, by the toolmarks left inside the firearm by the tools used to manufacture the firearm.

Association of Firearm and Tool Mark Examiners, *Theory of Identification as it Relates to Tool Marks: Revised*, 43 AFTE Journal 287 (2011) (*AFTE Revised Theory of Identification*).

Several concepts are helpful to an effort to understand that definition. First is the best known non-match. Each trained examiner has their own best known non-match, which is, as Todd explained, "the most agreement that you've ever seen in two [bullets or cartridge cases] that you know came from different [firearms]." The best known non-match is based on the examiner's "personal identification criteria," which each examiner develops subjectively through their training: As Alessio explained, any fully trained examiner has "seen enough matching and non-matching marks to develop their own personal identification criteria." Those criteria exist in the mind of each examiner, and the AFTE method does not require the examiner to articulate or apply them in any objective way. In the definition quoted above, "the best agreement demonstrated between toolmarks known to have been produced by different tools" refers to the examiner's best known non-match.

The next useful term is best known match. Unlike the best known non-match, which the examiner relies on in every case, the best known match is case specific, although it still requires application of the examiner's personal identification criteria. The AFTE method requires the examiner to test-fire the firearm to obtain "knowns," which are bullets or cartridge cases that the examiner knows were fired from the firearm in question. As Todd testified, the examiner must "identify what the pattern is for that particular firearm." To do that, the examiner puts "two of the known standards on the comparison microscope, one on each side," and they "compare the individual characteristics, or when they're oriented in the same orientation, we're going to find the pattern, and we're going to basically do what's known as a pattern match." The best known match is the examiner's assessment of the amount of agreement that the examiner finds between the patterns on two knowns test fired from the firearm in question. The examiner documents the parts of the pattern that they find significant for the best known match analysis, but, at the OEC 104 hearing, neither Todd nor Alessio testified

to any objective criteria for measuring the amount of similarity between two knowns. Rather, the examiner relies on their subjective personal identification criteria to identify the degree of agreement that they find between the knowns so that they can mentally compare that degree of agreement to the degree of agreement they see between the knowns and the unknowns. In the AFTE theory definition quoted above, "agreement demonstrated by toolmarks known to have been produced by the same tool" refers to the best known match.

The concepts of the best known non-match and the best known match suggest a meaning for the concept of "sufficient agreement," which is central to the AFTE theory but whose meaning is not fully articulated by the definition set out above. The definition expressly states the consequences of the examiner finding sufficient agreement: "Sufficient agreement" on "the unique surface of two toolmarks" "enables opinions of common origin to be made." That is, when the marks on two items are in sufficient agreement, an examiner may opine that the marks were created by the same firearm and, thus, that that firearm fired the unknown bullets or cartridge cases. The definition further explains that "[t]he statement that 'sufficient agreement' exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility." In other words, the reason that the examiner can opine on the common origin of the marks is because, when the examiner sees sufficient agreement, they consider it practically impossible that two different tools made the marks.

Although sufficient agreement is the linchpin of the AFTE theory—it is the standard by which an examiner reaches their conclusion—the AFTE's theory does not indicate how an examiner is to decide whether sufficient agreement exists.[7] However, separately, in its glossary, the AFTE has

[7] Confoundingly, the AFTE theory's standard for identification is "sufficient agreement," but the theory discusses the degree of correspondence that an examiner finds between toolmarks in terms of *significance* rather than sufficiency:

> "Agreement is *significant* when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool."

defined "identification" in a way that indicates that agreement is "sufficient" when it exceeds the examiner's best known non-match and is consistent with the best known match: "Identification" means "[a]greement of all discernible class characteristics and sufficient agreement of a combination of individual characteristics where the extent of agreement exceeds that which can occur in the comparison of toolmarks made by different tools and is consistent with the agreement demonstrated by toolmarks known to have been produced by the same tool." Association of Firearm and Tool Mark Examiners, *AFTE Glossary*, 94 (6th ed 2017), *available at* https://afte.org/wp-content/uploads/2024/11/AFTE_Glossary_Version_6.091922_FINAL_COPYRIGHT.pdf (accessed May 5, 2025) (defining "Range of Conclusions Possible When Comparing Toolmarks"). In this case, both Alessio and Todd testified that sufficient agreement occurs when the marks on a bullet or shell case exceed the best known non-match and are consistent with the best known match.

2. *The AFTE method*

With that background, we turn to the AFTE method, which implements the AFTE theory by providing steps that examiners follow to decide, based on the markings on bullets and cartridge cases, whether they were fired from a particular firearm (referred to as a "match" or "identification" with the firearm) or not fired from the particular firearm (referred to as a "non-match" or "exclusion"). In addition to a match or a non-match, the examiner may reach a result of "inconclusive" or, based on the initial examination of the item, which may be damaged, "unsuitable for comparison."

Todd testified to the steps of the AFTE method. The examiner begins by analyzing the unknowns, that is, the bullets or shell cases that the examiner wants to identify as having been fired from a particular firearm or not. The examiner documents the condition of the items and their class characteristics, which include the caliber of the bullets and the types of marks that will be left on all bullets or cases fired from some particular type or brand of firearm.[8]

AFTE Revised Theory of Identification, 43 AFTE Journal at 287 (emphasis added).

[8] The method also recognizes subclass characteristics, which are "common to a group of guns within a certain make or model, such as those manufactured

Then the examiner creates the knowns by test firing the firearm in question. Once the knowns are created, the examiner notes their class characteristics. If they do not match the class characteristics of the unknowns, the result is an exclusion.

Todd explained that, if all of the discernible class characteristics are in agreement, the examiner moves on to a microscopic comparison. The examiner begins by comparing the two test fires under a comparison microscope to establish the best known match, which, as noted above, is case specific. The examiner places two knowns on the microscope, which allows the examiner to view them side by side. Todd explained the process of pattern matching:

> "So a collection of features would be a pattern, if we're talking about those striated marks that kind of look like a bar code. Then you have a series of lines; they're going to have high spots and low spots which translate visually as bright or dark from where the light is shining and catching on that metal.
>
> "And then you're also going to have different widths. There might even be some movement to the line, so instead of being perfectly straight they might have some—some wiggle to them, some curve to them. And so that's what we're searching for is a pattern, and then we're trying to identify that same pattern in the same location on the other known.
>
> "Once we have that pattern matched up, then you would continue comparing the rest of that surface area to establish how much more agreement or if there's additional patterns that are present that would be unique to that firearm. And that's basically establishing your best known match for this particular firearm."

Once the examiner has established the best known match, they swap one of the knowns with an unknown, orient the items the same way, and consider whether the pattern observed on the knowns is present on the unknowns.

---

at a particular time and place. An example would include imperfections on a rifling tool that imparts similar toolmarks on a number of barrels before being modified either through use or refinishing." *Abruquah v. State*, 483 Md 637, 658-59, 296 A3d 961, 974 (2023) (internal quotation marks and citations omitted). Distinguishing subclass characteristics from individual ones can be difficult.

The next step, which occurs at the same time as the comparison step, is for the examiner to evaluate the agreement between the knowns and the unknown. Todd explained that she considers whether the amount of agreement she sees is consistent with the best known match and exceeds her own personal best known non-match:

> "[W]hat I'm looking for, for the AFTE theory of identification, I am—in order to say that it's a common source, I have to have agreement that was consistent with what I established when I was looking at my knowns. And I have to have agreement that exceeds what's known as the best known non-match. And that is something that's established through training and experience as well as all the research that's been done in the field. So as I'm agreeing—as I'm evaluating these things and establishing how much agreement I see, then I'm trying to meet those two things, basically answer those two questions: Does it exceed your best known non-match and does it—is it consistent with your best known match?"

Using that process, the examiner reaches a conclusion about whether there is sufficient agreement between the knowns and the unknown. Consistent with the definition of the AFTE theory, which explains that "[c]urrently the interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner's training and experience," *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287, Alessio testified at the hearing that, although the marks that the examiner is comparing are objectively observable, the ultimate determination—that is, the comparison of the degree to which the marks match—is subjective:

> "[PROSECUTOR:] How much of the AFTE technique involves documenting objectively observable marks?
>
> "[ALESSIO:] Most of it. It's *** like I talked about—those class characteristics, those measurable features. Then we get to those individual marks. But those individual marks are also measurable features that are reproducible in the fact that they occur in the same place, in the same frequency, the same position, and can be repeated by creating extra test fires to show that those measurable features are there and being reproduced over and over and

> over again. So I think it's extremely objective until you get to the last part where a human examiner has to render a decision."

Thus, the subjective part is the examiner's evaluation of and decision on whether the level of agreement between the knowns and the unknown exceeds the examiner's best known non-match and is consistent with the best known match for the firearm in question.

The final step of the AFTE method is verification. A second examiner repeats the process and reaches a conclusion. If it is the same as the first examiner's conclusion, the result is ready for reporting.

In addition to providing testimony, both parties also cited scientific literature on the AFTE theory and method. The state introduced several studies and papers, most published in the AFTE Journal, concluding that the AFTE method is "an accurate and precise method for determining a common source for bullets and cartridge case[s] for firearms collected from casework." Erich Smith, *Cartridge Case and Bullet Comparison Validation Study with Firearms Submitted in Casework*, 36 AFTE Journal, 130, 132 (2004); *see also, e.g.*, Angela Stroman, *Empirically Determined Frequency of Error in Cartridge Case Examinations Using a Declared Double-Blind Format*, 46 AFTE Journal, 157, 160 (2014).

Defendant cited a 2009 report from the National Research Council of the National Academy of Sciences (NRC) and a 2016 report from the President's Council of Advisors on Science and Technology (PCAST). *See* NRC, *Strengthening Forensic Science in the United States: A Path Forward* (2009) (*NRC Report*), *available at* https://www.ojp.gov/pdffiles1/nij/grants/228091.pdf (accessed May 5, 2025); PCAST, *Report to the President, Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (2016) (*PCAST Report*), *available at* https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed Apr 30, 2025). The *NRC Report* concluded that "[s]ufficient studies have not been done to understand the reliability and repeatability of the methods" of toolmark identification, *NRC*

Report at 154—that is, as the *PCAST Report* paraphrased, the NRC concluded that "the foundational validity of the field had not been established." *PCAST Report* at 105. The *NRC Report* noted that the AFTE Theory "is the best guidance available for the field of toolmark identification, [but] does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence." *NRC Report* at 155; *see also id.* at 153-54 (under the AFTE theory, "the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates").

Similarly, the *PCAST Report* provided a view from the scientific community on the topic of a court's ultimate determination under *Daubert* and *O'Key*, which is "'the scientific validity—and thus the evidentiary relevance and reliability—of the principles that underlie a proposed submission.'" *O'Key*, 321 Or at 305 (quoting *Daubert*, 509 US at 594-95); *accord PCAST Report* at 42 (same). The *PCAST Report* explained that, in light of the *Daubert* inquiry, "[i]t is the proper province of the scientific community to provide guidance concerning scientific standards for scientific validity." Ultimately, the report concluded that,

> "To establish foundational validity for a forensic feature-comparison method, the following elements are required:
>
> "(a) a reproducible and consistent procedure for (i) identifying features in evidence samples; (ii) comparing the features in two samples; and (iii) determining, based on the similarity between the features in two sets of features, whether the samples should be declared to be likely to come from the same source ('matching rule'); and
>
> "(b) empirical estimates, from appropriately designed studies from multiple groups, that establish (i) the method's false positive rate—that is, the probability it declares a proposed identification between samples that actually come from different sources and (ii) the method's sensitivity—that is, the probability it declares a[n] *** identification between samples that actually come from the same source."

Id. at 65.

The *PCAST Report* distinguished between objective and subjective feature-matching methods. It took the view that subjective methods can theoretically be valid for *Daubert* purposes, but that they require particularly rigorous empirical validation to avoid presenting subjective conclusions based on methods that have not been empirically verified—and thus that have unknown and unknowable probative value—as being based on science:

> "For subjective methods, procedures must still be carefully defined—but they involve substantial human judgment. For example, different examiners may recognize or focus on different features, may attach different importance to the same features, and may have different criteria for declaring proposed identifications. Because the procedures for feature identification, the matching rule, and frequency determinations about features are not objectively specified, the overall procedure must be treated as a kind of 'black box' inside the examiner's head.
>
> "Subjective methods require careful scrutiny, more generally, their heavy reliance on human judgment means that they are especially vulnerable to human error, inconsistency across examiners, and cognitive bias. In the forensic feature-comparison disciplines, cognitive bias includes the phenomena that, in certain settings, humans (1) may tend naturally to focus on similarities between samples and discount differences and (2) may also be influenced by extraneous information and external pressures about a case."

PCAST Report at 49.[9]

---

[9] The report explained that, while some types of subjective judgments may be scientific absent empirical validation—for example, "[b]ased on experience, a surgeon might be scientifically qualified to offer a judgment about whether another doctor acted appropriately in the operating theatre"—opinions based on measurement and comparison of physical qualities (like AFTE identifications) require empirical validation in order to have any probative value at all:

> "'[E]xperience' or 'judgment' cannot be used to establish the scientific validity and reliability of a metrological method, such as a forensic feature-comparison method. The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, is not a matter of 'judgment.' It is an empirical matter for which only empirical evidence is relevant."

Id. at 55.

The *PCAST Report* summarized the requirements for scientific validity and reliability: "For a metrological method [like the AFTE method] to be scientifically valid and reliable, the procedures that comprise it must be shown, based on empirical studies, to be repeatable, reproducible, and accurate, at levels that have been measured and are appropriate to the intended application." *Id.* at 47. It defined the relevant terms:

> "By 'repeatable,' we mean that, with known probability, an examiner obtains the same result, when analyzing samples from the same sources. By 'reproducible,' we mean that, with known probability, different examiners obtain the same result, when analyzing the same samples. By 'accurate,' we mean that, with known probabilities, an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives). By 'reliability,' we mean repeatability, reproducibility, and accuracy."

Id.

Within that framework, the *PCAST Report* further developed the *NRC Report*'s critique of firearms identification, the AFTE theory, and the research that purported to support it. With respect to the existing research studies, it explained,

> "Although firearms analysis has been used for many decades, only relatively recently has its validity been subjected to meaningful empirical testing. Over the past 15 years, the field has undertaken a number of studies that have sought to estimate the accuracy of examiners' conclusions. While the results demonstrate that examiners can under some circumstances identify the source of fired ammunition, many of the studies were not appropriate for assessing scientific validity and estimating the reliability because they employed artificial designs that differ in important ways from the problems faced in casework."

PCAST Report at 106.[10] After discussing the research in more detail and explaining the problems with the design of

[10] The report raised several issues regarding the design of existing studies, but the most pervasive was that, other than the Ames Laboratory study, which we describe below, the existing studies were not "black box" studies because they did not require examiners to make multiple *independent* identification decisions.

all but one of the existing studies, the report ultimately concluded that, with respect to the *Daubert* standard, "firearms analysis currently falls short of the criteria for foundational validity." *Id.* at 112. It elaborated:

> "The early studies indicate that examiners can, under some circumstances, associate ammunition with the gun from which it was fired. However, as described above, most of these studies involved designs that are not appropriate for assessing the scientific validity or estimating the reliability of the method as practiced. Indeed, comparison of the studies suggests that, because of their design, many frequently cited studies seriously underestimate the false positive rate.
>
> "At present, there is only a single study that was appropriately designed to test foundational validity and estimate reliability (Ames Laboratory study). Importantly, the study was conducted by an independent group, unaffiliated with a crime laboratory. Although the report is available on the web, it has not yet been subjected to peer review and publication.
>
> "The scientific criteria for foundational validity require appropriately designed studies by *more than one group* to ensure reproducibility. Because there has been only a single appropriately designed study, the current evidence falls short of the scientific criteria for foundational validity. There is thus a need for additional, appropriately designed black-box studies to provide estimates of reliability."

PCAST Report at 111 (emphasis in original; internal footnote omitted).

As noted in the foregoing quote, the *PCAST Report* identified and described a 2014 "black box" study conducted by the Ames Laboratory at Iowa State University. David P. Baldwin, Stanley J. Bajic, Max Morris, and Daniel S. Zamzow, *A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons*, Ames Laboratory-USDOE

PCAST Report at 106. Studies of the AFTE method commonly use closed or partially open sets of items, requiring examiners to conduct a matching exercise in which each match or exclusion that they identify narrows the pool of other possible matches and exclusions. In other words, as the *PCAST Report* noted, those studies work similarly to a game of Sudoku, "where initial answers can be used to help fill in subsequent answers." *Id.* That structure makes those studies very dissimilar to real-world case work.

Technical Report #IS-5207 (2014) (*Ames I*), *available at* https://www.osti.gov/servlets/purl/1987636 (accessed Apr 30, 2025). The *PCAST Report* generally approved of the design of *Ames I*, although it noted some ways in which the study conditions still were dissimilar to case work. It also noted that, although all the participating examiners were members of the AFTE who had been suitably trained and employed in law enforcement, the results suggested that "the false positive rate is highly heterogeneous across the examiners." *PCAST Report* at 110.[11]

In 2020, a report on a second Ames Laboratory black-box validation study of the AFTE method conducted in response to the *PCAST Report* was released. Stanley J. Bajic, L. Scott Chumbley, Max Morris, and Daniel Zamzow, Ames Laboratory-USDOE Technical Report #ISTR-5220, *Report: Validation Study of the Accuracy, Repeatability, and Reproducibility of Firearm Comparisons* (2020) (*Ames II*).[12] In that study, 173 examiners conducted 20,130 comparisons, some of bullets and some of cartridge cases. *Id*. at 2. In keeping with the AFTE Range of Conclusions, the examiners were asked to classify bullets and cartridge cases using four conclusions: "Identification, Inconclusive, Elimination, and Unsuitable, with three possibilities for qualifying an Inconclusive decision." *Id.* at 13. The researchers divided the comparisons, allocating them "to assess accuracy (8,640), repeatability (5,700), and reproducibility (5,790) of the evaluations made by participating examiners." L.S. Chumbley, M.D. Morris, S.J. Bajic, D. Zamzow, E. Smith, K. Monson,

---

[11] The report summarized the results of the *Ames I* study:

> "Among the 2178 different-source comparisons, there were 1421 eliminations, 735 inconclusives and 22 false positives. The inconclusive rate was 33.7 percent and the false positive rate among conclusive examinations was 1.5 percent (upper 95 percent confidence interval 2.2 percent). The false positive rate corresponds to an estimated rate of 1 error in 66 cases, with upper bound being 1 in 46. (It should be noted that 20 of the 22 false positives were made by just 5 of the 218 examiners—strongly suggesting that the false positive rate is highly heterogeneous across the examiners.)"

See also Ames I at 11 (noting that "each examiner has his or her own false identification probability, and *** these probabilities vary substantially").

[12] Like the *Ames I* study, it does not appear that the *Ames II* study has ever been published or peer reviewed. Defendant's expert testified that *Ames II* was available on the Ames Laboratory Department of Energy website for some period of time but is no longer available there.

G. Peters, *Accuracy, Repeatability, and Reproducibility of Firearm Comparisons Part 1: Accuracy* 1 (2021) (*Accuracy Report*). As to accuracy, the study had rates of false positives around 2 percent and sensitivity (the probability that examiners identify as a match a pair that comes from the same source) of around 75 percent.[13] Alan H. Dorfman & Richard Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives*, 5 Forensic Sci Int'l: Synergy, 100273, Table 3 (2022), *available at* https://www.sciencedirect.com/science/article/pii/S2589871X22000584?via%3Dihub (accessed May 14, 2025); *see Ames II* at 34, Table V (providing data); *Accuracy Report* at 19-20 (discussing sensitivity). The rate of false negatives was somewhat higher, 3.6 percent for bullets and 2.3 percent for cartridge cases. Dorfman & Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives*, 5 Forensic Sci Int'l: Synergy, 100273, at Table 3.

---

[13] The *Ames II* study report included the inconclusive results in the denominator of its calculation of the false positive rates rather than calculating the rates based only on the conclusive results, as the *PCAST Report* explained is necessary. *PCAST Report* at 51-52 ("When reporting a false positive rate to a jury, it is scientifically important to calculate the rate based on the proportion of conclusive examinations, rather than just the proportion of all examinations. This is appropriate because evidence used against a defendant will typically be based on conclusive, rather than inconclusive, examinations. To illustrate the point, consider an extreme case in which a method had been tested 1000 times and found to yield 990 inconclusive results, 10 false positives, and no correct results. It would be misleading to report that the false positive rate was 1 percent (10/1000 examinations). Rather, one should report that 100 percent of the conclusive results were false positives (10/10 examinations).")

We agree with PCAST that the better method of calculation is to disregard inconclusives rather than including them in the denominator for false positive rates. All of the information necessary to recalculate the rate of false positives and negatives compared to the number of conclusive results is in the *Ames II* study report; thus, we rely on the method identified as scientifically appropriate by PCAST. *See* Alan H. Dorfman & Richard Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives*, 5 Forensic Sci Int'l: Synergy, 100273, Table 3 (2022), *available at* https://www.sciencedirect.com/science/article/pii/S2589871X22000584?via%3Dihub (accessed May 14, 2025) (recalculating *Ames II* rates of error in a variety of ways; when treating inconclusives as irrelevant, *Ames II* data demonstrates "different source," *i.e.* false positive, error rates of 2.04 percent for bullets and 1.86 percent for cartridge cases and "same source," *i.e.* false negative, error rates as 3.67 percent for bullets and 2.31 percent for cartridge cases). However, we are unable to determine the appropriate confidence interval. *See id.* at 2.2.1 (noting, in any event, that the *Ames II* confidence intervals "despite their sophistication, are not well-grounded").

In the accuracy section of the *Ames II* study, as in *Ames I*, examiners' rates of error varied widely:

> "[H]ard errors were made by 34 of the 173 examiners when examining bullets; 36 of 173 for cartridge cases. Three participants made both kinds of errors. Analysis of the data collected in the first round of the study showed that the six most error-prone examiners account for 33 of the 112 errors—29%[—]while 13 examiners account for almost half of all the hard errors (54 of 112 errors)."

Accuracy Report at 11-12.

The *Ames II* investigators also assessed how often examiners agree with other examiners when they repeat the same comparison without knowledge of the first examiner's conclusion (reproducibility) and how often they agree with themselves when asked to perform the same comparison a second time (repeatability). The results indicated that examiners disagreed with other examiners between 22 percent and 69 percent of the time (depending on whether the items were matches or non-matches; how the three types of inconclusives were grouped; and whether the examined items were bullets or cartridge cases), *id.* at 45-52, and examiners disagreed with themselves between 14 percent and 38 percent of the time, depending on the same variables. *Id.* at 40-45.

3. United States v. Adams

Before the pretrial motions in this case, defendant was tried on a federal charge arising from the same incident at issue here. *See United States v. Adams*, 444 F Supp 3d 1248, 1251 (D Or 2020). In the federal case, the district court, Judge Mosman, granted defendant's motion *in limine* to exclude scientific testimony of an AFTE examiner, Gover, who examined the same firearm and cartridge cases at issue here. The record before the federal district court—primarily testimony from Gover—was somewhat different from the record in this case, as we discuss further below. Notwithstanding the different record, however, we find the district court's reasoning persuasive. Accordingly, we describe it in some detail.

The court began by explaining the relevant federal legal standard for the admission of scientific expert evidence.

It explained that *Daubert* requires scientific testimony to be "reliable." "In order to qualify as science, a proposition 'must be derived by the scientific method.' The expert's assertion need not be verified as an objective certainty, but it must have been derived by a verified scientific process in order to meet the necessary standard of evidentiary reliability." *Adams*, 444 F Supp 3d at 1257 (quoting *Daubert*, 509 US at 593-94; internal citation omitted). *Daubert* sets out five factors that overlap with the factors that we use under *Brown* and *O'Key*, which assist the court in determining whether the scientific evidence is sufficiently reliable to be relevant and helpful to the jury. *O'Key*, 321 Or at 306-07.

Judge Mosman recognized that, "at a superficial level," the AFTE testimony could be viewed as satisfying the *Daubert* requirements:

> "Replicability? Check. Other forensic examiners trained in ballistics comparisons can perform an examination using the same basic methodology. In fact, double check: a forensic examiner in this very case did a second comparison and came to the same conclusion. Error rate? Check. Testing seems to indicate an error rate hovering around two percent. Testing and standards? Check, at least as to some of what is done. Publication? Check. Journals put out by associations of forensic examiners have published hundreds of relevant articles. Acceptance? Check. All reputable forensic examiners accept the ballistics comparison method used in this case as valid."

Adams, 444 F Supp 3d at 1258.

However, the court observed that, "when you look beneath the surface," "the problems with [the AFTE] methodology begin to emerge." *Id.* It explained that "[i]t is worth remembering, at this point, that what we are analyzing here is the admissibility of *scientific* evidence. It is certainly possible to know things—and to know them so well that you are almost never wrong—without resorting to scientific methods." *Id.* (emphasis in original).

> "Imagine, for a moment, a school to train *** art experts. It is given the name, 'The Academy for the Science of Art Examination.' Everyone there is trained to follow a particular protocol, with several objective elements of

> scientific examination: carbon dating, microscopic examination, x-rays, and the like. The Academy establishes a journal, in which its graduates regularly publish papers describing their work in peer reviewed articles. There is regular testing in which examiners are sent two works of art and told to determine which one is a forgery, a test on which the examiners have only a two percent rate of false positives—*i.e.* of misidentifying original works as forgeries. And they all agree that what they are doing is science. The only rub is this: when all the objective work is done, and it comes time to decide whether what they are looking at is fake, they are told to call upon all the experience they have, and the feeling that true art invokes in them, and then make a call by listening to their inner voice telling them what is right. The ultimate support for their conclusions, then, is almost entirely subjective and inscrutable.
>
> "Like any analogy, it does not map perfectly over our facts. But the central point is this: If you have elements of objective examination, and a lot of training, and you form a like-minded group to review each other's work and publish papers, and if you maintain standards to guide the objective elements of the process, and ensure that a certain protocol is followed, and if you are frequently correct in your assessments, then even if you cannot explain how you reach your ultimate conclusions, and even if therefore your process fundamentally is not scientific, you can pass a superficial reading of the *Daubert* factors."

Id. at 1258-59.

The court noted that, even if expert evidence based on a witness's training and experience may be admissible as non-scientific expert evidence, it remains important that evidence not derived from science not be presented to the jury cloaked with the persuasive appeal of science. *Id.* at 1259. Then the court evaluated each of the *Daubert* factors, concluding, "On each factor, when one looks more closely at [Gover's] testimony and considers the underlying purpose of the inquiry, the methodology used in this case has problems." *Id.*

The court began with testability, which includes falsifiability and replicability. *Id.* Although the court determined that the AFTE method is falsifiable—the proposition that a particular gun fired a particular bullet or shell case is capable of being proved false by features that exclude

a match between the two—the court concluded that the method is not replicable, because, as a result of its ultimate subjectivity, it cannot be applied the same way twice:

> "In order to be replicable, a method must be objective enough that someone else not associated with the case could duplicate it and get the same results. That standard is very problematic here. Much of what Mr. Gover did was, in fact, objective. *** But when it comes to his ultimate conclusion of a match, Mr. Gover's inquiry was not objective.
>
> "The first problem is with the baseline comparator that Mr. Gover used—the 'best known non-match' as a comparator. This requires an examiner to compare the degree of correspondence on his split screen against the best possible non-match comparison that anyone is aware of between consecutively manufactured firearms. *** But Mr. Gover could not define this baseline in any objective way, nor could he explain the role it played in the actual comparison he made in this case. He apparently just kept a memory of the baseline in his head and then made a comparison unguided by any objective standards or benchmarks."[14]

Id. at 1260 (internal citation omitted). After providing an excerpt of Gover's testimony illustrating the problem, the court continued,

> "With no objective benchmark below which we would know that a shell casing was not fired from a particular firearm ***, replicability begins to fall apart.
>
> "Throughout Mr. Gover's testimony, this lack of any objective standard became increasingly evident. It is worth explaining at the outset that this flaw in Mr. Gover's testimony appears to derive from the AFTE methodology itself, not from any deficiency in Mr. Gover's understanding of his work. AFTE requires an examiner to find 'sufficient agreement' between crime scene shells and test fired shells from the firearm in question in order to determine a match.

[14] The district court seems to have considered the best known non-match to be the closest degree of agreement that any examiner anywhere has ever seen between items fired from different firearms. The record in our case is clear, however, that the best known non-match is the closest degree of agreement that the particular examiner has ever seen between items fired from different firearms. That understanding is consistent with Judge Mosman's observation that Gover "kept a memory of the baseline in his head." That is what Todd testified that she does as well—her subjective best known non-match is a product of her training and experience, and it is not something that can be articulated or produced in written form.

Here is Judge Garaufis of the Eastern District of New York explaining why 'sufficient agreement' is not an objective standard:

"'First, the sufficient agreement standard is circular and subjective. Reduced to its simplest terms, the AFTE Theory "declares that an examiner may state that two toolmarks have a 'common origin' when their features are in 'sufficient agreement.'" *PCAST Report* at 60. "It then defines 'sufficient agreement' as occurring when the examiner considers it a 'practical impossibility' that the toolmarks have different origins." *Id.* The NRC Report notes that the AFTE Theory "is the best guidance available for the field of toolmark identification, [but] does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence." *NRC Report* at 155. Without guidance as to the extent of commonality necessary to find "sufficient agreement," the AFTE Theory instructs examiners to draw identification conclusions from what is essentially a hunch—a hunch "based on the examiner's training and experience," *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287—but still a hunch.

"'Moreover, the application of this circular standard is "subjective in nature *** based on the examiner's training and experience." *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287. Ostensibly, one hundred firearms toolmark examiners could hold one hundred different personal standards of when two sets of toolmarks sufficiently agree, and all one hundred of these personal standards may accord with the AFTE Theory. Further, because the standard itself offers so little guidance on when an examiner should make an identification determination, some examiners may decide that the two sets of toolmarks were made by the same tool while others determine the toolmarks to be inconclusive and still others decide the toolmarks were made by different tools. To emphasize, these one hundred examiners could come to these contradictory conclusions without a single examiner running afoul of the AFTE Theory.'

"*United States v. Shipp*, No. 19-cr-029-NGG, 2019 WL 6329658 *13 (E.D.N.Y. Nov. 26, 2019).

"In other words, the AFTE 'sufficient agreement' standard is a tautology that doesn't *mean* anything."

Id. at 1261-62 (brackets and second omission in *Shipp*; emphasis in *Adams*; internal citation omitted).

After again excerpting Gover's testimony, the court continued:

> "Mr. Gover could not say that the person who checked his work, and who relied on the same methodology, applied the same standard in reaching the same conclusion. He could not be sure what threshold [the second examiner] used to decide that the shell casings were fired from the Taurus—*i.e.* that Mr. Gover's conclusion was correct. Not only is the AFTE method not replicable for an outsider to the method, but it is not replicable between trained members of AFTE who are using the same means of testing.
>
> "If this were truly a scientific inquiry, [Gover's testimony indicating that he could not conclusively know what standard another examiner applied] would not be possible. If a cancer researcher sought a second opinion from another cancer researcher in order to reach a diagnosis, both people would be able to say with certainty what the other person was looking for and why. If their conclusions deviated, they would be able to pinpoint the points of disagreement and why those data points were meaningful.
>
> "Over and over, Mr. Gover failed to do this. He could not explain which data points he looked at or why they were meaningful to him. And this is not purely a fault of Mr. Gover. There is no evidence in this record or elsewhere that the AFTE method relies on any scientific standard that would explain to an examiner like Mr. Gover how to interpret the data he sees in any kind of objective way. What he is actually doing is applying his training and experience to make a subjective conclusion about what he sees before him, just like the art expert in [the earlier-described] example. The AFTE method is therefore not replicable—and not testable—because it cannot be explained in a way that would allow an uninitiated person to perform the same test in the same way that Mr. Gover did. This factor weighs heavily against admissibility under *Daubert*."

Id. at 1263-64 (footnote omitted).[15]

[15] We understand the court's reference to an "uninitiated person" to mean someone seeking to apply the AFTE method without resorting to unarticulated and unmeasurable standards drawn from their own training and experience—not someone entirely untrained in the method.

After concluding that the AFTE method is not replicable, the court considered error rates. *Id*. at 1264. Treating the rate of false positives as the relevant error rate, it noted that the error rates shown for the AFTE method vary depending on the type of study, with black box studies—which reduce confounding variables better than the other types of studies but still do not mimic real-world conditions[16]—showing a rate of false positives of 2.2 percent. *Id*. at 1264-65. Given the significance of AFTE identification testimony for convictions, the court pointed out that that error rate could be expected to result in a wrongful conviction in one of every 46 cases. *Id.* at 1264. It also noted that, because of the design of the relevant studies, real-world error rates may significantly exceed those reported in the studies:

> "The incentive structure for the testing process is also concerning. It appears to be the case that the only way to do poorly on a test of the AFTE method is to record a false positive. There seems to be no real negative consequence for reaching an answer of inconclusive. Since the test takers know this, and know they are being tested, it at least incentivizes a rate of false positives that is lower than real world results. This may mean the error rate is lower from testing than in real world examinations."

Id. at 1265 (footnote omitted). That is, in the studies, when an examiner finds a close case, they are incentivized to record it as inconclusive—a result that is effectively counted as a correct result in many studies—rather than a match, which, if incorrect, will affect the error rate in the study. In real-world conditions, that incentive is absent: There is no way to check rates of false positives, and, as the record in this case indicates, examiners do not keep records of their

[16] In particular, the then-existing black box studies were not "test-blind," that is, the examiners were aware that they were being tested rather than doing routine case work. *See* Dorfman & Valliant, *Inconclusives, Errors, and Error Rates in Forensic Firearms Analysis: Three Statistical Perspectives*, 5 Forensic Science International: Synergy, 100273, at § 2.3; *see also PCAST Report* at 59 ("PCAST believes that test-blind proficiency testing of forensic examiners should be vigorously pursued, with the expectation that it should be in wide use, at least in large laboratories, within the next five years.").

More recently, at least one small test-blind study has been conducted. *See* Maddisen Neuman *et al*., *Blind Testing in Firearms: Preliminary Results from a Blind Quality Control Program*, 67 J Forensic Sci 964 (2022). In that study, 11 examiners made no false identifications or false exclusions in the course of 570 comparisons that were presented as if they were ordinary cases.

results (which, if they were kept, could show whether rates of inconclusive results are lower in case work than in studies). Ultimately, the court found the rate of error to weigh against admission. *Id.*

Moving on to the third *Daubert* factor, peer review, the court noted the purpose of that factor in the overall inquiry:

> "The question of whether a methodology has been subjected to peer review is a question of whether a methodology or hypothesis has been published to the scientific community for the purpose of detecting substantive flaws. *Daubert*, 509 US at 593. If a methodology has been published but not for this purpose, this factor is not satisfied in favor of admissibility."

Id.

The court rejected the idea that the AFTE Journal, which the government argued shows that the method is subject to peer review, satisfies that factor for two reasons: because the AFTE Journal "is a trade publication, meant only for industry insiders, not the scientific community," and, more importantly, because "the purpose of publication in the AFTE Journal is not to review the methodology for flaws but to review studies for their adherence to the methodology." *Id.* at 1265-66. That "is not the purpose that *Daubert* sets out for peer review." *Id.* at 1266.

The court held that the "standards and quality control" factor was satisfied; the record before it showed that there were standards and quality control measures in place, although that did not mitigate the fundamental problem with subjectivity and consequent lack of replicability. *Id.* As to the last factor, "general acceptance in the broader scientific community," the court observed,

> "The AFTE method that Mr. Gover uses has been widely accepted within his own community of technical experts. But it has been heavily criticized by other members of the broader scientific community for failing to yield reproducible results or a precisely defined process. In fact, [the *NRC Report* and *PCAST Report*] suggest to me that the widespread acceptance within the law enforcement community may have created a feedback loop that has inhibited the

> AFTE method from being further developed. *** Here, where the scientific community at large disavows the theory because it does not meet the parameters of science, I cannot find that the AFTE method enjoys 'general acceptance' in the scientific community."

Id. (internal citations omitted).[17]

In conclusion, the court noted that comparison analysis like the AFTE method, undertaken by a trained examiner, may be effective at identifying matches, but the problem is that, from what was in the record before the court, the analysis is based on training and experience—ultimately, hunches—not science:

> "Even at its worst, comparison analysis has a very low rate of error and yields results that cannot be random. But it is not clear that those results are the product of a *scientific* inquiry. Nothing in Mr. Gover's testimony explains how or why he reached his conclusion in any quantifiable, replicable way. It is possible that the AFTE method could be expressed in scientific terms, but I have not seen it done in this case, nor elsewhere."

Id. at 1266-67 (emphasis in original). Accordingly, the court granted defendant's motion to exclude the AFTE identification evidence. *Id.* at 1267.

4. Brown *and* O'Key *factors*

Overall, we find Judge Mosman's analysis highly persuasive, and we adopt the portions discussed in this opinion for purposes of our own analysis under *Brown* and *O'Key*. We see no need to expound further on the factors of peer review, standards and safeguards, and the technique's acceptance in the scientific community. We provide some additional discussion, however, regarding testability (and the overlapping *Brown* factor of the extent to which the technique relies on the subjective interpretation of the expert) and error rates, as well as *Brown*'s additional factors of the existence of specialized literature and the novelty of the invention.

---

[17] *Accord PCAST Report* at 55 ("[S]cientific validity of a method must be assessed within the framework of the broader scientific field of which it is a part (*e.g.*, measurement science in the case of feature-comparison methods). The fact that bitemark examiners defend the validity of bitemark examination means little.").

Initially, we note that, although Judge Mosman—and, in this opinion, we—occasionally refer to the fundamental problem with the AFTE method as relating to subjectivity, we understand the problem to go beyond the fact that the expert must apply some judgment in the course of reaching a conclusion. As the *PCAST Report* suggests, a method may be scientifically valid—assuming that the method's error rate and sensitivity can be empirically shown—even if it requires some degree of subjective judgment by the expert. *PCAST Report* at 49. Here, however, the problem is not that the examiner must apply some judgment in reaching their conclusion; rather, it is that, because there are *no* objective standards or criteria at all—the only question is whether, based on the particular examiner's past experience and memory of the pattern replications between knowns, there is a match—the procedure is not "reproducible" or "consistent"; each examiner who applies it is actually doing something different. *Id.* at 48 (a scientifically valid method requires "a reproducible and consistent procedure" for identifying features, comparing features, and determining whether the samples match).

As noted above, the record in this case differs in some ways from the record before the federal district court in *Adams*. We discuss those differences as they relate to the *O'Key*/*Daubert* factors and the factors identified in *Brown*. As explained below, we conclude that the state has failed to show that the AFTE identification evidence is scientifically valid.

i. Testability/Replicability and the Extent to Which the Technique Relies on the Subjective Interpretation of the Expert

As Judge Mosman noted, in the federal case, "Mr. Gover could not say that the person who checked his work, and who relied on the same methodology, applied the same standard in reaching the same conclusion." *Id.* at 1263. That was a problem because scientific evidence should be based on a method or analysis applied in a quantifiable way, such that two people who have done it correctly will have done the same thing. *Accord O'Key*, 321 Or at 301 n 19 (in order to be

"reliable," a method should "produce consistent results when replicated").

In this case, Alessio testified that, in his view, fully trained examiners do use the same standards, and that is evidenced through the low error rates shown in validation studies conducted over many years:

> "[PROSECUTOR:] And do all examiners correctly applying [the] AFTE methodology apply the same standard for declaring something as a match?
>
> "[ALESSIO:] Yes. I think it's—like the validation studies and the low error rate show that with the proper training—
>
> "*****
>
> "[ALESSIO:] *** I think the validation studies show that firearm examiners with the proper training, that rigorous training, when they've developed that personal identification criteria, can very accurately determine common-source origins of tool marks, and the validation studies show that with the very low error rate in the false positives.
>
> "*****
>
> "[PROSECUTOR:] Would other examiners correctly applying the AFTE methodology agree when a level of agreement exceeds a best known non-match?
>
> "[ALESSIO:] Yes.
>
> "*****
>
> "[PROSECUTOR:] Even though there's no numeric attached, why is the standard of exceeds the best known non-match able to be consistently applied amongst examiners?
>
> "[ALESSIO] Again, I think it goes back to those validation studies. We—we tested it and—
>
> "*****
>
> "So just the low error rates, the misidentifications, the false positives are really what we're concerned with. Did somebody identify something to a gun that didn't fire it? And those low error rates over all the years of the studies show that that training method and that the developing of

> their personal criteria and how close it is from examiner to examiner with the proper training can be."

We understand the state to suggest that the district court's reasoning—that the AFTE method is not replicable because each examiner applies their own subjective and unarticulated standards rather than uniform objective standards—carries less force in this case because of that difference between the records. We disagree.

Although Alessio testified that all trained examiners correctly applying the methodology use the same standards for identifying a match, he also readily acknowledged that those standards are subjective and unquantifiable. Nor could he reasonably testify otherwise, as the AFTE theory expressly notes the subjectivity of the examiner's determination, *AFTE Revised Theory of Identification*, 43 AFTE Journal at 287, and the AFTE has not articulated any quantifiable standard.

The subjectivity and lack of quantifiability of the standards undermines Alessio's inference that, because studies have indicated that examiners often reach correct results, all examiners must be applying the same standard. As Judge Mosman noted, when practitioners of a replicable method disagree as to the result in a given case, the reason for the disagreement can be identified—with the consequence, in a forensic context, that the trier of fact can assess the persuasiveness of the various views. That is not the case with the AFTE method. If two examiners were to disagree on whether something was a match, each would only be able to explain that, according to their personal identification criteria, the degree of agreement was or was not sufficient to result in a match. *See Shipp*, 2019 WL 6329658 at *13 ("Ostensibly, one hundred firearms toolmark examiners could hold one hundred different personal standards of when two sets of toolmarks sufficiently agree, and all one hundred of these personal standards may accord with the AFTE Theory.").

The same problem confronts the factfinder when there is no disagreement between examiners: Although an examiner may show photographs or point to individual features that do or do not match, the significance of

the similarities or dissimilarities depends entirely on the examiner's subjective criteria, based on their own particular training and experience—to what extent they have seen that degree of similarity or dissimilarity in matches and non-matches in the past.

Indeed, in this case, Todd presented a detailed summary of her comparison process, highlighting numerous similarities that she identified between the knowns and the unknowns. However, the significance of those similarities could only be evaluated through the lens of her personal identification criteria:

> "[PROSECUTOR:] And what area were you looking at here for—or what were you looking at here for a match?
>
> "[TODD:] So again I'm looking at those lines, the striated tool marks that came from the aperture shearing; the spatial relationships of them; the widths of them; the fact that they have correspondence basically from where they start at the bottom on the left side, right above that black area which is just a shadow from that firing pin impression.
>
> "So from the bottom of it, up to the top where it kind of whites out, there's correspondence all the way through that. So that would be consistent with what I saw in my test fires. *So again, if we're always equating this back to evaluating and trying to answer two questions, the first one is, is it consistent with your best known match, which would be your test fires; and the second one is does it exceed our best known non-match.*"

The trier of fact has to simply take the examiner's word for it that the standard is satisfied: The answer to the question "how do you know the agreement is sufficient?" is, ultimately, "I decided that it was sufficient."[18]

---

[18] That situation is particularly problematic because, as the district court noted in *Adams*, "[i]f someone cloaked in the powerful robe of science tells a jury he knows something, and the reason why amounts to 'trust me, I'm a scientist,' then the government gains most of the power of scientific evidence at none of the cost. It becomes very difficult to engage such a witness in meaningful cross examination." 444 F Supp 3d at 1259 (slip op at 15). *Accord PCAST Report* at 53 ("Why is it essential to know a method's false positive rate and sensitivity? Because without appropriate empirical measurement of a method's accuracy, the fact that two samples in a particular case show similar features has no probative value—and, as noted above, it may have considerable prejudicial impact because juries will likely incorrectly attach meaning to the observation.").

The state contends that the fact that an expert's conclusion based on the AFTE method is subjective does not mean that it is not scientifically valid. It relies on the Supreme Court's discussion in *O'Key* of subjectivity in the horizonal gaze nystagmus (HGN) test. 321 Or at 318. There, the court noted that an officer's testimony regarding the officer's observation of the defendant's nystagmus was subjective in the sense that the defendant could not meaningfully dispute the officer's observations:

> "The HGN test is based on subjective visual observation. The officer who administers the test has no physical sample to take to a laboratory. A defendant is not able to have an expert examine the evidence. Test conditions cannot be duplicated, and the test results cannot be verified. Thus, a defendant cannot contradict much of the officer's testimony."

Id. (internal citations omitted).

That type of subjectivity is dispositively different from the problem with subjectivity in the application of the AFTE method. Neither the AFTE theory nor the AFTE method prescribes or quantifies what the examiner is looking for; the examiner is looking for sufficient agreement, which is defined only by their own personal identification criteria. The subjectivity is in every decision implementing the procedure: which features are relevant, how they should be compared, and what they indicate. By contrast, as the court explained in *O'Key*, the HGN test quantitatively measures nystagmus—that is, there is no subjectivity in the determination of the relevance or significance of the movement of the person's eyes. *See* 321 Or at 296 (HGN evidence rests on the proposition "that there is a causal relationship between consumption of alcohol and the type of nystagmus *measured by the HGN test*" (emphasis added)); *id.* at 294-95 (describing the regulations and literature that provide objective standards for measurement). The HGN test is subjective only in the sense that the officer's visual perception of the person's eye movement cannot be recorded or confirmed at a later time—which, the court noted, presents a situation no more problematic than testimony about other observable facts that cannot be independently confirmed later. *Id.* at 318 ("[O]bservation by an officer of the presence of nystagmus is no more subjective than the observation by an officer of other

indicia of alcohol impairment, such as swaying, staggering, having bloodshot eyes, or using slurred speech."). The court did not, as the state contends, approve as scientifically valid a method, like the AFTE method, whose *application* is subjective and cannot be articulated. *Accord State v. Trujillo*, 271 Or App 785, 798, 353 P3d 609, *rev den*, 358 Or 146 (2015) (finding that retrograde extrapolation mainly involves "objective application of a [mathematical] formula" and thus does not require subjective interpretation, which favored admissibility of the scientific evidence); *State v. Reed*, 268 Or App 734, 745-46, 343 P3d 680, *rev den*, 357 Or 551 (2015) (application of an assessment tool to determine ability to consent to sexual conduct required "some subjective evaluation" of the assessed person's responses to questions, but the expert's subjective interpretation was verified by interviews with others).

Returning to Alessio's testimony that, because studies show low error rates, all examiners must apply the same standards, we also disagree with the factual premise underlying his inference. In his view, because studies conducted over many years show that examiners rarely make false positive identifications, all examiners must be using the same standards. But when members of the scientific community have reviewed that body of studies, they have rejected the studies conducted "over many years" that he relied on. In PCAST's words, the studies that existed in 2016, other than *Ames I*, "involved designs that are not appropriate for assessing the scientific validity or estimating the reliability of the method as practiced." *PCAST Report* at 111.

Finally, the data that are directly on point do not support Alessio's view that examiners apply the same standards. As explained above, in the *Ames II* study, investigators found that, upon repeating a comparison, examiners disagreed with other examiners between 22 percent and 69 percent of the time and disagreed with themselves between 14 percent and 38 percent of the time. *Ames II* at 45, 52. Those data indicate that, contrary to Alessio's testimony, different examiners do not reliably use the same standards to determine how something should be categorized.[19]

[19] The data also call into question whether even the same examiner is actually using the same standard across cases.

The state contends that, although those rates show that examiners are not "perfect[ ]," results are "more reproducible and repeatable than 'would be attributed to chance.'" (Quoting *Ames II* at 40). In the state's view, that shows that the examiners' opinions "must be based on something objective," which is a good enough basis on which to provide scientific testimony to jurors.

We disagree. Our ultimate question is whether the method actually *measures* what it purports to measure and whether it reliably produces the same results when replicated, that is, applied to the same items in the same way. The AFTE method does not measure against anything beyond the particular examiner's subjective standards, and for that reason, it also cannot be applied the same way by more than one examiner. Ultimately, it is different to say that a method is valid and replicable—it operates according to a consistent standard that governs the process and decision-making of examiners *and* it reaches accurate, repeatable, and reproducible results—than that a group of practitioners can often make correct identifications based on subjective standards arrived at through training and experience. *Adams*, 444 F Supp 3d at 1263-64 ("What [an AFTE examiner] is actually doing is applying his training and experience to make a subjective conclusion about what he sees before him, just like the art expert in [the earlier-described] example.").

In sum, the testability factor and *Brown*'s additional factor, "the extent to which the technique relies on subjective interpretation of the expert," 297 Or at 436, weigh very heavily against the possibility that AFTE identification evidence is scientifically valid such that it may be presented to the jury as evidence derived from science.

ii. Potential Rates of Error

The state contends that the potential rate of error is low and, consequently, weighs in favor of admitting AFTE identification evidence as scientific evidence. It disagrees with the district court's reasoning in *Adams*, contending that, even considering only the Ames studies, the error rate is low enough to justify admission. It also points out that *Adams* was decided before the *Ames II* study was published and asserts that *Ames*

II is a second black-box study that, now that it exists, fills in as the single additional black-box study that PCAST identified as necessary to validate the AFTE method.

For his part, defendant contends that an additional problem exists with rates of error that goes beyond the issues identified in the *PCAST Report*. In his view, the two Ames studies do not reflect accurate rates of error because of the way they account for results of "inconclusive." He argues that the state has not met its burden to show that the error rate of the AFTE method is low because no study has correctly identified the rate of error for the AFTE method.

Defendant asserts that we cannot rely on the reported error rates in any of the existing studies because none of the studies recognizes or accounts for the fact that some examiners identify as inconclusive pairings that other examiners identify as matches and that, when that occurs, one of the two examiners must be making a mistake. In other words, one examiner cannot be correct that there is insufficient information to arrive at an identification or exclusion if another examiner is correct that there *is* enough information (and, based on that information, reaches a correct conclusion). The Ames studies both focused only on "hard errors," *i.e.* false identifications and false exclusions. As we have calculated their error rates, explained above, they treat answers of "inconclusive"—including answers of "inconclusive" given for pairs that other examiners have counted as identifications or exclusions—as immaterial.

We agree with defendant that the fact that two answers are both treated as correct is problematic, but we view the problem as resulting, at the most basic level, from the AFTE method's inherent subjectivity, that is, the fact that each examiner's standard for sufficient agreement is different. Because each examiner is applying their own version of the AFTE method, with individual error rates and sensitivity, they will sometimes reach different results based on the same comparison. We agree with defendant that that cuts against the conclusion that AFTE evidence is scientific.

However, we do not view that as a problem specific to rates of error.[20] In our view, the central question for the rate-of-error factor is the frequency of false positives—how often an AFTE examiner will testify that a given item was fired from a particular firearm when, in fact, it was not fired from that firearm. On that question, the two Ames studies provide some, albeit imperfect, information about how often examiners make false identifications.

The state points to the two Ames studies as indicating that the rate of false positives of the AFTE method is low. That is somewhat true, as a matter of averages. As the district court observed, the 2.2 percent false positive rate from *Ames I* translates to a potentially wrong conviction in one out of every 46 cases. As we have noted, we are unable to calculate the false positive rate at a 95 percent confidence interval from *Ames II*; however, the observed rate was around 2 percent.

Ultimately, on this record, we conclude that the rate of error factor is neutral. Even if we were to agree with the state that the Ames studies show low rates of error and accept that those rates are representative of average rates of error in case work, because every examiner applies a different standard for the AFTE method—resulting in the fact that, as the Ames studies indicate, each examiner has a different rate of error and even a different error probability—identifying an average error rate across all examiners is not particularly helpful to the trier of fact, who, on this record, lacks information that would allow them to determine where on the error spectrum the testifying examiner falls.[21]

---

[20] We also recognize that, as defendant points out, the problem with incentives that Judge Mosman identified—in a study, examiners are incentivized to classify close cases as inconclusive, thereby avoiding a match or exclusion that may be hard error—makes ignoring inconclusives a problematic study-design choice in any study that is not test blind. Further, we do not disagree that, ultimately, the inconclusives may include some mistakes, where, under an examiner's subjective criteria, there is enough information to make a conclusive determination, but the examiner nevertheless classifies the comparison as inconclusive. As explained in the text, notwithstanding those issues, we find it more practical, given the currently available data, to focus on the rate of false positives.

[21] As the *PCAST Report* observes,

"Demonstrating that an examiner is capable of reliably applying the method is crucial—especially for subjective methods, in which human judgment plays a central role. From a scientific standpoint, the ability to apply a method

iii. The existence of specialized literature

As mentioned previously, both parties cite scientific literature regarding the AFTE method, although the content of that literature—whether it criticizes or lauds the AFTE method—varies. As the Supreme Court explained in *Brown*, this factor does not hinge on the substantive content of the cited specialized literature, but rather on there being "no dearth of literature available on [a] controversial subject." 297 Or at 433. Here, the contrast between the parties' views on the AFTE method is reflected in the cited literature. We find that this factor favors admissibility.

iv. The novelty of the invention

The evidence in this record shows that the basic premise of toolmark comparison has been used since the late 1920s and that the AFTE theory of identification was "adopted in 1992." Given that the method has been formally included in the field of forensic science for over thirty years, we find the AFTE to not be novel, and thus this factor favors admissibility.

5. *AFTE identification evidence is based on practical expertise, not science, and therefore will not assist the jury when it is presented as being based on science.*

Having considered each factor separately, we return to the overall inquiry under OEC 401 and 702, considering "the probative value of the proffered scientific evidence" and "how that evidence might impair rather than help the trier of fact." *O'Key*, 321 Or at 299. The factors outlined in *Brown*

---

reliably can be demonstrated only through empirical testing that measures how often the expert reaches the correct answer."

PCAST Report at 56. That reference to empirical testing appears to contemplate more than the routine proficiency testing that examiners take. *See id.* at 68 (noting that "easy [proficiency] tests are favored by the community, with the result that tests that are too challenging could jeopardize repeat business for a commercial vendor" (internal quotation marks omitted)); *see also* Hofmann, Carriquiry, & Vanderplas, *Treatment of Inconclusives in the AFTE Range of Conclusions*, 19 Law, Probability, and Risk 317, 343 (2020) (proposing use of "decision-specific error rates" in court, which would take into account "examiner-specific probabilities of incorrect conclusions as well as examiner-specific historical decisions"). In this case, Todd testified that she had not made any errors on her competency testing, which occurs before an examiner is certified, and that she had participated in a study (which was a closed-set study) and had not made any errors. She also acknowledged that she did not keep data on her previous comparisons.

and *O'Key* are ultimately nonexhaustive and were meant to be dynamic enough to address the nuances of various proffered scientific techniques. *Id.* at 305 (explaining that the proper inquiry is a "flexible" inquiry into the evidentiary relevance and reliability of the principles underlying the proposed method, "not on the conclusions that they generate"); *see also Brown*, 297 Or at 417-18 ("What is important is not lockstep affirmative findings as to each factor, but analysis of each factor by the court in reaching its decision on the probative value of the evidence under OEC 401 and OEC 702." (Internal footnote omitted.)).

Here, both parties agree that the AFTE theory of identification is relevant in this case, and thus we need not make any further inquiry under OEC 401. However, after applying the principles articulated in *Brown* and *O'Key*, we conclude that AFTE identification evidence in this case was not helpful to the jury. The subjective nature of the AFTE method, and the related fact that each practitioner has a different rate of error, makes it impossible for the court to "ensure that the persuasive appeal" of the evidence to the jury "is legitimate." *O'Key*, 321 Or at 291. We therefore conclude on this record that the state did not carry its burden of showing that the AFTE method is scientifically valid.

C. *Harmlessness*

The state argues that, because the AFTE identification evidence was cumulative of other "more probative" evidence, any error was harmless. We disagree. The state contends that surveillance camera footage from the crime scene, which was presented to the jury, "leave[s] no room for reasonable doubt that defendant is the same person as the shooter" in the footage. Thus, the state argues, "the jury [was] unlikely to have rested its verdict on the toolmark identification." The state's theory of harmlessness would require us to reweigh the evidence, which we may not do. *State v. Davis*, 336 Or 19, 32, 77 P3d 111 (2003) (In reviewing the record to evaluate whether error was harmless, "we do not determine, as a factfinder, whether the defendant is guilty. That inquiry would invite this court to engage improperly in weighing the evidence and, essentially, retrying the case, while disregarding the error committed at trial, to determine

whether the defendant is guilty. Rather, when we review the record, we do so in light of the error at issue. We ask whether there was little likelihood that the error affected the jury's verdict."); *see also, e.g.*, *State v. Ramirez*, 310 Or App 62, 67, 483 P3d 1232 (2021) ("[W]e do not usurp the role of the fact-finder and determine if defendant is guilty or reweigh the evidence." (Internal quotation marks omitted.)). We cannot say that there is little likelihood that admission of the AFTE identification evidence affected the verdict.

II. MOTION TO CONTROVERT

In defendant's second assignment of error, he contends that the trial court erred by denying his motion to controvert search-warrant affidavits and suppress evidence obtained pursuant to those warrants. In support of his contention, defendant relies on Article I, section 9, of the Oregon Constitution, ORS 133.693, and the Fourth Amendment to the United States Constitution.

Defendant filed a pretrial motion to controvert certain facts in the affidavits in support of the state's warrants, and to suppress evidence discovered as a result of the warrants. He argued that the affidavits left out material evidence concerning an alternative suspect—specifically, an eyewitness description of the shooter's clothing that differed from what defendant was wearing in the surveillance video—and that, when reevaluated considering that material omission, the affidavits did not establish probable cause. The trial court denied that motion for two reasons: First, the court found that the omission of the eyewitness description was not a material omission. Second, the court held that, even if the omission were material, the magistrate granting the warrant would have reasonably concluded that "the facts as alleged with all reasonable inferences establish probable cause."

On appeal, defendant asserts that the court erred in both its determination that the omission was not material and its conclusion that, even if the omission was material, probable cause nevertheless existed. Assuming without deciding that defendant is correct on the first point, we disagree with him on the second point. The omission of the

eyewitness account of defendant's clothing did not undermine the inference that defendant was the shooter depicted on the surveillance video because, as the trial court found, the description was not significantly inconsistent with defendant's clothing as depicted on the video and, consequently, did not undermine the affiant's belief that the video indicated that defendant was the shooter.[22] Accordingly, we affirm the denial of the motion to controvert.

III. CONCLUSION

Because the court erred in admitting the AFTE identification evidence and that error was not harmless, we reverse and remand.

Reversed and remanded.

[22] To the extent that defendant takes issue with the wording of the trial court's conclusion that "the facts as alleged with all reasonable inferences establish probable cause," we understand the court to have described the correct standard in imprecise terms rather than applying an incorrect legal standard.