Navigating the Minefield of Pragmatism in Forensic Practice

Science, Law, and Forensics

William A. Tobin

(804) 448-3955
wtobin@feintl.com
wtobin@nexet.net
Science, Law, and Forensics

“What is Forensic....?”

(804) 448-3955  William A. Tobin  wtobin@feintl.com
wtobin@nexet.net
“CSI” Comparison Evidence

- hairs & fibers
- glass fragments
- firearms/toolmarks
- bullet lead compositions
- fingerprints
- paint chips
- GSR (gunshot residue)
- mineralogical debris (e.g., soil)
Riddle of the Day:

Q: What do the following have in common?

A. Lead pencils  
B. Tin foil  
C. Tooth Fairy  
D. Easter Bunny  
E. Forensic science
Riddle of the Day:

Q: What do the following have in common?

A. Lead pencils
B. Tin foil
C. Tooth Fairy
D. Easter Bunny
E. Forensic science \(^{(\text{non-DNA})}\)

A: They’re all fictitious entities.
“Forensic science” is not the infallible practice portrayed by Hollywood
(CSI, NCIS, inter alia)
‘Forensic science’ is considered an oxymoron by mainstream scientific community

“Forensic scientists” are generally not scientists
Definitions of ‘Science’

• The intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through methodical observation and experiment

• systematically organized body of knowledge on a particular subject
Contrary to Hollywood’s portrayal, forensic science is comprised of few scientists. Instead, forensic science is mostly comprised of specialized technicians who simply employ or operate the same procedure or instrument over and over again. . . Moreover, the vast majority of firearms and fingerprint examiners do not even possess an undergraduate degree of any kind, let alone a science degree. Instead, they are sworn law enforcement officers with little to no scientific background or training. . . . “Many autopsies are now performed by individuals without needed training in general pathology and forensic pathology.”

The forensics profession lacks a truly scientific culture—one with sufficient written protocols and an empirical basis for the most basic procedures. This results in an environment in which misconduct can too easily thrive. Stated another way, forensic science needs more science. . . . On a systemic level, perhaps the best example is the [FBI] laboratory, considered the country’s premier crime lab. A 1997 Inspector General’s report on the [FBI] lab found scientifically flawed testimony, inaccurate testimony, testimony beyond the competence of examiners, improper preparation of laboratory reports, insufficient documentation of test results, scientifically flawed reports, inadequate record management & retention, and failures of management to resolve serious and credible allegations of incompetence. . . .

“At present, forensic science is virtually unregulated, with the paradoxical result that clinical laboratories must meet higher standards to be allowed to diagnose strep throat than forensic labs must meet to put a defendant on death row.”

Many of the forensic techniques used in courtroom proceedings... rest on a foundation of very weak science, and virtually no rigorous research to strengthen this foundation is being done. Instead, we have a growing body of unreliable research funded by law enforcement agencies with a strong interest in promoting the validity of these techniques. This forensic “science” differs significantly from what most of us consider science to be.

Scientists vs. Technicians

**Scientist:** A person who studies natural phenomena and develops theories in a systematic manner through hypothesis testing via the scientific method.

**Technician:** A person with a mid-level understanding of theory and high-level mastery of technique.
Forensic Lab Staffing & Supervision

- examiners: barely trained technicians
- 96% bachelor degrees (or less)
- 3% master’s degrees
- 1% Ph.Ds

With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.¹

Specific Source Attributions
(aka, ‘Individualizations’)
Specific Source Attributions (aka, ‘Individualizations’) 

“Exhibit [A] is a ‘match’ to Exhibit [B]”

“C came from D”

“Item ‘Q’ originated from Item ‘K’”

(where ‘Q’ = questioned and ‘K’ = known)

“The Q1 bullet was fired from the K1 firearm”

Note: such opinions are inherently probabilistic
Specific Source Attributions (aka, ‘Individualizations’) 

• Unique to forensic inquiry
• Inherently probabilistic
• Typical assumptions:
  • ‘representativeness’
  • uniqueness
  • discernible uniqueness
  • probative value
Specific Source Attributions
(aka, ‘Individualizations’)

“The Individualization Fallacy in Forensic Science Evidence”,
Michael J. Saks & Jonathan J. Koehler

“Individualization Claims in Forensic Science: Still Unwarranted”,
Jonathan J. Koehler & Michael J. Saks
75 Brook. L. Rev. 1187, 1187–88 (2010)
Definition

probative value

Evidence which is sufficiently useful to prove something important in a trial; the ability of an item of evidence to make a relevant disputed point more or less true.
How Were Forensic Practices Developed?
How Forensic Practices Developed

- Intuitively
- By observational study
- Using simple enumeration
- With no design of experiments
- No studies of reliability confidence
Development Characteristics of Forensic Practices

- Developed by *inductive* process...
- ...by nonscientists
- Not developed by scientific method
- No comprehensive, meaningful testing
- Insular community *(peer-review implications)*
- Absence of cross-discipline input
- Absence of extrajudicial interest
- Note: admissibility ≠ validation
  
  (“implicit testing” implications of *stare decisis*)
Definitions

- **reliability**: same result over repeated measurements
- **validity**: measures what it purports to measure
- **accuracy**: conformity to a true value
- **precision**: (a) least variation from set (b) refinement of measure
- **repeatability**: same experimenter replication
- **reproducibility**: different experimenter replication
Precision vs. Accuracy

precision

accuracy
Caution: Beware ‘pathological science’ & Fallacy of the Converse (blue Chevy Novas & lightbulbs)
The Scientific Method

Pathological science

A term believed to have been first used by Nobel Laureate Irving Langmuir in his presentation at a colloquium at The Knolls Research Laboratory, December 18, 1953.

Characterizes situations, with no implications of dishonesty, where people are influenced into mistaken beliefs or false interpretations of results resulting from a lack of understanding about how humans can deceive themselves and be led astray by subjective influences, wishful thinking, cognitive biases, or unforeseen interactions between or among input variables known as “threshold interactions.” *

Keystones of Scientific Method

Falsifiability
Protocol
Parameters of Detection
Rules of Parameter Application
Repeatability
Reproducibility
Scientific Method

- Scientific Method aims to eliminate subjectivity

- No science allows inference based on 100% subjectivity ("training & experience")

- F/TM (aka, ‘ballistics’), bitemarks, *inter alia*, are virtually 100% subjective practices
Firearms/toolmarks identification is virtually a 100% subjective pattern-matching practice with no objective indicia for examinations or validation, and no feedback loop in casework.
AFTE Theory of Identification

“... toolmarks enables opinions of common origin to be made when the unique surface contours of two toolmarks are in sufficient agreement.”

“... [t]his sufficient agreement is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges, and furrows. Specifically, the relative height or depth, width, curvature, and spatial relationship of the individual peaks... Agreement is significant when it exceeds the best agreement demonstrated between two toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that sufficient agreement exists between two toolmarks means that the likelihood another tool could have made the mark can be considered a practical impossibility.

The current interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience.”
In other words...

The examiner is advised to think back to best matching non-match she can remember. If she can’t remember a better match than what she is seeing now, it’s a “practical impossibility” that the current match is coincidental.
Firearms/Toolmarks Identification Is Not Science

• No protocol
• No parameters of detection
• No rules of parameter application
• 100% subjective practice
• No repeatability
• No reproducibility
• No science allows 100% subjectivity
• No science allows validation based on ‘training & experience’
Why Lengthy Admissibility?

- 1-dimensional domain development
- No extrajudicial scrutiny
- No process dissection
- Focus on instrumentation, not “task at hand”
- Extensive analytical technique papers
- Entrenchment with nuclear reactor
- *Laissez-faire* judging
Methods of Inference

- Deductive
- Inductive
- Abductive
Methods of Inference

- **Deductive**: general to specific
- **Inductive**: specific to general
- **Abductive**: inference to best explanation
Deduction

ONLY when you can characterize entire population (sample pool)

Logic process: general → specific

Example: blue Chevy Nova metaphor
Induction

all other sampling regimes

Logic process: specific $\rightarrow$ general

**NOTE**: All forensic practices are based on empirical induction.

**Example**: shoe, car purchases
Deduction vs. Induction

Deductive
- low risk of error,
  forensically unrealistic

Inductive
- Much higher vulnerability to error
- All forensic practices are inductive
- Inferences must be appended with probabilistic expression
Logical Argument Classifications

**Induction**

**Caution 1**: Most vulnerable to fallacious reasoning from various fallacies of presumption, such as false dichotomy, *petitio principii*, suppressed evidence, *inter alia*.

**Caution 2**: There exists no combination of samples that will allow 100% confidence (as exists for deductive reasoning).
Induction

• Cannot “prove” inductive hypothesis by accumulation of positive instances from a population (simple enumeration)

• “Training & experience” is unacceptable basis for scientific proof or validation
Risks of Induction

• May be circumstance-acceptable if limits understood (e.g., shoes, car)

• ‘Proving’ vs. corroborating hypothesis

• Infer future outcomes from past (instances of which we have had no experience resemble those of which we have had experience)
Example of risk relating to simple enumeration

Assume:

- 100,000 firearms in region
- 100 pairs are confusingly similar
- 100 firearms examiners at crime lab
- Each conducts 10 pairwise exams
- Every day for 10 years

[See Karl Popper’s, “The Problem of Induction” re black swans & falsifiability]

Example of risk relating to simple enumeration

Assume:
100,000 firearms in region
100 pairs are confusingly similar
100 firearms examiners at crime lab
Each conducts 10 pairwise exams
Every day for 10 years

After 3,650,000 pairwise examinations over 10 years, there remains a 93\% chance none of the 100 pairs of similar firearms were examined!

[See Karl Popper’s, “The Problem of Induction” re black swans & falsifiability]
Flawed Logic Processes
fallacy:

mistaken belief, especially based on unsound judgment
Fallacies of Presumption
(premises presuming what they purport to prove)

- False dichotomy
- Suppressed evidence
- *Petitio principii* (begging the question)
Fallacies of Presumption

• False dichotomy

Presenting two nonjointly exhaustive alternatives as if they were jointly exhaustive, then eliminating one, leaving the other as inference (conclusion)
Fallacies of Presumption

- Suppressed evidence

Arguer ignores relevant evidence that outweighs presented evidence and entails very different inference.
Fallacies of Presumption

- **Petitio principii** (begging the question)

  truth of the conclusion is assumed by the premises

  Arguer uses phraseology that tends to conceal the questionably true character of a key premise
AFTE Theory of Identification

“... toolmarks enables opinions of common origin to be made when the unique surface contours of two toolmarks are in sufficient agreement.”

“... [t]his sufficient agreement is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges, and furrows. Specifically, the relative height or depth, width, curvature, and spatial relationship of the individual peaks... Agreement is significant when it exceeds the best agreement demonstrated between two toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that sufficient agreement exists between two toolmarks means that the likelihood another tool could have made the mark can be considered a practical impossibility.

The current interpretation of individualization/identification is subjective in nature, founded on scientific principles and based on the examiner’s training and experience.”
Q: How do you evaluate a forensic practice?

A: Like eating an elephant
Necessary Phases of Forensic Practice

- Evaluations/Analyses
- Comparisons ("matching", "grouping")
- Likelihood (probative value)
- Inference (conclusion)
Evaluating Phases of Forensic Practice

- Insure each “if” in every “if-then” step is valid *
- Verify that progression between each phase of logic process is valid (“then”)

“Then” becomes next “if”

- Is conclusion based on valid logic or unjustified extrapolation? *
- What premises/assumptions are required for validity of inference?
Evaluating ‘Validation’ Studies
Evaluating Purported ‘Validation’ Studies

- What is the hypothesis to be tested in each study?
- What is the population of interest (target population)?
- What is the study sample population?
- Does the sample population coincide with the target population?
- Are the study inferences within acceptable limits of the experimental (sample) frame?
- Are the opinions offered in judicial proceedings supported by the claimed “validation studies”? In other words, does the purpose for which the “validation studies” are offered to the court, express or implied, exceed the limitations of the study(s)?
Numerous studies are presented as “validation” studies, including various “10-gun studies”, one or more with hundreds of respondents.
“Analysis of Experiments in Forensic Firearms/Toolmarks Practice Offered as Support for Low Rates of Practice Error & Claims of Inferential Certainty”


http://ssrn.com/author=1521077
Internal vs. External Validity

**Internal validity**: basic minimum without which any experiment is uninterpretable (“Did experimental treatments make a difference in this specific experimental instance?”)

**External validity**: generalizability. (“To what populations, treatment variables, *inter alia*, can this effect be generalized?”)
Some of the reasons for no external validity:

- Do not mirror real-world conditions or evidence
- Pristine samples (undamaged)
- Pre-examined/selected for clarity
- No collaborative effort monitoring
- No access to ground truth (except in proficiency tests)
- Deductive, not inductive, inference process
- No parameters of detection
- No rules of parameter application
- 100% subjective declaration (Miller & Neel: Type III error)
- Not double-blind or blind
- “Inconclusives” skyrocket (+40% in one CTS test)
- Samples irrelevant for universal application
- “Laughingly easy” (Scotland Yard & U.S. examiners)
Mistaken beliefs or false interpretations of results resulting from a lack of understanding about how humans can deceive themselves and be led astray by subjective influences, wishful thinking, cognitive biases, or unforeseen interactions between or among input variables known as “threshold interactions.”
Two Required Premises for Validity of Specific Source Attributions

• Uniqueness
• Repeatability

1Repeatability and reproducibility have different and specific meanings in the true scientific method to denote experimental outcome than they are used in forensic firearms-toolmarks (F/TM) practice to denote persistence in transfer of observed characteristics. In firearms identification practice, the terms are used somewhat interchangeably to denote the persistence in transfer of characteristics from tool to work piece or other receptor material rather than a scientific process itself.
Uniqueness

• How unique is ‘unique’?
• Two issues:
  • Ethereal, idealistic: existence of
  • Pragmatic: discernible
“Proficiency” Testing

- Are the results of proficiency testing claimed as rate of error for *Daubert*?
- Is “proficiency” testing in-house?
- Does “proficiency” testing cover ALL phases of forensic process?
- Who selects test samples?
- Fantasyland samples or real-world?
- Who grades/evaluates test results?
- Have you ever failed a proficiency test?
- Anyone in your unit ever failed test?
- Blind or double-blind testing?
Vulnerabilities of Proficiency Testing

• Neither double-blind nor blind
• Deductive test regime
• Respondents’ comments on public record that they insultingly and laughably easy
• “Inconclusives” not incorrect responses (rate of “inconclusive” rises dramatically (quadrupled in one study) during proficiency testing as compared to casework)
• No control over collaborative efforts
• Samples do not mirror real-world casework [provided in “pristine condition” (undamaged)]
• Test samples preexamined to insure clarity
• *inter alia*
Forensic System Challenges

- No feedback loop; no benefit from experience
- “Expert” syndrome; feel like they *have to* have opinion; some just can’t say “I don’t know”
- Lack of defined, relevant academic program
- Examiners inexperienced in real-world science, deductive & inductive logic, scientific methodology
- Inherent bias as law enforcement laboratories
  - “us vs. them”
  - camaraderie, “blue wall” peer pressure loyalty exceeds scientific loyalty
Forensic System Challenges

- How do you validate a 100% subjective practice?
- Institutional loyalty exceeds scientific loyalty
- Desire to please client; “pushing” inference beyond scientifically supportable limits
- Desire for notoriety, recognition; “hotdogging”
- No testimony monitoring program internally
- No testimony monitoring program externally
- Internal “reviewers” of similar backgrounds (resulting in “confirmation bias”)
- Supervisor, managers are bureaucrats; career & “not on my watch” concerns
Observer effects

- Expectation bias
- Confirmation bias
- Contextual bias
Observer Bias Question:

Does Position of Expert’s Client Affect Judgment & Decision-Making?
Does Position of Expert’s Client Affect Judgment & Decision-Making?

In asbestos damage study, plaintiff-hired radiologists found evidence of possible asbestos-related abnormalities in 95.9% of the cases presented.*

When independent radiologists at Johns Hopkins University reinterpreted the same x-rays, they found abnormalities in just 4.5% of the cases.

“Hypothesis Testing of the Critical Underlying Premise of Discernible Uniqueness in Firearms/Toolmarks Forensic Practice”

53 Jurimetrics Journal 121-142 (2013)

http://ssrn.com/author=1521077
“Analysis of Experiments in Forensic Firearms/Toolmarks Practice Offered as Support for Low Rates of Practice Error & Claims of Inferential Certainty”


http://ssrn.com/author=1521077
Navigating the Minefield of Pragmatism in Forensic Practice

William A. Tobin

(804) 448-3955  wtobin@feintl.com  wtobin@nexet.net