

SUPREME COURT OF THE STATE OF NEW YORK
COUNTY OF BRONX Part 33

-----X

THE PEOPLE OF THE STATE OF NEW YORK DECISION

Frye

Motion

Against

IND. NO.

267/2018

A. M. & MICHAEL ROSS,

Defendants.

-----X

April A. Newbauer, J. _

Defendants filed motions pursuant to *Frye v. United States*, 293 F. 1013 (D.C.Cir. 1923) to preclude a ballistics expert from testifying that shell casings found at a crime scene matched a firearm found in a car the defendants occupied. Defendants argued a forensic comparison of the shell casings to test fires from the gun would lack "scientific validity and general acceptance in the relevant scientific community." *Id.* The People opposed the motions, arguing that toolmark examination has been a widely accepted practice in forensic science and is thus not a proper subject of a *Frye* hearing. The court ordered a hearing on the questions of what scientific community is the relevant one when considering expert testimony on toolmark examination and whether toolmark evidence has general acceptance within that community.¹

Issues presented

Is the human eye capable of determining whether marks on casings or bullets are the result of being fired from a particular gun? Does the forensic training given to toolmark examiners make them reliable expert witnesses even though their opinions are inherently subjective? The answers apparently depend on which scientific disciplines comprise the "relevant scientific community." Ballistics analysts perform toolmark comparisons by placing a

cartridge casing or bullet from a crime scene and a cartridge or bullet test fired from a gun under a comparison microscope within the same field of view. Forensic scientists maintain that by examining toolmarks under a microscope, a trained examiner can determine whether there is "sufficient" agreement to reliably identify a particular firearm as the source of the toolmarks. Researchers in traditional scientific disciplines-- including study design and research methodology, statistics and psychology--are unified in their view that toolmark identification is just a practice in search of a science and is not reliable. For the reasons stated below, I find that both groups in this contentious debate together comprise the "relevant scientific community" that may offer valuable guidance on toolmark examination. Some expert toolmark testimony will be permitted to the extent that there is enough general agreement, if not universal acceptance² of its reliability.

Types of toolmarks

Unquestionably, firearms leave marks on shell casings and bullets when weapons are fired. So-called "class characteristics" are marks manufacturers intend to imprint on the bullets or casings in order to brand their products. These distinguishing marks can be made known to forensic examiners in the field through training offered by the firearms manufacturers. Examples of class characteristics are interior right or left twists, lands and grooves of a known size or parallel lines on the breech face of the firearm. Forensic toolmark examiners complete an extensive training course under supervision of their professional association (AFTE) to become skilled in recognizing class characteristics.

"Subclass characteristics" are also discernible surface features of the bullet or cartridge resulting from manufacturing, but they are unintentional. Subclass characteristics are caused when a machining operation unexpectedly leaves marks that carry across a batch or batches of firearms. These incidental features of the manufacturing process can be significant and may gradually evolve over time.³ Manufacturers do not routinely report subclass events and

or maintain a database of subclass patterns. There is simply no telling when a subclass arises or how many firearms are in a subclass. (Tr. 661)⁴ As noted by the People's expert, Detective Jonathan Fox, the NYPD first identified subclass issues by observing crime scene evidence where the source of markings could not be located. Some time later similar markings were observed-- apparently by coincidence--on different caliber cartridge heads of unfired cartridges.⁵ Most likely by tracing back through the manufacturer, investigators were able to categorize the markings as part of a subclass rather than mistakenly characterizing them as unique.

Thus it follows that subclass characteristics would ordinarily be unknown to a toolmark examiner. This big reveal lies at the heart of the critique of toolmark identification even though it was not the only focus of the *Frye* challenge. When tool examiners state--as the People's expert witness did--that they are looking for the "unique characteristics" that are left behind during the manufacturing process of the firearm, they really mean unique *as far as they know*. Examiners in the field would be utterly unaware of subclass characteristics present in a batch or batches of firearms. Subclass imperfections may appear as a thick heavy line or other type of mark. As such they may mimic the third class of toolmarks, "individual characteristics." These are stray marks on shell casings or bullets which are caused by any number of phenomena, including manufacturing errors affecting only a single firearm, but more commonly environmental factors such as damage to the casing or bullet by contact with other objects (vehicles, etc.). Fouling of the gun interior through repeated firing or slight damage as a result of cleaning can also result in stray markings.

Expert testimony

The People first called Jennifer Lady, a quality assurance manager in the NYPD laboratory. She was qualified as an expert in the accreditation procedures and quality assurance procedures at the police laboratory. Lady has been responsible for maintaining standards and assuring continued accreditation of the police lab with the New York Commission on Forensic

Science, the American Society of Crime Lab Directors (ASCLD), and more recently the National Accreditation Board (ANAB). Lady testified that to maintain accreditation, the NYPD laboratory is subject to ongoing monitoring by ANAB, which sends volunteers from other forensic laboratories to conduct a full assessment "of every single accreditation requirement," including supplemental requirements that NYPD adds voluntarily. NYPD has a current accreditation certificate in effect since 2018. There are also no open "nonconformances", or deviations from accepted standards.

Lady was in charge of overseeing proficiency testing for NYPD firearms examiners. However, she acknowledged that she was not qualified to perform a technical assessment of the microscopy section of the NYPD lab because she does not have the requisite base knowledge to make an assessment. Lady testified that ballistics examiners take proficiency tests created by external forensic laboratories as well as in-house tests. She indicated the tests are designed to mimic case work but did not elaborate on the design or implementation of the tests.

The People also called Detective Jonathan Fox, a twenty-two year veteran of the NYPD and an operability tester and microscopist⁶ for the ballistics section of the police laboratory since completing training in 2007. He was qualified as an expert in general ballistics operability, and in microscopic firearm and toolmark analysis. Fox's understanding of toolmark identification was acquired through AFTE training and years of experience. According to AFTE theory, firearms identification is a discipline of forensic science with its primary concern to determine if a bullet, cartridge case or other ammunition component was fired by a particular firearm. Quoting from the AFTE Technical Procedures Manual(P's Ex 6), Detective Fox stated that microscopy was: "[a]n empirical comparative analysis that can determine if a striated or impressed mark was produced by a particular tool. Empirical, he said, meant "based on the verifiable by observation or experience rather than theory or pure logic."

When asked to describe the range of conclusions a firearms examiner

might reach when conducting an analysis, Detective Fox testified that he could conclude cartridge casings were fired from the same weapon by either : 1)"sufficient agreement" of the individual and class characteristics of the firing pin; 2) "sufficient individual characteristics and class characteristics of the breech face"; or 3) sufficient agreement of both. Fox testified that a finding of sufficient agreement is subjective and there is no across-the-board standard as to what is "sufficient agreement" in his field.

The People's final expert was Todd Weller, a forensic science consultant and former criminalist for the Oakland Police Crime Laboratory. Weller, a member of AFTE and a contributor to forensic science oversight panels, was qualified as an expert in firearm and toolmark examination and study design as it is applied in forensic science. Weller has testified in multiple courts on toolmark analysis and is an impassioned advocate for AFTE toolmark identification theory.

In 1992, AFTE promulgated and adopted a range of conclusions for toolmark examiners as follows:

-Identification, described as "agreement of all discernible class characteristics and *sufficient agreement* ⁷ 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";

-Inconclusive Type A, described as agreement of all discernible class characteristics and some agreement of individual characteristics, but insufficient for identification;

-Inconclusive Type B, described as right in the middle; and

-Inconclusive Type C, described as agreement of all discernible class characteristics and some agreement of individual characteristics, but sufficient for an elimination;

-Elimination, described as significant disagreement of discernible class characteristics and/or individual characteristics; and finally

-Unsuitable for examination, usually described as item so damaged it doesn't have any firearm marks remaining.

Weller claimed that the process of firearm identification through toolmarks is not purely subjective because "a lot of observations lead up to the final conclusion." Weller stated the objective elements are measuring the bullet's diameter, counting the number of lands and grooves, whether they twist right or left, and whether under a microscope the analyst could detect the presence of striae that line up. When asked what "sufficient agreement" is, Weller referred back to the AFTE definition. (Tr. 615) Identification under AFTE theory has also been described as the "significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours." *See United States v. Taylor*, 663 F Supp2d 1170, 1177 (D. NM 2009). As a number of courts have pointed out that the AFTE standard is circular—an identification can be made upon sufficient agreement, and agreement is sufficient when an identification can be made. *Id.* at 1177, *citing United States v. Monteiro*, 407 F Supp2d 351 (D. Mass 2006).

Weller could not specify the frequency of manufacturing glitches causing subclass characteristics, or the number of firearms affected by subclass characteristics. Still he maintained, ..."there's a lot of literature and training involved in identifying subclass characteristics to make sure you don't use them for a conclusion or identification."⁸ When asked how an examiner distinguishes between subclass and individual characteristics, Weller offered an elliptical response:

So class is a measurable feature, as I talked about before, or observable feature, hemispherical firing pin, caliber, numbers of lands and grooves. Subclass marks have their own characteristics. They tend to be uniform marks. So, for example, on a breech face, the marks will carry across the entire breech face surface with no break of those features. So when examining a tool, that's what you are going to look for, is relatively coarse, large marks that are going to carry across the entire tool surface. There is another area of firearms that has a propensity for subclass characteristics. Specifically, groove impressions sometimes can have

subclass characteristics. So examiners are trained not to rely on that unless they have the barrel to examine and they can rule out subclass characteristics.⁹

While not qualified as an expert in statistics, Weller was permitted to discuss the various studies the People introduced in terms of their acceptance by forensic scientists. Most of the toolmark industry studies report a high degree of accuracy by trained forensic examiners. Accredited laboratories across the United States rely on the methods these studies purport to validate. However Weller was not able to opine on the reliability of the studies' methodology, foundational validity or conclusions from a research perspective.

Defense witnesses Dean David Faigman of Hastings Law School and Dr. Nicholas Scurich of the University of California, Irvine, were qualified as experts in the related fields of scientific research methodology and study design. Both witnesses have interdisciplinary appointments at their respective universities. Professor Faigman offered a detailed critique of the conception and design of the studies undertaken by forensic scientists to support the validity of toolmark examination.¹⁰ First, Faigman underscored the difference between DNA analysis, which arose out of mainstream science and was rigorously tested by peer-reviewed academicians, and the so-called identification sciences including toolmark analysis, which were developed in police laboratories and became part of their field practice. The forensic practices had never been subjected to peer review and tested in studies using traditional research protocols. Faigman outlined various deficiencies in the toolmark industry study designs: failure to select an appropriate group of forensic examiners to participate; failure to establish a control group; inability to control and describe the level of difficulty of the tests; failure to mandate test conditions and provide oversight; and inadequate peer review of the results. While he acknowledged that all studies have some limitations, Faigman views the toolmark studies as uniquely combining obscurity of design with lax implementation and oversight. According to Faigman, only

the independent Ames Laboratory study (Def. Ex.K) was structured in a way to yield valid results. Faigman emphasized what has been noted by some courts: that while it is sometimes appropriate for scientists to evaluate their own methods through a peer review process, establishing scientific validity also requires scientific evaluation by non-affiliated researchers. *See People v. Thompson*, 65 Misc3d 1206(A)(SupCt Kings Co 2019)(Dwyer, J.).

The Ames study adhered to what is commonly referred to as "black box" design, in which the researchers tested and arranged all the ammunition in advance so the correct answers (either a match or no match) were known to them. Volunteer examiners were asked to determine whether a given pair of bullets or cartridges matched, and then to move on to the next set ("pairwise comparison"). Both defense experts found this design more likely to generate accurate data than the numerous studies in which examiners were given a number of known cartridges or bullets and asked whether any of them matched a number of unknowns ("set-to-set" design). The set-to-set design allowed examiners (who were unsupervised) unlimited time and ability to make cross-comparisons of the knowns and unknowns to reach their conclusions. It was not known which comparisons were made. In the pairwise design in contrast, the examiner's conclusion on one pair is independent of any conclusion as to the next pair comparison.

The results of the Ames pairwise study suggested that participants made positive identifications fairly successfully within the confines of the experiment.¹¹ However, approximately thirty percent of the original volunteers did not return a completed study, and it is not known why. More importantly according to Faigman, even among those examiners who completed the study, the number of "inconclusive" answers rose significantly ("shot through the roof" as he put it) compared to other studies. This may suggest that the test designed by an independent lab was more challenging than the industry designed studies. Further, Faigman testified that mainstream scientists would consider the high rate of inconclusive findings--twenty percent of the total--to be error, since the examiners failed to identify

what were designed to be known outcomes (match or no match).¹² Faigman admitted that not all scientists would agree on what kind of error the inconclusive results represented (*see* "error rates" below) and therefore how to view a twenty percent error rate. The Ames Laboratory study has not been published or replicated. According to Faigman, its real import is to underscore the need for many more carefully designed studies testing the accuracy of toolmark identification. Faigman has been involved in an effort to bridge "mainstream" science and forensic science and secure funding for proper methodological study of toolmark examination. As he was quick to point out, the importance of the categorization of the inconclusives as errors is more important when assessing their implications for the accused defendant than in a research laboratory: "[t]he real question...is what are the consequences of making a mistake."

The defense's other expert, Dr. Nicholas Scurich, has an academic background in psychology, and his current focus is on research and design methods and how to evaluate the outcomes of studies involving some level of human decisionmaking. Dr Scurich was qualified as an expert in research methods including psychometrics,¹³ and the study of human judgment and decisionmaking. Scurich agreed with Faigman that the toolmark tests designed by forensic scientists were highly problematic because of the test design and unclear criteria for participation ("sampling problems"). Scurich testified that it takes care to design and evaluate a study involving human judgment because the sample of participants may affect the outcomes, participants may interpret instructions differently or may simply stray from the instructions. Scurich also rejected the set-to-set design studies as not providing the ability to measure results, because of the myriad of unknown comparisons that participants may have made to arrive at their answers. He testified that this key feature of the forensic set-to-set designed studies also rendered them unlike most actual ballistics examinations in the field.

Error Rates

The question of the proper way to view the "error rates" in the studies

became hotly contested and emerged as a major area where forensic and mainstream science parted company. The defense argued that examiner error rates of up to 35% in the studies suggest a fundamental problem with subjective toolmark analysis:

Q.: [Dean Faigman], why is it important to establishing error rates when determining the foundational validity of a discipline?

A: Well, the error rates really define the weight that you want to give...All applied science is probabilistic...We never say a hundred percent certain[ty]. Even DNA gives you a random match probability which is a probabilistic statement about the DNA profile. And so the error rate tells you what are the costs potentially, will give you what are the costs of making a mistake....Now, the problem with firearms is we don't know have (sic) those data in terms of what the protocols or what the methods are actually doing. So the best we can possibly have is really the black box studies which can tell us what the error rate is under conditions that are as close to your case work as possible so that we know what their cost of making a mistake might be if we rely on a firearms expert that says it's a match, it's not a match or it's inconclusive. So error rates are key. They are, I would say, the cornerstone of foundational validity. They're the cornerstone of the foundation.

Todd Weller countered that the sheer number of studies with similar results verify the success of trained examiners in matching bullets or casings to firearms. Weller acknowledged that as in all forensic disciplines, toolmark examination has some "subjectivity of interpretation." He pointed out that DNA analysis—the gold standard forensic discipline—was also somewhat subjective and open to interpretation, particularly where there are multiple contributors or low copy DNA. To address the seemingly high error rates in the studies, Weller defended the studies' treatment of inconclusive answers as correct answers, due to the examiners' cautious approach and experience operating in real world conditions; he praised examiners for not jumping to conclusions. Weller disagreed with Faigman and Scurich's conclusion that

because the test designers predetermined the answers to specific questions would be either right or wrong answers, any inconclusive result must be considered an error. Scurich additionally pointed out that if inconclusive answers were not treated as errors, an examiner could safely declare all comparisons inconclusive and have made no errors. The Ames authors noted that not every participant answered every question. Participants were theoretically avoiding the more difficult questions. Scurich even classified inconclusive answers in this context as false positives because of their implication for real cases in which defendants should be exonerated but are not because of inconclusive results. The larger question is whether any forensic validation studies have sufficiently low error rates and generate high enough confidence intervals¹⁴ that mainstream scientific community would consider valid.

Because no witness was qualified as an expert in statistics, the court proposed, without objection, to call Dr. Heike Hofmann as a court witness. Dr. Hofmann holds a doctorate in statistics and is Professor of Statistics at Iowa State University, an institution at the forefront of research in toolmark examination. Dr. Hofmann is a co-author of the 2018 article "Automatic Matching of Bullet Land Impressions," introduced at the hearing as a prosecution exhibit (People's 36).¹⁵ The study examined the feasibility of 3D surface measurements of impressions on bullets and cartridges, thus automating the comparison to determine if they were fired from the same barrel or different barrels.

Dr. Hofmann attacked the stated error rates in the forensic toolmark studies. She testified that the determination of a true error rate was very tricky to determine from these semi-field studies because of the multiple sources for error. Dr. Hofman expressed particular concern over evidence quality and the lack of controls over the examiners taking the tests. These factors were not taken into account in calculating error rates in most studies, even the Ames Laboratory study. Hofmann also agreed with the defense experts that the studies which had a closed set, where a cartridge case must match or not

match one other piece of evidence are known to underestimate the error rate. Finally, she testified that by excluding inconclusive results, the error rates were artificially deflated.

***Frye* Analysis**

The test under *Frye* is whether the proffered scientific techniques, when properly performed, generate results accepted as reliable within the relevant scientific community. *Parker v. Mobil Oil Corp.*, 7 NY3d 434 (2006), citing *People v. Wesley*, 83 NY2d 417 (1994). A *Frye* hearing on toolmark examination has never been held in New York State before. As one federal district court judge recently commented, courts have in "cursory fashion" tended to identify toolmark examiners as the relevant community, summarily determine that AFTE theory is generally accepted, and permit firearm identifications through expert testimony. See *United States v. Shipp*, 422 F.Supp 3d 762 (EDNY 2019). See also *United States v. Green*, 405 F.Supp2d 104, 122 (Dist Ct. Mass 2005): "[a]lthough the scholarly literature is extraordinarily critical, court after court has continued to allow the admission of this testimony." In 2010, the court in *People v. Givens*, 30 Misc3d 475 (SupCt Bx Co 2010)(Webber, J.), declined to hold a *Frye* hearing on toolmark testimony, noting that toolmark identification was a generally accepted forensic practice. The court cited *United States v. Monteiro*, 407 FSupp2d 351, in which a district court had rejected the defendant's request to exclude toolmark examiner testimony under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993).

In opposing the motions for a *Frye* hearing, the People pointed out that toolmark and firearm identification has been accepted in courtrooms for years. Nothing about the practice is novel, so they argue it was not subject to a *Frye* challenge. See *People v. Brooks*, 31 NY3d 939 (2018).¹⁶ The Court of Appeals largely discounted this argument in a recent case, finding the trial court had abused its discretion in failing to conduct a *Frye* hearing on low copy DNA. Writing for the court, Judge Fahey said:

Familiarity does not always breed accuracy, and our *Frye* jurisprudence accounts for the fact that evolving views and opinions

in a scientific community may occasionally require the scrutiny of a *Frye* hearing with respect to a familiar technique. There is no absolute rule as to when a *Frye* hearing should or should not be granted, and courts should be guided by the current state of scientific knowledge and opinion in making such determinations.

People v. Williams, 35 NY3d 24 (2020). The high court had previously expressed this precept in terms of the court's need to examine the scientific foundation for admission of evidence as it would the foundation of any evidence. See *People v. Wesley*, 83 NY2d at 422; *Parker v. Mobil Oil Corp.*, 7 NY3d at 447. See also *Matter of Floyd Y.*, 22 NY3d 95 (2013). In their motion papers, the defendants persuasively argue that toolmark identification procedures qualify as novel within the meaning of *Frye* because that they have never been scientifically tested, citing the NRC and PCAST critiques. The motions were granted and the hearing was held before the trial, in January 2020.

The People adhere to the view that the relevant scientific community should be limited to forensic scientists and their established conclusions. Yet *Frye* demands an unbiased, objective review by those with no professional interest in its acceptance. *People v. Williams*, 35 NY3d 24. The professional standing and livelihood of forensic scientists depends on the validity of AFTE theory. See *United States v. Tibbs*, 2019 DC Super. LEXIS 9. Certainly this came across in the testimony of Mr. Weller, a professional consultant and frequent expert witness for the prosecution. The targeted use of AFTE theory by law enforcement investigators, under pressure and with potential for confirmation bias, limits the degree of intellectual rigor and detachment that counts as neutral scientific expertise. See *United States v. Shipp*, 422 F.Supp.3d 762.

In the quest to determine the relevant scientific community for *Frye* purposes, rarely do the experts fall into such cognizable camps, forensic practitioners on one side and academic researchers on the other. Until the publication of the National Research Council (NRC) report in 2009, forensic science was clearly the only scientific community that counted in criminal investigations and courtrooms. In the next decade, however, toolmark

analysis and other forensic practices came under a national lens. The NRC report contrasted the weighty research underpinning DNA analysis with the void in scientific studies validating toolmark identification. Both the NRC and PCAST reports thus jump-started court review of traditional forensic practices. At this point, the results are mixed. Some courts still permit a testifying ballistics examiner to recite the ‘reasonable degree of ballistics certainty’ standard. *See United States v. Johnson*, 875 F3d 1265 (9th Cir. 2017)(reviewing the district court under an abuse of discretion standard). But other courts have found that testimony too misleading. *See United States v. Glynn*, 578 FSupp2d 567, 574-575 (SDNY 2008)(limiting ballistics examiner to state that a match was "more likely than not."); *see also United States v. Ashburn*, 88 FSupp3d 239 (EDNY 2015). The NYPD laboratory itself has now turned away from the ‘reasonable degree of scientific certainty’ standard in drawing its conclusions about ballistics and to the "sufficient agreement" language consistent with AFTE guidelines. Most courts conclude that although the studies have flaws and the error rates are simply too hard to calculate, because of the rigor of examiner training and the assumption that error rates are low, some expert toolmark testimony should be permitted. *See United States v. Johnson*, 2019 WL 1130258. But a number of district courts hearing *Daubert* ¹⁷challenges have broadened the relevant scientific community to take into account the contrasting views of mainstream researchers. Consequently, the scope of permissible expert toolmark testimony is narrowing overall. .

In a *Frye* analysis the relevant scientific community does not have to be unanimous. *See Matter of State of New York v. Hilton C.*, 158 AD3d 707 (2d Dept 2018). The Court of Appeals has expressed as much: "[a]s with any other type of expert evidence, we recognize the danger in allowing unreliable or speculative information (or "junk science") to go before the jury with the weight of an impressively credentialed expert behind it. But, it is similarly inappropriate to set an insurmountable standard that would effectively deprive [parties] of their day in court. It is necessary to find a balance between these

two extremes." *Parker v. Mobil Oil Corp.*, 7 NY3d 434, 447 (2006). A scientific standard that is subjective, "inherently vague" and interpretative warrants serious scrutiny. *See United States v. Glynn*, 578 FSupp2d 567. [Forensic scientists and AFTE themselves have not taken a head-in-the-sand attitude](#) lately but appeared to recognize that toolmark identification, while time honored in the field and the courtroom, is under attack and must withstand scrutiny by other researchers. The People submitted numerous exhibits demonstrating that forensic science has picked up the pace and virtually raced to publish favorable studies. Where forensic science has fallen short, however, is that these validation efforts have been afflicted by fatally flawed study designs and subpar quantitative and qualitative measurements. General acceptance under *Frye* requires the testimony of neutral scientists experts. *See People v. Williams*, 35 NY3d 24; *see also Burton v. State*, 884 So2d 1112 (Ct Appeals Fla 2004). Considering the voluminous hearing record, including the testimony of expert witnesses and the exhibits,¹⁸ the relevant scientific community has been shown to include forensic scientists, researchers in scientific methodology and statistics, and to the extent that human perception and judgment continue to be involved in toolmark examination, psychologists as well. These disciplines all contribute to determining what is accepted science in the firearm and toolmark identification arena.

This broad community has no consensus on the more subjective findings and conclusions of toolmark examiners. The vague "sufficient agreement" standard and the circular reasoning needed to arrive at a firearm identification has no acceptance in mainstream science. Moreover, the language employed by the prosecutions's experts has the potential to mislead jurors. A jury may well be inclined to conclude that any "matching" marks on shell casings are unique to the firearm when that simply cannot be within the examiner's knowledge. (*See* discussion of subclass characteristics, *supra*). "Sufficient agreement" is too easily equated with complete agreement. Such an opaque or fuzzy concept may cause a jury to speculate rather than

understand nuances inherent in toolmark identification.

The defendants sought preclusion of ballistics examiners entirely, arguing that their professional bias and abject indifference to scientific method renders them akin to astrologists. But as with many types of applied science, practical experience must have some relevance and value. There should be some agreement in the relevant scientific community on the value of that experience. *See People v. Oddone*, 22 NY3d 369 (2013). And there is, even among mainstream scientists. Toolmark examiners are not the forensic equivalent of astrologists. No expert disputed the evidence that individual manufacturers intentionally brand their firearms with tools, and that these signature markings can be taught to examiners. Like the embossed indentations on an Oreo cookie, they are instantly recognizable to a trained examiner, and capable of assisting the fact finder to the extent of often, though not always, ruling in or out a particular firearm manufacturer.¹⁹ The People have met their burden to establish *Frye* admissibility as to class characteristics. *See Marso v. Novak*, 42 AD3d 377 (1st Dept 2007). It would be farcical to preclude experienced ballistics experts from rendering any opinion about known manufacturing marks. There is a consensus, or at least not all that much disagreement, to allow examiners to express an opinion on toolmarks that are class characteristics. The defendants' challenge to all expert testimony on toolmarks is therefore rejected.

Mainstream scientists really focus their critique elsewhere—on whether examiners should be expressing opinions on perceived individual characteristics. Even if an expert is using reliable principles to examine for class characteristics, there is little reliable basis for extrapolating further from other marks seen under a microscope. The expert's opinions must be limited if there is simply too great an analytical gap between the data and the opinion proffered. *See Cornell v. West 51st Realty, LLC*, 22 NY3d 762 (2014). At a foundational level, beyond comparing class characteristics forensic toolmark practice lacks adequate scientific underpinning and the confidence of the scientific community as whole.

A significant flaw in the forensic method is the potential for subclass characteristics to mimic individual characteristics and obscure the true reason for what may appear to the examiner to be a unique match: "...[b]ullets fired from different guns may have significantly similar markings, reflecting class or subclass, rather than individual characteristics." *United States v. Taylor*, 663 F.Supp2d 1170,1177 (Dist NM 2009). Both the literature and the forensic science expert confirmed that subclass characteristics remain an unknown for the examiner under ordinary circumstances. Such a void can lead to an erroneous conclusion that there is "agreement" or "consistency" if the examiner mistakes a subclass characteristic for an individual one on discharged shell casings or bullets..

Courts must be wary when scientific evidence is offered to prove a defendant's guilt. The problem of a microscopist putting expert conclusions before the jury is not solved by the opportunity to cross examine or present a countervailing expert on methodology. Experts enter upon the jury's province, since the expert--and not the jury--draws conclusions from the facts. *Matter of Floyd Y.*, 22 NY3d 95, 106, *citing People v. Cronin*, 60 NY2d 430, 433 (1983). The testimony of "sufficient agreement of a match" or even the language "consistent with" goes to the heart of the question of guilt. Doubts as to admissibility under *Frye* are best resolved to minimize the chances of a wrongful conviction. *Cf. Sybers v. State*, 841 So2d 532 (1st DisCt Fla 2003).

Fortunately, evidence submitted at the hearing suggests positive change is coming. Research into 3D technology promises a potential for automatic analysis of the surface topographies of bullets, such that the current problems of subjectivity and quantifying the marks and impressions more precisely. (People's Ex. 31, 36). As Dr. Hofmann observed, "[m]atching bullets is clearly not a one-step process, but rather a sequence of data analysis tasks each deserving attention. As there is no scientific standard in place at this point in time, our intent is to explain an approach to addressing these tasks, while documenting all steps and providing all code so that other

researchers and forensic scientists can reproduce and expand on our findings. **Science may well eliminate the interpretation, guesswork and biases associated with visual forensic examinations.** " (Emphasis added.)²⁰

Conclusions

The People may call an expert to testify as to whether there is evidence of class characteristics that would include or exclude the firearm at issue. The ballistics examiner may explain the reasons for an opinion that class characteristics are present or not present to the jury. In addition, the examiner may, if s/he believes the class characteristics are the same, indicate that the firearm cannot be ruled out as the source of the shell casings. The examiner may further explain what is done with instruments, e.g. the process of using a comparison microscope, describe verbally and/or show the jurors photos of the relevant evidence, including shell casings and test fires.

The examiner may not, however, offer qualitative opinions on matters not adequately supported by the defined relevant scientific community. Specifically, the examiner may not opine on the significance of any marks other than class characteristics, as the reliability of that practice in the relevant scientific community as a whole has not been established. Moreover, any opinion based in unproven science and expressed in subjective terms such as "sufficient agreement" or "consistent with" may mislead the jury and will not be permitted.

The defendants' motions to preclude testimony by a forensic toolmark examiner are granted in part and denied in part. The People may proffer their NYPD ballistics detective as an expert in firearm and toolmark examination for the testimony on class characteristics as described above.

This constitutes the decision and order of the court.

ENTER,

Dated: June 30, 2020
Bronx, New York

HON. APRIL A. NEWBAUER
Acting Supreme Court Justice

¹Federal and other state courts have heard similar challenges to toolmark testimony, and the parties discussed in considerable detail the history

of the arguments on both sides in their motion papers. In addition, the parties proffered expert affidavits from some of the prior cases. The defendants stressed commentary from the National Research Council Report, a 2009 analysis critical of forensic practices. They also cited the 2016 report of the President's Council of Advisors on Science and Technology (PCAST), which purported to speak for "mainstream science" in criticizing toolmark examination as lacking in precision and scientific validity. The People argued that the FBI, American Society of Crime Lab Directors (ASCLAD) and Association of Firearm and Toolmark Examiners (AFTE) vehemently rejected these criticisms, and Attorney General Loretta Lynch rejected the recommendations of the PCAST authors. (*See People's Affm in Opp*, pp.33-34.) While these sources were broadly taken into consideration, the rulings here primarily focus on the evidence adduced at the hearing.

²General acceptance does not mean absolute acceptance in the relevant scientific community. There will always be sticklers and doubters. Unanimous endorsement is not necessary. *People v. Middleton*, 54 NY2d 42 (1981). What is important is to allow a properly trained expert to assist the jury in drawing conclusions without misleading the fact finder by either grandiose conclusions or deceptively vague testimony.

³AFTE Glossary definition. (*See People's Ex. 4.*)

⁴Defendant's motion quotes manufacturing studies showing that thousands of firearms may have identical or similar imperfections and produce similar markings on the bullets fired through them. *See Rowe, W.F.*, "The Use of Statistics in Forensic Science" (CRC 1991), Affirmation of Kyla J. Wells, Esq.

⁵*See*, "Subclass Characteristics on CCI Speer Cartridge Case Heads," *People's Ex 4.*

⁶Detective Fox explained the comparison microscope is essentially two microscopes combined through an optical bridge, which allows the microscopist to examine two different pieces of evidence side by side.

⁷Emphasis added to note this phrase is the shorthand examiners will frequently use to signify identification.

⁸Fr: testimony of Todd Weller (Tr. 612).

⁹Multiple sources submitted by the People observe that there may very well be confusion between subclass and individual characteristics, leading to inappropriate identifications. Detective Fox was a co-author of one such article, but he was not asked about his analysis or conclusions.

¹⁰Professor Faigman and his company Jurilytics have been designated a technical advisor to courts and found to have provided "a high level academic peer review of expert reports." *See TL Wallace Construction v. McArthur, et al.*, 234 So3d 312 (SupCt Miss 2017).

¹¹Even in the Ames study (also called the "Baldwin study" after its author) there were unresolved design issues such as the unknown and uneven way the test may have been administered, and that multiple participants never returned the test and were not accounted for in the results.

¹²The other defense expert as well as mainstream researchers noted in the People's exhibits (Dr. Itiel Dror, Massachusetts Institute of Technology, Dr. Glenn Langenburg, Def. Ex G, Dr. Michael Salyards Ex. I).

¹³Dr. Scurich defined psychometrics as the study of creating and evaluating measurements for the outcomes of studies so that the information will prove useful.

¹⁴A confidence interval is a range within which (based on prior statistical results) a parameter--such as the percentage of inconclusive answers--will fall within a definite range of values. A confidence level is the probability the parameter will fall within the range. Evidence at the hearing suggests that research scientists look to arrive at a confidence level of about 95% for meaningful results. High confidence intervals can only be arrived at by testing a large quantity of data. One problem in the forensic studies was that sampling sizes were too small to generate a statistically meaningful

confidence interval.

¹⁵Dr. Hofmann's most recent article takes as its starting point the court's decision in *United States v. Green*, 405 F.Supp2d 104 (D. Mass 2005), in which a federal district court denied a *Daubert* challenge to expert testimony as a whole but still limited the scope of the testimony because of concerns over the scientific validity of firearm identification.

¹⁶Absent a novel or experimental technique, a *Frye* hearing is generally unwarranted. *People v. Brooks*, 31 NY3d 939. However, a court may always examine whether there is too great an analytical gap between analytical data and the scientific opinion sought to be offered. *Id.*

¹⁷*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 US 579 (1993). Under *Daubert* and Rule 702 of the Federal Rules of Evidence, a federal court operates as a gatekeeper to ensure that expert testimony is based on sufficient facts and is the product of reliable principles and methods to assist the jury. Admissibility is guided by five factors: 1) whether a theory or technique can be (and has been) tested; 2) whether the theory or technique has been subjected to peer review and publication; 3) the theory's or technique's known or potential rate of error; 4) the existence and maintenance of standards controlling the technique's operation; and **5) whether the theory or technique is generally accepted within the relevant scientific community** (*Frye* standard). New York is one of seven remaining states that use only *Frye* as the standard for admitting expert testimony. 4 NY Practice, Commercial Litigation in NY State Courts 30.14 (4th Ed.).

¹⁸ Some exhibits were admitted for limited purposes.

¹⁹As Detective Fox testified, Glock weapons are manufactured by building the firearm around a mandrel which is then removed, making the interior surface smoother than when a broaching tool is used to drill out the center of the barrel. Detective Fox stated that the majority of the time when a Glock is involved he finds the data inconclusive. (Tr. 156)

²⁰*See* Hare, E., et al., "Automatic Matching of Bullet Land impressions," *The Annals of Applied Science*, Vol. 11, No. 4 (2017)(People's Ex. 36).