DECLARATION OF CLIFFORD SPIEGELMAN IN SUPPORT OF DEFENDANT JOSEPH BLACKNELL'S MOTION TO EXCLUDE FIREARMS AND TOOLMARK IDENTIFICATION EVIDENCE OR, IN THE ALTERNATIVE, FOR A *KELLY* HEARING

I, Clifford Spiegelman, depose and state as follows:

1. I received an undergraduate degree in 1970 in Economics, Math, and Statistics from the State University of SUNY in Buffalo. In 1973, I received a Master of Science degree in 1973 in Managerial Economics from Northwestern University. In 1976 I was awarded a PhD. degree in Applied Mathematics/Statistics from Northwestern University. After obtaining my PhD, I served as an Assistant Professor of Statistics at Florida State University. Since 1990, I have been a Professor in the Department of Statistics, and currently a Distinguished Professor at Texas A&M University.

I am a member of the American Statistical Association (ASA), the world's largest community of statisticians; collectively, ASA's members "serve in industry, government, and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare." About the American Statistical Association, available at http://www.amstat.org/about/index.cfm. I have served on the Executive Committee of the ASA's Section on Physical and Engineering Sciences. I have also served as President for the South East Texas Chapter of the ASA.

Further, I have been a Managing Board Member for the Institute for Studies in Science and Law since 2007, and have served on the Board of Trustees of the National Institute of Statistical Sciences since 2008. I have been selected for honors and awards including Fellow of the Institute of Mathematical Statistics (1990), Fellow of the American Statistical Association (1992), elected member of the International Statistical Institute (1993), Statistics in Chemistry Award for best paper (2002 and again in 2008) and the Jerome Sacks Award for Outstanding Cross-Disciplinary Research (2007).

In 2003, I was a member of a National Research Council (NRC) Committee that evaluated and issued a report on Forensic Comparative Bullet Lead Analysis.

I have published in journals on both theoretical and applied statistical sciences on theories of error, statistical significance, engineering standards, and calibration and many other topics.

2. Like other scientists, statisticians such as myself are experts in the scientific method and its application. As a statistician, I am also an expert in experimental design, i.e.

how a study must be constructed to answer a given question. "[S]tatistical techniques . . . help determine what questions should be asked and what answers can be obtained given the available data" for every forensic technique. Budowle et al, A Perspective on Errors, Bias, and Interpretation in the Forensic Sciences and Direction for Continuing Advancement, *J Forensic Sci* at 6 (2009).

To the extent that a firearms examination – or any forensic examination, for that matter – purports to extrapolate the results of a single side-by-side comparison to a statement that expressly states or implies knowledge of the entire population of firearms, it necessarily includes inferential statistics; that is, the process of obtaining information about a larger group (population) from the study of a smaller group (sample). Thus, whenever the firearms analysis culminates in an identification, then the relevant scientific community also includes statisticians. It is undeniable that a statistician does not need to have analyzed bullets to assess whether the data that firearms examiners collect in a case is of the type that can be inferred to the larger population of all firearms.

3. Firearms examiners believe that each firearm imparts unique marks onto bullets and cartridge cases. However, I have read scores of articles on firearms identification, and I am not aware of any literature that comes close to justifying a claim of uniqueness in the firearms comparison discipline. Even assuming that bullets and casings can exhibit unique marks, there is no scientifically valid evidence that firearm examiners can, under their current methodology, come to a reliable conclusion that a bullet or cartridge case was fired from a particular gun. As I explain below, this is my opinion upon reviewing the relevant literature, and it is the belief of the scientific community as encapsulated in the 2008 NRC report "Ballistic Imaging" and the 2009 NRC report "Strengthening Forensic Science in the United States: A Path Forward," which was in turn endorsed by the American Statistical Association (discussed in more detail below).

4. The scientific method, which is the only method of hypothesis testing that is generally accepted by scientists, imparts reliability on a scientific process by allowing for results that are repeatable and reproducible. For firearms/toolmarks examinations, repeatable results mean that the same examiner on different days would reach the same conclusion. Reproducible results mean that different examiners would reach the same conclusions. If a methodology does not have these features, to a high degree of certainty, the methodology cannot be relied on to

reach accurate results.

In order to ensure reliability, the scientific method must be applied to any purported claims of "identification" with respect to bullets and casings. Without repeatability and reproducibility, one cannot know if the same examiner would reach the same conclusion on different days, or whether a different examiner would be likely to reach the same conclusion.

5. Currently the error rate for firearms examinations is unknown. The "science" behind zero or near-zero error rates claimed by firearm and toolmark examiners is not defensible. The largest studies that I have found in the references provided by the prosecutor use only a few types of guns, a couple of types of ammunition, and the examiners are told they are being tested instead of examiners working in a blind fashion. These separately inadequate "studies" (really, as I discuss later, these are not studies but proficiency tests) fall far short of genuinely approximating an error rate for the discipline.[1] Without information on the error rates associated with each of the range of firearms and manufacturing techniques (this information is not reported in the literature, as far as I have seen), it is impossible to link marks on a bullet to a particular firearm. The flawed studies cannot be sensibly cobbled together to support identity statements.

6. Thus, neither firearm examiners' claims of uniqueness of markings on bullets and casings, nor their claims that they can distinguish the marks on one firearm from the marks on all other firearms, nor their claims of an error rate approaching zero, are supported by the scientific method, or scientifically valid studies, both of which are required for identity claims to be generally accepted by scientists. For identity claims to be generally accepted, firearm examiners would need to have matching criteria, and would need studies showing that these matching criteria reliably produce an identification to a particular source and no others. This is a belief held by me and generally held by highly esteemed scientists, as reflected by recent National Research Council reports.

7. Two recent publications of the National Academy of Sciences' National Research

[1] Indeed, it is well known that firearms examiners make mistakes on the routine and extremely easy (e.g. examiners are asked to compare bullets fired from two completely different classes of firearm) Collaborative Testing Service (CTS) proficiency tests. A paper by Gryzbowski et al states there is a 1.4% (.6% for incorrect IDs) error rate for these simple proficiency tests. This is astoundingly high given the tests' simplicity. Of course, these proficiency tests are not a substitute for real error rate studies. To serve as a reliable measure of a discipline's error rate, a proficiency test must match the difficultly of the actual casework encountered by examiners.

Council, "Strengthening Forensic Science in the United States: A Path Forward" and "Ballistic Imaging" state – and I agree – that the studies and data accumulated by firearms examiners to date do not demonstrate that firearms impart unique marks on bullets and cartridge cases, or that firearms examiners can reliably connect marks on a bullet or cartridge case to a particular firearm. In other words, the statistical foundations for firearm toolmark identification have not been established. Firearms examiners have failed to employ statistical techniques that could help determine what questions should be asked and what answers can be obtained given the available data.

8. Specifically, the authors of Ballistic Imaging explained that "[u]nderlying the specific tasks with which the committee was charged [i.e., assessing the feasibility of a national ballistics database] is the question of whether firearms-related toolmarks are unique: that is, whether a particular set of toolmarks can be shown to come from one weapon to the exclusion of all others. Very early in its work the committee found that this question cannot now be definitively answered." Ballistic Imaging at 3. This lead the committee to their one and only finding, featured prominently and in boldface:

> **Finding: The validity of the fundamental assumptions of uniqueness and reproducibility of firearms-related toolmarks has not yet been fully demonstrated.** Ballistic Imaging at 3; 81.

As the committee was careful to note, it was not "the function of this committee to assess the general validity of firearms identification and toolmark examination" – in other words, the committee did not determine whether or not firearm examinations are valid. Ballistic Imaging at 81. Indeed, that would require the kind of validity testing described above, as yet not undertaken: "A significant amount of research would be needed to scientifically determine the degree to which firearms-related toolmarks are unique or even to quantitatively characterize the probability of uniqueness." Ballistic Imaging at 3. "At a minimum, assessing the general validity and uniqueness of toolmark evidence would require a much wider range of gun and ammunition selections and firing conditions than was supported in our experimentation through NIST (see Chapter 9). It would also require precise quantification of the myriad sources of variability inherent in the firing of a gun (see Chapter 2). In short, it would be a major undertaking, requiring a sustained program of research over many years." Ballistic Imaging at

18. This is, in essence, what I have described above as necessary to validate the field.

In fact, the Ballistic Imaging committee considered taking it upon themselves to attempt to fill the enormous information gap left by the current research. After Ann Davis, AFTE's president at that time, presented information and literature on firearms and toolmarks to the committee, the committee members asked if AFTE would be willing to participate in a human firearm toolmark error rate study. Ms. Davis indicated that she (and AFTE) was not interested. Because the large scale study the committee envisioned would be impossible without AFTE's participation, and because it was unclear how far the committee could stray from their charge of assessing the feasibility of a national ballistics database, the matter was dropped.

Rather, the committee's analysis ended with the finding that the existing research does not show that firearms identification is reliable. Their finding is an unavoidable judgment of the scientific community, whether intended for the courts, Congress, or other scientists: right now, the firearms toolmark discipline is not based in science, and it would require a considerable amount of research "if the basic premises of firearms identification are to be put on a more solid scientific footing." Ballistic Imaging at 82. And while "stopping short of commenting on whether firearms toolmark evidence should be admissible" as a general matter, the Ballistic Imaging committee was quite firm that "[c]onclusions drawn in firearms identification should not be made to imply the presence of a firm statistical basis when none has been demonstrated." Id. at 82. An identity statement to any degree of certainty would imply a statistical basis that simply does not exist. While not expressly stating as much, the committee opined that identity statements, at least, should be better supported: "statements on toolmark matches (including legal testimony) should be supported by statements the work that was done in the laboratory, by the notes and documentation made by examiners, and by proficiency testing or established error rates for individual examiners in the field and in that particular laboratory, but should not overreach to make extreme probability statements." Id. at 82-83.

9. The more recent NRC "Strengthening Forensic Science" report builds on the NRC "Ballistic Imaging" report, and is even more emphatic in its criticism of firearms and toolmarks.

The committee emphasizes that firearm examiners do not follow the scientific method: "A fundamental problem with toolmark and firearms analysis is the lack of a precisely defined process. . . . This AFTE document, which is the best guidance available for the field of toolmark identification, 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." Strengthening Forensic Science at 155. "Sufficient studies have not been done to understand the reliability and repeatability of the methods" that do exist. Id. at 154. After their review of the literature, they found that "the scientific knowledge base for toolmark and firearms analysis is fairly limited." Id. at 155. "The fact is that many forensic tests—such as those used to infer the source of toolmarks . . . —have never been exposed to stringent scientific scrutiny." Id. at 42. The committee members criticize the lack of standards and lack of scientific basis for error rate determination in no uncertain terms: "the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for estimation of error rates." Id. at 153-54. They note that "[b]ecause not enough is known about the variabilities among individual tools and guns, we are not able to specify how many points of similarity are necessary for a given level of confidence in the result." Id. at 154. In other words, the committee is saying that since the variability between firearms (and the technique by which they were manufactured) is basically unknown, it's not possible to assign any degree of certainty to the claims that firearms examiners are making.

In sum, the committee found that the firearms discipline had not demonstrated that they could link marks on bullets and cartridge casings to a particular firearm. They agreed, however, that they could link marks to a class of weapons. Id. at 154.

10. Further, the authors of "Strengthening Forensic Science" have been unambiguous in their statements about the effects the report should be having in courts. Judge Harry Edwards, co-chair of that committee, was livid when he heard that prosecutors were representing that he had suggested the Report was inapplicable to admissibility determinations and publicly clarified the committee's viewpoint:

> As I explained to the Senate Committee, because the Report presents "findings about the current status of the scientific foundation of particular areas of forensic science," it would be "no surprise if the report is cited authoritatively" by the courts in their assessment of particular cases.
>
> Why was that my prediction? Because it seemed quite obvious, at least to me, that if a particular forensic methodology or practice, once thought to be scientifically valid, has been revealed to lack validation or reliability, no prosecutor would offer evidence derived from that discipline without taking the new information into account and no judge would continue to admit such evidence without considering the new information regarding the scientific validity and reliability of its source.

11. The US National Academy of Sciences is composed of about 2000 leading scientists including all Nobel Laureates in the sciences and medicine. The Academy speaks for the US science community. The Academy seeks highly-regarded and carefully-chosen scientists to serve on standing committees (such as the Committee on Science, Technology, and the Law), as well as empaneled subcommittees such as "Strengthening Forensic Science" and "Ballistic Imaging." Members of committees are unpaid and take their assignments seriously. Selection criteria include excellence in science, relevance of qualifications, and integrity. Both the "Ballistic Imaging" and "Strengthening Forensic Science" committees counted statisticians, engineers, and materials scientists among their ranks, along with other scientists and forensic practitioners. Committee members are informed both by invited guests – for example AFTE president Ann Davis in the case of the "Ballistic Imaging" committee and current SWGGUN member and former AFTE president Peter Striupaitis in the case of the more recent "Forensic Science" committee – and by reading relevant literature.

I have read over 50 papers and one book in firearm toolmarks, visited an automated firearm toolmark lab and co-wrote a preproposal (which will be discussed at greater length later in this affidavit) for a firearm toolmark identification study that was well received by the NIJ. With this background, I am well aware that one can get a very good sense of the firearms comparison field by reading several key papers and discussing the issues with key leaders such as Ann Davis and Peter Striupaitis (both of whom, along with AFTE and SWGGUN, were encouraged to – and did – provide any scientific literature supporting the field to the committee). That is exactly what the NRC committee did. As Judge Edwards described, the committee "carefully considered any peer-reviewed, scientific research purporting to support the validity and reliability of existing forensic disciplines. Additionally, we invited experts in each discipline to refer us to any pertinent research. Committee members and staff spent countless hours reviewing these materials. And before the Report was released, it was peer-reviewed by outside experts in the fields of science, law, and forensic practice."

The papers that the committee reviewed overwhelmingly show a pervasive lack of appreciation for the scientific method, and a lack of understanding of the proper design of a validation study (I detail a few of the numerous failures that plague so-called "validation studies" later in this affidavit).

12. The NRC Strengthening Forensic Science report has been well received in the scientific community. For example, the American Statistical Association's Board of Directors

(BOD) unanimously endorsed[2] the NRC report "Strengthening Forensic Science in the United States: A Path Forward". Toolmarks was one of the specific areas considered by the BOD before the vote. I was one of 6 ASA subcommittee members who provided a position memo to the BOD to help with the vote.

13. As long as the methodology details used in the pattern matching are carefully articulated and there are valid error rates for false identifications and missed identifications, pattern matching is an acceptable scientific technique. Unfortunately, as I and the NRC have found, this is not the case. As Biasotti (a firearms examiner whose attempts to bring a statistical foundation to the discipline are cited in the NRC Strengthening Forensic Science report) has written, "I know a match when I see one" is not good science. Automated toolmark identification (which powers ballistic databases) uses well-defined algorithms. These automated algorithms are testable. It is also possible to test the methods used by human firearm examiners, though they first would have to articulate their matching criteria (which they currently do not do).

No one would want medical drugs or devices to be approved for marketing without well-defined and articulated experiments that emulate actual use. Indeed, the FDA would never approve a drug that had been so minimally tested and with such vague standards that the prescribing information instructed physicians to prescribe whatever amount they felt was "sufficient" without specifications for patient's age, size, medical history and other medications. In order for their methodology to be generally accepted as reliable by the scientific community, the same standards must apply to forensic tests. Transparency rather than indeterminate, inarticulated, and vague matching criteria is required by science and the scientific method.

14. It has been suggested that firearms examiners be allowed to testify to a match to a "reasonable scientific certainty" or a "practical certainty," with both of these described as communicating a "high level of certainty," along the lines of "the likelihood of another firearm having fired these cartridges is so remote as to be considered a practical impossibility." These are exactly the kinds of conclusions that I, the NRC and the general scientific community agree firearms examiners should not be permitted to draw given the stark failures in their methodology and testing outlined above.

[2] The only exception to a unanimous vote was some of the government employees who said they may have a conflict of interest and abstained.

15. The community of statisticians and mathematicians, and scientists in general, do not generally accept the ability of the protocols, procedures, and methodology used by firearm and toolmark examiners to provide a scientifically acceptable estimate that a bullet was fired from a particular firearm. As the NRC put it in "Strengthening Forensic Science", the firearms comparison discipline, like other subjective pattern-matching disciplines, has not yet "been rigorously shown to have the capacity to consistently and with a high degree of certainty support conclusions about 'individualization' (more commonly known as 'matching' of an unknown item of evidence to a specific known source)." Strengthening Forensic Science at 87. Firearms examiners in particular have not only failed to show a "high degree of certainty" in their individualization/identification conclusions, they've failed to show that they reach their conclusions to "a[ny] given degree of confidence." Id. at 155. In other words, firearms examiners have not shown that they can reliably identify a particular firearm, period.

An identity statement to any degree of certainty implies a statistical basis that the scientific community agrees simply does not exist. "Conclusions drawn in firearms identification should not be made to imply the presence of a firm statistical basis when none has been demonstrated." Ballistic Imaging at 82. Not only are "absolute certainty" statements inappropriate, but other formulations – "reasonable scientific certainty" or "practical certainty," both of which communicate a very high level of confidence (just short of absolute) in the result – are equally unsupported by science and unwarranted given the available data.

16. For now, the only scientifically supported opinion, and the only opinion that the scientific community endorses, is markings on a bullet or cartridge case were left by a certain class of firearm, to which the suspect firearm belongs: "The committee agrees that class characteristics are helpful in narrowing the pool of tools that may have left a distinctive mark." Id. at 154. I believe that firearms examiners will be able to establish that they can narrow the pool further, to a subclass of weapons (i.e. same manufacturing batch), however, those studies have not yet been done. They would not be difficult to perform after the community specified its matching criteria to the standards of the mainstream scientific community.

17. Certain "validation studies" were reviewed by the NRC (see e.g. Strengthening Forensic Science at 155 and fn. 65, where the largest scale study of those provided was discussed), taken into account, and still the committee's conclusion was that identity claims were

not warranted. As the NRC pointed out, these "validation studies" "suggest a heavy reliance on the subjective findings of examiners rather than on the rigorous quantification and analysis of sources of variability." Strengthening Forensic Science at 155. That is because these are not validation studies at all, but rather proficiency tests: the studies do not make any attempt whatsoever to characterize the matching criteria across the spectrum of firearms and manufacturing techniques.

Rather, they merely test the ability of (for the majority of these studies)[3] a small group (8 or 9) of examiners in the FBI firearms unit to correctly match bullets or cases back to the correct firearm. The papers simply state that the examiner got it right or wrong, without explaining what matching criteria were used in each (or any) case. They typically test only one type of firearm and hence only one type of manufacturing technique. The tests do not mimic case conditions: they test pristine bullets, fired into a water tank, rather than damaged bullets. The results cannot be reliably extrapolated to other firearms or manufacturing techniques, nor can they be reliably extrapolated to damaged ammunition. Some of these "studies" aren't even looking at firearms.

Significantly, in every one of these cases the examiner was aware that he or she was participating in an "error-rate" study, and knew the goal of the study; i.e., these studies were not conducted in a blind fashion, as is required by good science. The rate of "inconclusive" findings in these studies is extremely high, and may indicate that the examiners, aware of the effect that an incorrect result would have on the discipline, were inordinately, and perhaps uncharacteristically, cautious. For example, in Erich Smith's study, examiners (again, a small group from the FBI) failed to correctly exclude 335 out of 352 true cartridge cases exclusions and 352 out of 352 (i.e. every single) true bullet exclusions, despite the fact that "[t]he bullets and cartridge cases were examined using a comparison microscope to confirm that individual microscopic marks of value were reproduced during test fires." These inconclusive results (and the high levels of inconclusives in other studies) were ignored as harmless in the tabulations and not factored into the calculated error rate (0%). This failure to account for all the data accumulated is unscientific, and the opinion that inconclusives are "harmless" is contrary to history.[4]

---

[3] One of the studies – the Hamby, Brundage study – has a larger number of participants than the others. This allows the error rate estimation for this study to be more precise than the others; however, it does not change the fact that this is a proficiency test and not a validation study. Only one type of firearm is examined – a firearm manufactured in 1985 – and only one manufacturing technique. This study tells us nothing about other manufacturing techniques or firearms.

[4] In the infamous case of Sacco and Vanzetti it is believed that the key evidence that resulted in conviction and ultimate execution was an inconclusive match that was presented as possibly coming from a weapon found in the

Another important consequence of the non-blind administration of these tests is, the test takers were aware that they were participating in a study of consecutively-manufactured firearms; i.e. they were put on notice that the markings from one bullet or casing to the next would have heightened similarity. This minimizes the relevance of these studies. The issue is "can examiners detect differences in consecutively manufactured firearms if they only see evidence from one?" In these studies the examiner gets test fires from many consecutively manufactured firearms and knows that they should look for the closest match out of all of those. In case work, the firearms examiner does not get test fires from other weapons that were made in the same batch. She is not afforded the same opportunity to compare and see which casing is most similar. This is not dissimilar to asking if "identical" twins can be identified. Analogously to the published studies the examiner gets to study both "identical" twins carefully before being tested. The scientific method requires that the examiner only see one twin and not be told that there is a twin before testing.

18. Both I and the NRC have attempted on different occasions to engage members of the firearms discipline to participate in a true and properly-designed error rate study. As mentioned above, Ann Davis, speaking on behalf of AFTE, indicated that she was not interested in participating in an error rate study conducted by the NRC as a component of the Ballistic Imaging report. The failure to obtain valid error rates for the pattern-matching component of error lies on the shoulders of the AFTE community in general.

---

possession of Mr. Sacco. The firearm examiner, Proctor, makes it clear that he knew the bullet could have been fired by any Colt pistol but his testimony gave a different impression. Q Have you and opinion as to whether bullet number 3 was fired from a Colt automatic pistol which is in evidence? To which I answered "I have". He then proceeded Q. And what is your opinion? "My opinion is that it is consistent with being fired by that pistol." Proctor's affidavit continues "That is still my opinion for the reason that bullet number 3, in my judgment passed through some Colt automatic pistol, but I did not intend to by that answer to imply that I had found any evidence that the so-called mortal bullet had passed through this particular Colt automatic pistol and the District Attorney well knew that I did not intend and framed his question accordingly. . . ." Similarly, in the assassination of Marin Luther King an inconclusive bullet was involved. James Earl Ray confessed and then tried unsuccessfully to retract his confession to killing Dr. King. The murder bullet was an inconclusive match to the rifle attributed to Mr. Ray.

Had the proposed research among the DCPDS, DCMPD, and a research team gone forward last year, as discussed later in this affidavit, methodology for placing a probability of a match (to a small group or subclass of weapons), of 20%, 50%, or 80% would likely have been possible. It could not go forward because the DCMPD declined participation. This would be helpful to history and possibly law enforcement. See a column that I wrote in the American Statistical Association newsletter AMSTATNEWS for details: http://magazine.amstat.org/blog/2010/03/01/scipolmar10/. Inconclusive findings can be misleading in their typical presentations to the courts. Inconclusives should never be presented in stronger terms than the make of weapon that fired the shot resulting in the evidence is the same as the defendant's weapon; if there is insufficient data to even say

Imaging report. The failure to obtain valid error rates for the pattern-matching component of error lies on the shoulders of the AFTE community in general.

19. I speak as a member of the mainstream scientific community who overwhelming believe:

a. Firearm toolmarks is being grossly oversold to the courts as far as its accuracy, and reliability.

b. That there is good value in firearm toolmark evidence if properly presented to the court. Research would have to be done to ascertain the strength of its relevance.

c. The forensic community in general and firearm toolmark trade unions such as AFTE has not cooperated as much as they could and should with the mainstream scientific community.

20. In conclusion, firearms examiners do not follow a generally accepted scientific methodology in reaching conclusions that a bullet or casing was fired from a particular firearm. Examiners cannot justify a zero percent error rate as it has already been demonstrated by proficiency testing that the rate is not zero. Neither can examiners justify a near-zero error rate, because the testing has not been done to demonstrate this. To achieve a generally accepted scientific methodology, firearm examiners need to produce detailed standard operating procedures and perform experiments that will support accurate and statistically supported probability statements indicating the significance and likelihood that an evidence bullet or casing was fired from an identified firearm. Accordingly, conclusions of identity to a particular firearm cannot be reliably be reached – in this case or any other – as there is significant dispute as to the methodology employed by firearms examiners and because of the lack of a proven statistical error rate and random match probability.

Signed under penalty of perjury this 21st day of November, 2011.

Clifford Spiegelman