No *Shepard's* Signal™
As of: March 17, 2020 4:10 PM Z

United States

United States District Court for the District of Oregon

March 16, 2020, Filed

Case No. 3:19-cr-00009-MO-1

Reporter
2020 U.S. Dist. LEXIS 45125 *

UNITED STATES OF AMERICA, V. ODELL TONY ADAMS, Defendant. MOSMAN,J., Defendant, Odell Tony Adams,

Notice: Decision text below is the first available text from the court; it has not been editorially reviewed by LexisNexis. Publisher's editorial review, including Headnotes, Case Summary, Shepard's analysis or any amendments will be added in accordance with LexisNexis editorial guidelines.

Core Terms

testing, error rate, toolmark, methodology, match, firearms, percent, training, false positive, convictions, scientific, forensic, studies, cases, scientific community, identification, bullet, peer, quality control, shell casing, inconclusive, threshold, examinee, fired, terms

Opinion

[*1] OPINION AND ORDER

"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 ofIdentification,* 43 AFTE Journal at 287-but still a hunch.

Moreover, the application of this circular standard **[*2]** is "subjective in nature ... based on the examiner's training and experience." *AFTE Revised Theory ofIdentification,* 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. This was evident throughout Mr. Gover's testimony as he struggled to explain what he was looking for in order to conclude the shell casings matched the Taurus. For example:

-OPINION AND ORDER

[PROSECUTOR]: **[*3]** So if there's not a numeric threshold when you're doing this identification, what assurances do you have that your conclusions will be consistent with the conclusions of other toolmark examiners?

[MR. GOVER]: Part of it is I'm using the same methodology, the same criteria for identification, and once I've completed doing my actual comparative analysis and I have drawn a conclusion, within our system, a second qualified firearms examiner will come in right behind me and look at that same evidence to either agree or disagree with my findings.

[PROSECUTOR]: And is it fair to say that AFTE has described the sufficient agreement as being a subjective standard

[MR. GOVER]: Yes. The interpretation of the objective observations, which is viewing those contours, surface contours of the toolmarks, being able to line up the peaks and valleys and ridges and furrows, that's - those are all objective observations. The interpretation of the observation is what is the subjective aspect of the firearms identification.·

[PROSECUTOR]: And when you're making your conclusion, is that based in part on your experience?

[MR. GOVER]: Based on experience, training, research provided by the Association of Firearm **[*4]** and Toolmark Examiners, current training, current validations that are going on now, as well as past experiments which date back decades, you know, being done by AFTE. So it's a combination of all of it.

Tr. [70] at 25-26.

[PROSECUTOR]: And was it your conclusion that there was sufficient agreement between those two?

[MR. GOVER]: Yes.

[PROSECUTOR]: And just for our benefit, can you sort of show your math, show us how you got there, in terms of just when you're looking at those two photos, what is it in particular that you're looking for that leads you to conclude, hey, there's sufficient agreement here?

[MR. GOVER]: I'm looking at the correspondence of those striated marks betweenmy test fire on the left and my unknown cartridge case on the right and evaluating that correspondence. And from what I can see from that correspondence, that exceeded the known level of individual characteristics that I would expect to see or correspond between two cartridge cases fired in two different firearms.

- OPINION AND ORDER

[PROSECUTOR]: I just want to go back over this one more time. I think we talked about this before. But why not, when doing your analysis to see if there's sufficient agreement, why not **[*5]** have a numeric threshold that if there's, say, seven striations that match, that's enough, but below that isn't good enough?

[MR. GOVER]: Why not?

[PROSECUTOR]: Yeah. Why isn't there a numeric - why can't we qualify this with numbers?

[MR. GOVER]: Part of it for me, to understand that quantifiable is almost another subjective aspect of when I'm looking at these furrows and ridges, what constitutes the consecutively - what they refer to as the consecutively manufactured or consecutively matching striae. So I don't have a complete answer to that and I've tried to formulate an argument as to why or why not, and I haven't, you know, heard a lot of other people talk about it."

Tr. [70] at 38-40.

[DEFENSE ATTORNEY]: .... And I'm sorry, the name of the person that did the review after you and confirmed?

[MR. GOVER]: Allesio.

[DEFENSE ATTORNEY]: Allesio. Now, the threshold that- so is it fair to say that Mr. Allesio has his own threshold in terms of what is sufficient agreement?

[MR. GOVER]: I don't-if he does, I would assume that it's pretty much the same, being as we have the same level of experience and type of training.

[DEFENSE ATTORNEY]: But you have no way of knowing that?

[MR. GOVER]: No.

[DEFENSE **[*6]** ATTORNEY]: Because it's based on his own perceptions and what he happened to retain?

[MR. GOVER]: It depends on his interpretation of the objective-

[DEFENSE ATTORNEY]: Maybe he read some literature that you didn't read, that kind of thing?

[MR. GOVER]: We both receive the AFTE Journal, so it's available for both ofus.

Tr. [70] at 46-47.

- OPINION AND ORDER

This last excerpt of testimony is particularly damaging to the admissibility of this

methodology under *Daubert.* 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 Mr. Allesio 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, such testimony 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 **[*7]** 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. 6 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 Malcolm

Gladwell's 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.

The full transcript of Mr. Gover's testimony is attached to this opinion as Appendix A.

- OPINION AND ORDER

B. Error Rates

The next *Daubert* factor that must be applied is the rate of error for the methodology **[*8]** in

question. It is important to note that whether a particular error rate is high or low has no

independent meaning outside of its context. Whether an error rate is "high" or "low" depends on

the use to which the testimony will be put. Here, the use of forensic toolmark testing is very

often the most critical factor in determining guilt or innocence-an endeavor that must tolerate

only a very small rate of error. Even an error rate of five percent, which would be low in many

contexts, would be unacceptably high where it drives a guilty verdict because it would mean that

one in twenty convictions could be wrong. 7 The question here is whether the error rate for

toolmark testing is acceptably low so as to be a reliable means of determining guilt or innocence.

The Government initially asserted that the error rate for toolmark comparison testing is

between .9 and 1.5 percent. Gov't. Resp. [57] at 9. But testing shows a range of outcomes, sometimes with an error rate as high as 2.2 percent. *United States v. Shipp,* No. 19-cr-029-NGG,

WL 6329658 *12 (E.D.N.Y., Nov. 26, 2019). If these all sound like low rates of error,

whose differences could not possibly be material, it is helpful to consider them in **[*9]** terms of

wrongful convictions, which is the correct framework for an error rate that measures only false-

positives-i. *e.* incorrectly identified matches. *See, e.g.,* President's Council of Advisors on Sci.

& Tech., *Forensic Science in Criminal Courts: Ensuring Scientific Validity ofFeature*

I recognize that the "one in twenty" figure would be an overstatement if the error rate

included both false positives and false negatives-i. *e.* resulting in false acquittals as well as false convictions. But as I explain below, I think the testing methods for the error rate of the AFTE method yield results almost exclusively focused on false positives. See, e.g., *PCAST Report,* 104-114 (discussing firearms testing). The PCAST Report was a "meta-study," which means it examined the results of all the fields tests that had been done and drew conclusions from all of those studies in context of each other. All of the studies cited in the PCAST Report focused on false positives when reaching an error rate for ballistics comparison testing.

- OPINION AND ORDER

Comparison Methods, 104-14 (2016) ("PCAST Report") (discussing firearms testing). A .9 percent error rate would lead to about 1 in 111 wrongful convictions. **[*10]** A 1.5 percent error rate

would mean that 1 in 67 convictions were wrong. And 2.2 percent would mean that 1 in 46 convictions were wrong. These are dramatically different rates of error when put into context.

What's more, the higher error rates tend to arise from the studies that most closely resemble the real-world conditions of toolmark testing. The lowest rates arise from the "closed-set" tests, which require the examinee to perform a matching exercise between two sets of bullets or shell casings. *Shipp,* 2019 WL 6329658 at *12 (citing *PCAST Report,* 106-11 (2016)). An examinee can "perform perfectly" ifhe simply matches each bullet to the standard that is closest.

Id. Further, each match narrows the field for further matches. *Id.*

The next highest error rates-about 2.1 percent-arise from partly closed sets. *Id.* (citing PCAST Report at 109). These tests also give the examinee a closed set of matches, but it also

includes two bullets or shells that do not have a match in the set. *Id.* The error rate from these

tests is "nearly 100-fold higher" than from the closed-set tests. *Id.*

Finally, the "black box" studies yield the highest error rates, about 2.2. percent. *Id.* (citing PCAST Report at 110-11). These tests presented **[*11]** each examinee with an unknown shell casing or bullet and three test fires from the same known firearm, which may or may not have been the source of the unknown casing or bullet. *Id.* These tests most closely resemble real-world

analysis-i. *e.* what Mr. Gover testified that he did in this case.

On the other hand, Mr. Gover testified that a study from the Ames Laboratory at Iowa State University found a 1 percent error rate from a test in which 218 examinees were given cartridges fired from 25 different firearms and used the methodology at issue in this case to

match the casings to the firearms. Tr. [70] at 27. The details of this study are not clear from Mr.

- OPINION AND ORDER

Gover's testimony, but this appears to have been a "closed-set" test, the type with the lowest

error rates on average.

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. 8 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 **[*12]** real world results. This may mean the error rate is lower from testing than in real world examinations.

It is hard to know exactly what to make of these results. It is possible that the error rate for toolmark testing is very low, but it is more likely that it is not. Assuming false positive test results lead to wrongful convictions, a wrongful conviction rate of 1 in 46 is far too high. The best test results would favor the government, but it is unlikely those tests reflect real-world error rates. The worst results favor Defendant. At most, then, this factor of the *Daubert* test is neutral as to both parties. In my opinion, it cuts somewhat in favor of Defendant.

C. Peer Review

At the outset, it is important to remember the purpose of this *Daubert* factor. 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 U.S. at 593. If a methodology has been published but not for this purpose, this factor is not satisfied in favor of admissibility.

In fact, the closed-set study discussed above yielded an "inconclusive" rate of 41.8 percent, and the black **[*13]** box study yielded an inconclusive rate of 33.7 percent. *PCAST Report* at 111. These results were not included in the "error rate."

- OPINION AND ORDER

Here, Mr. Gover testified that the methodology he uses is peer reviewed through the AFTE Journal, which exercises quality control over studies that are done and decides whether they are "worth publication." Tr. [70] at 31. The Journal does not evaluate the methodology itself, which has been established as the industry standard since 1992. *Id.* at 32. The only question the Journal asks is whether any studies being published used the correct, accepted methodology. *Id.* This does not amount to peer review, for two reasons. First, the AFTE Journal is a trade publication, meant only for industry insiders, not the scientific community. Second and more importantly, 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. In fact, the methodology has never changed, aside from a minor revision of terminology, in the 18 years Mr. Gover has worked as a forensic scientist. *Id.* at 32-33. This is not the purpose that *Daubert* sets out for peer review. This factor therefore favors Defendant. **[*14]**

D. Standards and Quality Control

Mr. Gover did identify some quality control mechanisms. For example, he takes an annual proficiency test. Tr. [70] at 33. The test yields only binary pass/fail results, not a rate of error, but it does give Mr. Gover some information about how well he can identify firearms matches. *Id.* Further, every forensic toolmark test is reviewed by a second examiner who either verifies or disagrees with the conclusion. *Id.* Then the results are put through a "technical review" process, in which another examiner reviews the notes taken during the comparison testing phase in order to make sure that the proper procedures were followed and that the examiners' conclusions are supported. *Id.* at 34. At a more general level, Mr. Gover and his colleagues also receive training and procedures manuals from AFTE that explain how to do comparison testing and what processes to follow. *Id.*

- OPINION AND ORDER

This amounts to quality control. If these enforcement mechanisms were applied to a scientifically testable methodology, it would be easy to say that toolmark comparison testing was held to a high standard and subject to quality control. This *Daubert* factor therefore favors the Government, **[*15]** although it is not dispositive.

E. General Acceptance

The fifth and final *Daubert* factor asks whether the methodology in question has been accepted in the broader scientific community. This is a difficult question to answer. The AFTE method that Mr. Gover uses has been widely accepted within his own community of technical experts. Tr. [70] at 35. 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. Def.'s Br. [55] at 6-7 (citing Nat'l Research Council, *Ballistic Imaging* 81 (National Academies Press 2008)); NRC Comm. on Identifying the Needs of the Forensic Sci. Cmty., *StrengtheningForensic Science in the United States: A Path Forward* 155 (2009)). In fact, these reports 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. On the other hand, widespread general acceptance within a given technical community could theoretically be sufficient. But that is more consistent with a *Kumho Tire* theory of expert testimony, not with *Daubert.* Here, where the scientific **[*16]** 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.

CONCLUSION

I want to be clear that my ruling, as expressed in the foregoing opinion, is limited by the testimony before me during the hearings held in this case. It is not an indictment of forensic

- OPINION AND ORDER

evidence or toolmark comparison analysis writ large. It is clear that Mr. Gover and his

colleagues are on to something. 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. 9

Therefore, for the reasons discussed above, Mr. Adams's Motion *in Limine* [55] is

GRANTED in part and DENIED in part. Mr. Gover's expert testimony is limited to the

following observational evidence: (1) the Taurus pistol recovered in the crawlspace of **[*17]** Mr.

Adams's home is a 40 caliber, semi-automatic pistol with a hemispheric-tipped firing pin, barrel

with six lands/grooves and right twist; (2) that the casings test fired from the Taurus showed 40

caliber, hemispheric firing pin impression; (3) the casings seized from outside the shooting scene

were 40 caliber, with hemispheric firing pin impressions; and (4) the bullet recovered from gold

Oldsmobile at the scene of the shooting were 40/l0mm caliber, with six lands/groves and a right

twist.

//

//

After listening at length to Mr. Gover's testimony, it appears there is an element of over-

inclusiveness at play. It seems likely that the AFTE method could be revised to rest on

quantitative factors. It seems equally likely that a more quantitative measure of sufficient agreement would result in a finding of inconclusive in cases that currently result in a match. Mr. Gover and his peers seem reluctant to impose quantitative restrictions on their methodology because it would fail to justify a match in those cases where the numerical standard isn't met, but the trained examiner has an impression--call it a hunch-that it is actually a match. But there are several settings in which law enforcement **[*18]** officials are required to leave their hunches at the courthouse door. It is only slightly inaccurate to say that the criminal justice system is designed to favor false negatives over false positives. To be admissible, in such a system, as scientific evidence, AFTE will have to shift away from hunches to numbers.

- OPINION AND ORDER

No evidence relating to Mr. Gover's methodology or conclusions relating to whether the

shell casings matched the Taurus will be admitted at trial.

IT IS SO ORDERED.

-1;

DATED this (*J* day of March, 2020.

- OPINION AND ORDER

End of Document