NIST Interagency Report
NIST IR 8510sup2

Communicating Forensic Findings: Current Practices and Future Directions

Workshop Presentations

Sandra L. Koch

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8510sup2

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Interagency Report
NIST IR 8510sup2

Communicating Forensic Findings: Current Practices and Future Directions

Workshop Presentations

Sandra L. Koch
Special Programs Office
Laboratory Programs

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8510sup2

March 2025

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Craig Burkhardt, Acting Under Secretary of Commerce for Standards and Technology and Acting NIST Director

NIST IR 8510sup2
March 2025

Presentations can be accessed at: https://www.nist.gov/news-events/events/2024/06/communicating-forensic-findings-current-practices-and-future-directions

Publication History
Approved by the NIST Editorial Review Board on 2025-03-07

How to Cite this NIST Technical Series Publication
Koch SL (2025) Communicating Forensic Findings: Current Practices and Future Directions: Workshop Presentations, *Supplemental Document to Communicating Forensic Findings: A NIST Scientific Foundation Review* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8510sup2. https://doi.org/10.6028/NIST.IR.8510sup2

Author ORCID iDs
Sandra L. Koch: 0000-0001-5872-2266

Contact Information
sandra.koch@nist.gov

June 25-26, 2024
Rockville, MD

Communicating Forensic Findings: Current Practices and Future Directions

Sandra L. Koch, Ph.D.
NIST Special Programs Office

NIST | NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Workshop Planning Team and Moderators

Sandy Koch
(chair)

Will Guthrie

Hari Iyer

Steve Lund

John Butler

Sanne Aalbers

Melissa Taylor

Sheila Willis

Meeting notes

- This meeting will not be recorded
- We are taking notes to write up a summary of what was presented and discussed to post on the website publicly afterward
- Thank you!
  - Darby Harris
  - Kelsey Johns
  - Holly Zhao

Day 1

Session 1: Framing the issues

Session 2: Communicating scientific evidence

- Forensic Report
- Expert Testimony
- Interpretation

Session 3: Practitioner perspectives

- Trace materials
- Footwear & Tire tread
- DNA
- Documents
- Digital/ Multimedia

Session 4: Group Discussion

Day 2

Session 5: Gaps

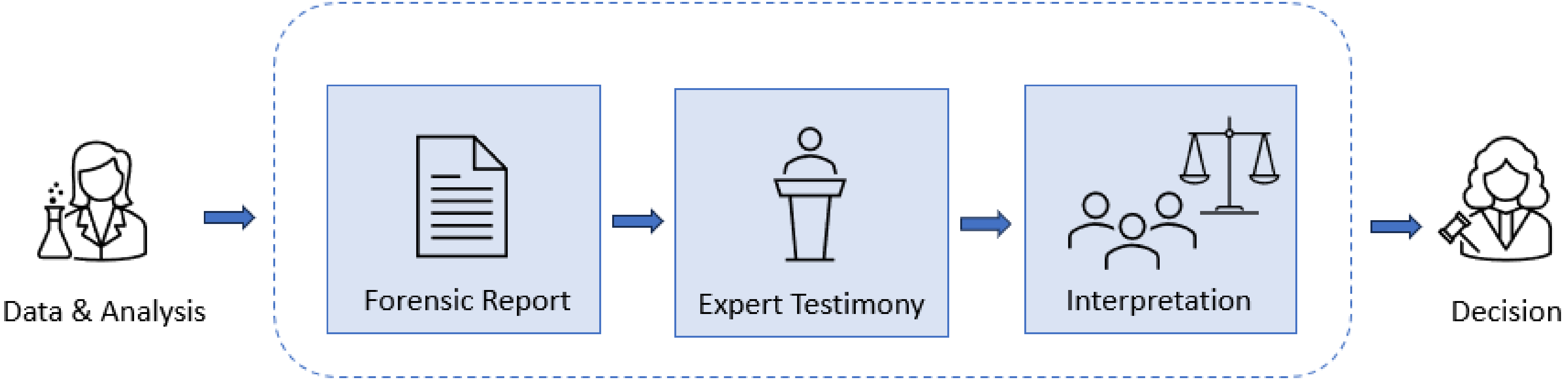
Session 6: European perspectives and practices

Session 7: Breakout discussions

Session 8: Why this Matters?

Wrap up review and What’s Next?

How are forensic findings communicated and understood? And what can be improved?



Input

What is reported?

What is included/excluded?

How is it reported (LR, RMP, verbal, etc.)?

What is presented?

Who is presenting?

How is it presented (proposition level, scale)?

What is understood?

How are results interpreted?

Outcome

Goals for Today's Workshop:

- To hear from a diverse group of speakers and stakeholders
- Time for discussion has been built into the schedule:
  - **Questions and comments are encouraged!**
  - **Please be open to different perspectives that will be shared today and contribute to discussions in a collegial manner**

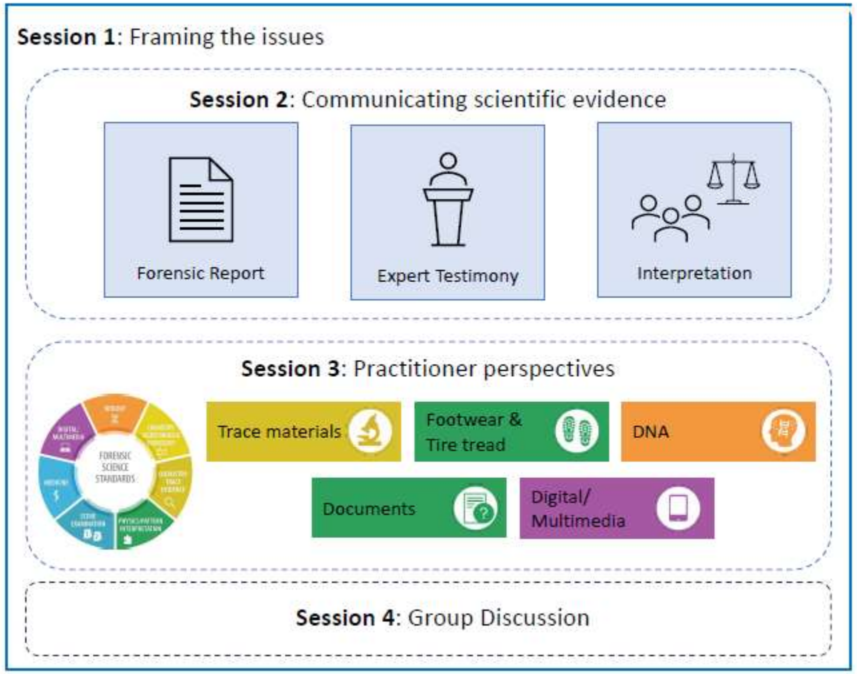
Goals for Today's Workshop:

- This workshop is intended to provide the focus for our next Scientific Foundation Review
  - And we will also be looking for potential participants to assist our project team in drafting the NIST report (there will also be an opportunity during the public comment period to provide feedback)
  - We have started to gather references so if there are publications that you want us to be made aware of, please feel free to send those citations or articles my way (sandra.koch@nist.gov)

Session 1: Framing the Issues

The speakers in this session were asked to set the stage for the rest of the workshop so we have a similar framework to understand the issues

- John Butler NIST Special Programs Office
- Judge J. Michael Ryan DC Superior Court
- Hal Stern UC Irvine/CSAFE
- Steve Lund and Hari Iyer NIST Statistical Engineering Department



Communicating Forensic Findings (CFF) Workshop

Rockville, MD

25 June 2024

NIST Scientific Foundation Reviews

John M. Butler, Ph.D.

NIST Special Programs Office

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Forensic Science Program

https://www.nist.gov/spo/forensic-science-program

Special Programs Office Shyam Sunder

Research at NIST in 8 focus areas:

- Biometrics
- Digital Evidence
- Drugs and Toxicology
- Evidential Statistics
- Firearms and Toolmarks
- Forensic Genetics
- Forensic Science Quality Assurance
- Trace Evidence

Robert Ramotowski

Standards efforts involve administering **OSAC**

22 forensic disciplines with **>800 participants** from across the community

OSAC REGISTRY

190 forensic science standards
(as of 4 June 2024)

OSAC LEXICON

>4,000 terms
organized by forensic discipline

OSAC IMPLEMENTER

>150 implementers
forensic science service providers

John Paul Jones

Foundation Studies


NISTIR 8225

NIST Scientific Foundation Reviews

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Associate Director of Laboratory Programs

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs
Director's Office

December 2020

U.S. Department of Commerce

National Institute of Standards and Technology

- DNA Mixture Interpretation
- **Digital Investigation Techniques**
- **Bitemark Analysis**
- Firearm Examination
- Footwear & Tire
- Communicating Findings (LR)

John Butler

Scientific Foundation Studies

https://www.nist.gov/forensic-science/interdisciplinary-topics/scientific-foundation-reviews

Goal: Identify the scientific foundations that support and underpin forensic methods and document and assess empirical evidence for the reliability of these methods using publicly available data and peer-reviewed literature.

a topic is selected

Initial Input
(Resource Group, **Workshop**, Interlab Study, etc.)

NIST Process

A Study Team Works

DRAFT Report

Consider Public Comments Received

FINAL Report

Public Comments on Draft Report

Our approach to conducting these studies, also known as technical merit evaluations, is described in NIST Interagency Report NISTIR 8225: ***NIST Scientific Foundation Reviews***

NIST Foundation Study Reports

NIST Internal Report
NIST IR 8354

Digital Investigation Techniques: A NIST Scientific Foundation Review

James R. Lyle
Barbara Guttman
Software and Systems Division

John M. Butler
Kelly Sauerwein
Christina Reed
Corrine E. Lloyd
Special Programs Office

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8354

November 2022

November 2022
84-pages

Digital evidence examination **rests on a firm foundation based in computer science.** Extensive testing of over 250 widely used digital forensic tools showed that **most tools perform their intended functions with only minor anomalies**.

NIST Interagency Report
NIST IR 8352

Bitemark Analysis:
A NIST Scientific Foundation Review

Kelly Sauerwein
John M. Butler
Christina Reed
Division 602 - Special Programs Office
Laboratory Programs
National Institute of Standards and Technology

Karen K. Reczek
Division 601 – Standards Coordination Office
Laboratory Programs
National Institute of Standards and Technology

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8352

March 2023

March 2023
42-pages

Supplemental Documents

114 pages — **NIST Interagency Report NIST IR 8352sup1**
CSAFE Bitemark Thinkshop Report
Center for Statistics and Applications in Forensic Evidence (CSAFE)

10 pages — **NIST Interagency Report NIST IR 8352sup2**
Standards and Guidelines in Forensic Odontology
Karen K. Reczek

13 pages — **NIST Interagency Report NIST IR 8352sup3**
Summary of Published Criticisms of Bitemark Foundations and Responses by Forensic Odontologists
John M. Butler

30 pages — **NIST Interagency Report NIST IR 8352sup4**
Bitemark Analysis Reference List
Kelly Sauerwein

Forensic bitemark analysis **lacks a sufficient scientific foundation** because the three key premises of the field are not supported by the data.

NISTIR 8351-DRAFT

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Melissa K. Taylor
Sheila Willis*
Special Programs Office
Associate Director of Laboratory Programs

Hari Iyer
Statistical Engineering Division
Information Technology Laboratory

Peter M. Vallone
Biomolecular Measurement Division
Material Measurement Laboratory

Rich Press
Public Affairs
Director's Office

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351-draft

June 2021

DRAFT
June 2021
250-pages

Received extensive public comments (~500 pages) that are being considered along with additional information since June 2021. **We will release a final report when completed**.

https://www.nist.gov/forensic-science/interdisciplinary-topics/scientific-foundation-reviews

Why Study Communicating Forensic Findings?

- Identified as a need in the DNA Mixture Interpretation Draft Report
  - **Key Takeaway #4.8**: We encourage a separate scientific foundation review on the topic of likelihood ratios in forensic science and how LRs are calculated, understood, and communicated.

- The planning committee felt it would be useful to expand this workshop to cover communicating forensic findings rather than simply discussing likelihood ratios

- NIST previously held **two workshops (in May 2016 and June 2017) on quantifying the weight of forensic evidence** with some helpful content and discussions to build upon…

Previous NIST Workshops
on Quantifying the Weight of Forensic Evidence

- **May 5-6, 2016**:
  - (Presentation slides available) https://www.nist.gov/itl/iad/image-group/tc-quantifying-weight-forensic-evidenceonline-proceedings
  - (Presentation videos available) https://www.nist.gov/news-events/events/ibpc-technical-colloquium-quantifying-weight-forensic-evidence
  - (Bibliography of 21 key articles) https://www.nist.gov/itl/iad/image-group/quantifying-weight-evidence-reading-material
- **June 27-29, 2017**:
  - https://www.nist.gov/itl/iad/image-group/technical-colloquium-quantifying-weight-forensic-evidence

Approach to Conducting These Foundation Studies

Our approach to conducting these studies, also known as technical merit evaluations, is described in NIST Interagency Report NISTIR 8225: ***NIST Scientific Foundation Reviews*** and generally follows these steps:

We are here

1. A forensic discipline, method, and/or practice is selected for study
2. Publicly available scientific literature and information are gathered
3. **A workshop may be held seeking input from members of the community**
4. Team of NIST scientists and outside experts meet, discuss, and draft report and supplemental documents
5. Information is shared and received at forensic conferences during the deliberation phase
6. Draft reports are made available for public comment along with supplemental documents and all public comments received are shared
7. After considering public comments, reports are finalized and made available on NIST website

https://www.nist.gov/forensic-science/interdisciplinary-topics/scientific-foundation-reviews

Thank you for your attention!

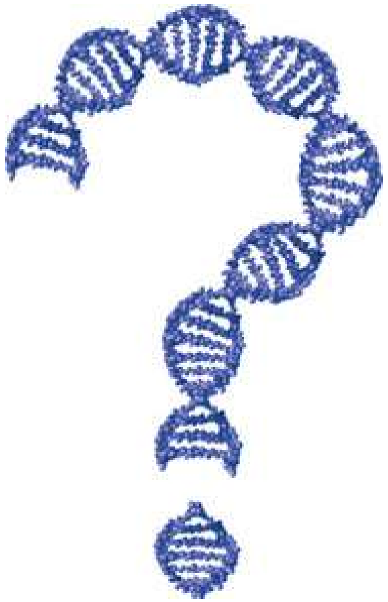
Acknowledgments:
- Congressional funding and NIST Special Programs Office
- Planning team: Sandy Koch, Sanne Aalbers, John Butler, Will Guthrie, Hari Iyer, Steve Lund, Melissa Taylor
- Logistics: Corrine Lloyd, Donna Ramkissoon, Dalia Travis, Pauline Truong
- Thank you for attending and participating over these two days

Sandy Koch
sandra.koch@nist.gov

John Butler
john.butler@nist.gov

https://www.nist.gov/spo/forensic-science-program

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE



Questions?

Feel free to contact us for further information

[1 slide 702]

Rule 702. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if the proponent demonstrates to the court that it is more likely than not that:

(a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on sufficient facts or data;

(c) the testimony is the product of reliable principles and methods; and

(d) the expert's opinion reflects a reliable application of the principles and methods to the facts of the case.

[SLIDE 2]

The amendment continues the practice of the original Rule in referring to a qualified witness as an "expert." This was done to provide continuity and to minimize change. The use of the term "expert" in the Rule does not, however, mean that a jury should actually be informed that a qualified witness is testifying as an "expert." Indeed, there is much to be said for a practice that prohibits the use of the term "expert" by both the parties and the court at trial. Such a practice "ensures that trial courts do not inadvertently put their stamp of authority" on a witness's opinion, and protects against the jury's being "overwhelmed by the so-called 'experts'.

[SLIDE 3]

Instruction 2.215 SPECIALIZED OPINION TESTIMONY

[FORMERLY EXPERT TESTIMONY]

In this case, [you will hear] [you heard] the testimony of [name of witness] who [will express] [expressed] opinions concerning [certain subjects; specify the subject(s), if possible]. If scientific, technical, or other specialized knowledge might assist the jury in understanding the evidence or in determining a fact in issue, a witness who possesses knowledge, skill, experience, training, or education may testify and state an opinion concerning such matters. You are not bound to accept this witness's opinion. If you find that the opinion is not based on sufficient education or experience, that the reasons supporting the opinion are not sound, or that the opinion is outweighed by other evidence, you may completely or partially disregard the opinion. You should consider this evidence with all the other evidence in the case and give it as much weight as you think it fairly deserves. [During the testimony of [an] expert witness[es] in this case, you have heard [an] [more than one] expert refer to information that was not otherwise introduced or admitted into evidence. This information is relevant only to explain what the expert[s] relied upon in forming his/her/their opinion[s]. You may not consider the expert's testimony to be evidence of the truth of that information. You may consider this information only for the purpose of evaluating the expert's opinion and not for any other purpose.]

1 Criminal Jury Instructions for DC Instruction 2.215 (2024)

[SLIDE DAUBERT QUESTIONS]

Questions to assess reliability—whether a theory or technique is scientific knowledge that will assist the trier of fact?

- Whether it can be (and has been) tested
- Whether the theory or technique has been subjected to peer review and publication
- What is the known or potential rate of error?
- Are there standards maintained which control the technique's operation?
- General acceptance can have some bearing: a reliability assessment permits identification of a relevant scientific community and determination of a particular degree of acceptance within that community.

[SLIDE 4 - TIMELINE]

1990s Daubert, Kumho Tire, new FRE 702

2000 Joseph P. Bono affidavit. US v. Curtis, 755 A.2d 1011 (D.C. 2000)

2006 NATIONAL ACADEMY OF SCIENCES: Strengthening Forensic Science in the US report

2010 In Re O.W., 09-DEL 1977 (Ryan, J. April 2, 2010)

2016 PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods

2019 US v. Tibbs, 2016 CF1 19431 (Edelman, J. Sept. 5, 2019)

2024 US v. Green, 2018 CF1 4356 (Okun, J. April 1, 2024).

[SLIDE 4A]

2000

Affidavit from Joseph P. Bono, the Director of the DEA Mid-Atlantic Laboratory "noted that tests and instruments that are properly used by qualified forensic chemists are incapable of producing a false positive."

US v. Curtis, 755 A.2d 1011, 1013 n.7 (D.C. 2000)

2006 NAS report

[SLIDE 4B]

2010

“At that hearing, Ms. H--------- testified that there is a zero percent (0%) error rate associated with the combined three test procedures used here to identify the unknown, seized substance[]”

While explaining that each of these tests used alone is presumptive, as distinct from confirmatory, [] Ms. H-------- nonetheless maintained their infallibility when used in concert. With the designation that these tests are merely presumptive, the DEA chemist acknowledged that there is some degree of inherent error calculable with respect to each of the tests when they are performed in isolation. That there is some distinct and additional degree of error calculable with respect to this analyst’s performance of each test is also without question.

[T]he assertion that the combination of the three tests in question used for the identification of marijuana is infallible, coupled with the claim by the DEA forensic chemist of her own zero percent (0%) error rate in conducting these tests, and her vague allusions to the existence of “different studies that have been introduced today and that are at my laboratory” supporting this claim, are sufficient to indicate a possible flaw in the testing procedures, thus warranting the government’s compliance with Respondent’s discovery request...

In Re O.W., 09-DEL 1977 (Ryan, J. April 2, 2010)

[SLIDE 4C]

2016 PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods

2019 US v. Tibbs, 2016 CF1 19431 (Edelman, J. Sept. 5, 2019)

2024 US v. Green, 2018 CF1 4356 (Okun, J. April 1, 2024).

[SLIDE 4D]

2019

Based largely on the inability of the published studies in the field to establish an error rate, the absence of an objective standard for identification, and the lack of acceptance of the discipline's foundational validity outside of the community of firearms and toolmark examiners, the Court precluded the government from eliciting testimony identifying the recovered firearm as the source of the recovered cartridge casing. Instead, the Court ruled that the government's expert witness must limit his testimony to a conclusion that, based on his examination of the evidence and the consistency of the class characteristics and microscopic toolmarks, the firearm cannot be excluded as the source of the casing.

US v. Tibbs, 2016 CF1 19431 (Edelman, J. Sept. 5, 2019)

[SLIDE 4E]

2024

First, the Court will not permit the examiner to state that his conclusions are to a 100% certainty, to a reasonable degree of scientific certainty, or based on a comparison to all other firearms or toolmarks. Second, the examiner will have to qualify his opinion by stating that his conclusions are not based on a statistically derived or verified measure and that there is not a generally accepted statistical method for conveying the weight of an identification. Third, the examiner will have to qualify his opinion by making clear that his opinion is based on his subjective determination of sufficient agreement in individual characteristics or random imperfections. And finally, the examiner will have to qualify his opinion by testifying that the relevant cartridge casings are "consistent with" having been fired from the firearm at issue, not that they were fired from the firearm at issue, and not even, as the Government has proposed, that there is extremely strong support for the proposition that the casings were fired from the firearm at issue.

US v. Green, 2018 CF1 4356 (Okun, J. April 1, 2024).

Communicating Forensic Findings: Framing the Issues

NIST June 25-26, 2024

Hal Stern
Department of Statistics
University of California, Irvine
sternh@uci.edu

UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

csafe
Center for Statistics and Applications in Forensic Evidence

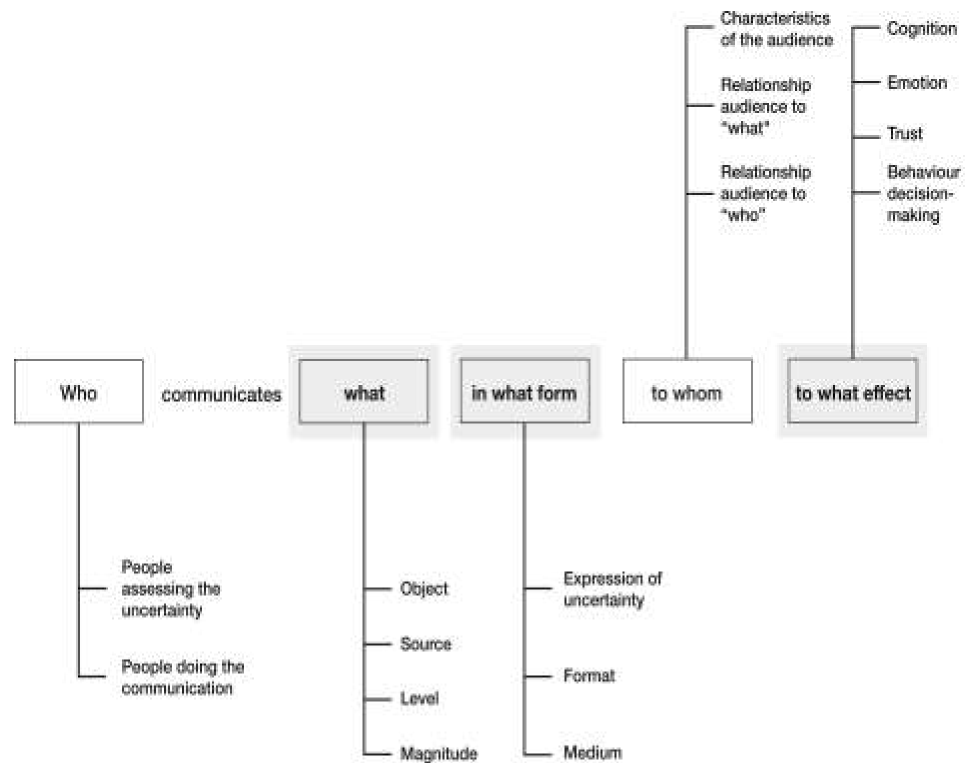
Communicating Uncertainty

- August 2014 NIST Federal Funding Opportunity :

  "A critical need in the forensic science research community is a **more thorough understanding and contextualizing of the uncertainty associated with scientific measurements** and/or analytical techniques. **Reporting uncertainty in forensic science measurements is currently an uncommon practice**, largely because the forensic science community demands an unequivocal conclusion of a binary analysis .."

Communicating Uncertainty

- van der Bles et al., 2019, Royal Society Open Science “Communicating Uncertainty About Facts, Numbers and Science”
- Framework:
  - Who is communicating?
  - **What are they communicating?**
  - **In what form is the uncertainty communicated?**
  - Communicated to whom?
  - Communicated to what effect?



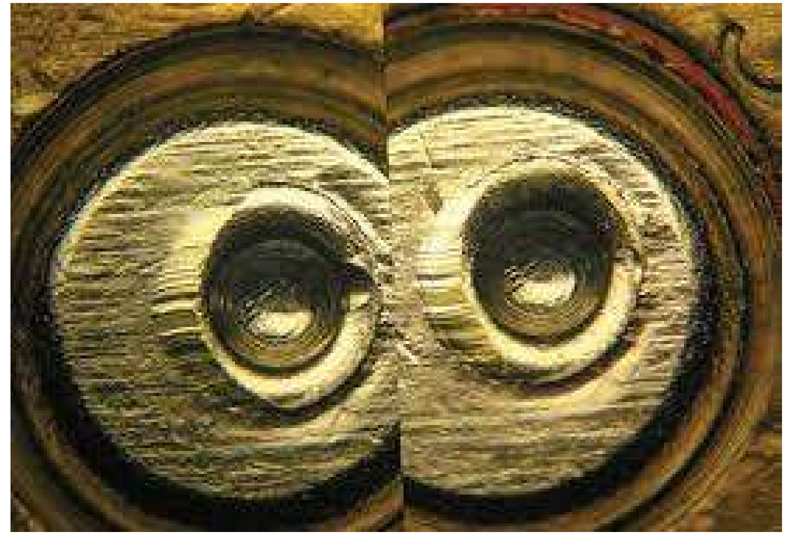
Communicating Uncertainty

- Who is communicating?
  - Forensic examiner
  - Attorney (prosecutor / defense)
- Communicated to whom?
  - Trier of fact (jury or judge)
    - 2018 – Only 2% of federal criminal cases went to jury trial
    - 2013-14 – Only 2% of felony cases in CA went to jury trial
  - Investigators
  - Attorneys
- Communicated to what effect?
  - Primarily about decision-making
  - Decision making by jury, investigators, attorneys

What is being communicated?

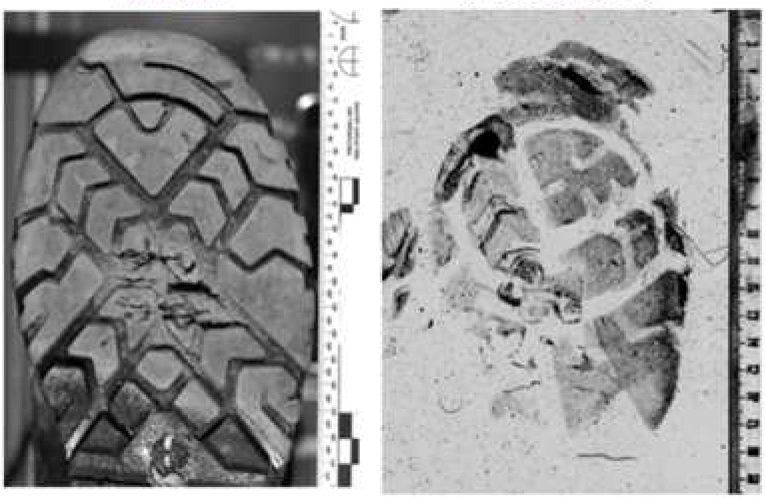
- The task of interest for purposes of this presentation:
  assess two items of evidence,
  one from a known source and one from an unknown source,
  to assess the proposition that the two samples originate from the same source

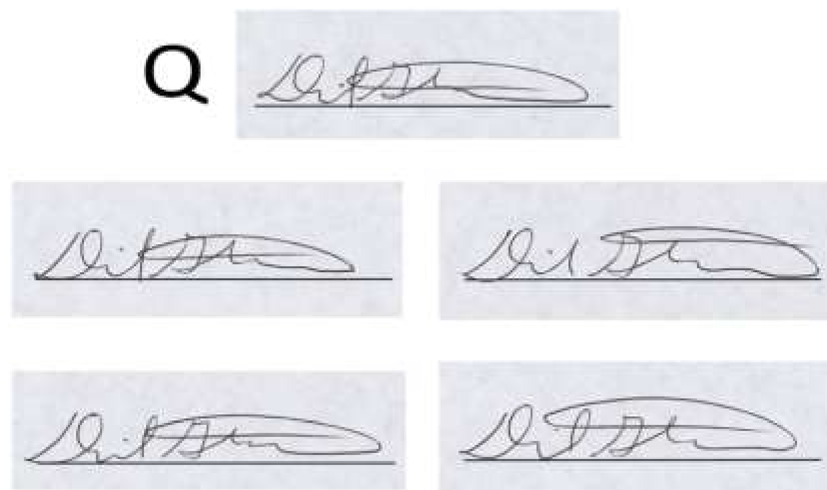
- Clearly, there are other scenarios
  - Digital evidence (collecting evidence)
  - Bloodstain pattern analysis (causal mechanism)
  - DNA mixture analysis (inclusion of suspect)



Known shoe

Crime scene impression



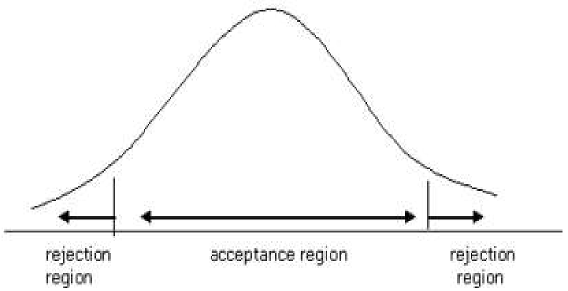


UCI Department of Statistics
Donald Bren School of Information & Computer Sciences

csafe
Center for Statistics and Applications in Forensic Evidence

In what form is the uncertainty communicated?

- Approaches
  - Expert assessment based on experience, training, use of accepted methods. Typically summarized by a categorical conclusion (e.g., identification / exclusion / inconclusive)
  - Two-stage procedure (see, e.g., Parker and Holford in the 1960s)
    - similarity (can the Q and K be distinguished)
    - discrimination (is the observed agreement a coincidence)



  - Likelihood ratio (or the closely related Bayes factor)

Forensic Evidence as Expert Opinion

- Status quo in pattern disciplines (fingerprints, shoe prints, firearms, toolmarks, questioned documents, etc.)
- Examiner analyzes evidence based on
  - Experience
  - Training
  - Use of accepted methods in the field
- Assessment of the evidence reflects examiner's expert opinion
- Conclusions typically reported as categorical conclusions
  - Identification, Exclusion, Inconclusive
  - Multi-category scales (e.g., questioned documents)
  - Potentially via OSAC-developed interpretation scales

Forensic Evidence as Expert Opinion

- Strengths and Weaknesses:
  - Conclusions can be easily understood
  - Black-box studies can be used to provide discipline-level performance data
    - Measure reliability (reproducibility/repeatability) and accuracy
    - For example: Ulery et al. (2011) latent print study found:

Nonmated pairs:	0.15% ID	11.14% Inconcl	88.71% Excl
Mated pairs:	61.37% ID	31.09% Inconcl	7.54% Excl

    - But these studies have limitations
      - Does not address individual case/expert
      - Imc; with “inconclusive” results
      - Studies vs real casework
  - Existing scales don’t address uncertainty (other than through “inconclusive”)
  - It is proving challenging to develop scales that integrate uncertainty assessment
    - Some support ? Strong support?

The Two-Stage Approach

- Stage 1 - Similarity
  - Statistical test or procedure to determine if the two samples "are indistinguishable", "can't be distinguished", "match", etc.
- Stage 2 - Discrimination
  - Assessment of the probability that two samples from different sources would be found indistinguishable
- Used in assessment of trace evidence (like glass)
- Conceptually many other disciplines appear to act in this way (e.g., a footwear examination)

The Two-Stage Approach

- Strengths and Weaknesses
  - Stage 1 is a natural thing to do for discrete / categorical variables (blood type, DNA alleles)
  - Stage 1 is more challenging when the evidence are summarized by quantitative measurements (e.g., element concentrations for glass)
    - Requires a statistical procedure of some sort (e.g., ASTM E2927 for glass)
    - The usual null hypothesis (samples can't be distinguished) seems to be the wrong starting point
    - A binary decision here (distinguished / not) can involve a loss of information
  - Stage 2 is difficult (what is the relevant population?)
  - Stage 2 is not usually provided in a quantitative way

The likelihood ratio (LR)

- A current focus of much attention in forensic science research is the likelihood ratio
- The LR is a statistical concept seen as a potential unifying logic for evaluation and interpretation of forensic evidence
- The LR already plays a role outside forensics in …
  - Statistical inference (hypothesis tests)
  - Evaluating evidence provided by medical diagnostic tests
- Europe has moved in this direction (ENFSI Guidelines and work of NFI)

The likelihood ratio (LR)

- E = evidence
- H_s = "same source" proposition (two samples have the same source)
  H_d = "different source" proposition (two samples have different sources)
- Bayes' Theorem

$$\frac{\Pr(H_s \mid E)}{\Pr(H_d \mid E)} = \frac{\Pr(E \mid H_s)}{\Pr(E \mid H_d)} \; \frac{\Pr(H_s)}{\Pr(H_d)}$$

"a posteriori" odds
in favor of same
source hypothesis

Likelihood ratio or
Bayes factor

"a priori" odds
in favor of same
source hypothesis

- Details: role of task-relevant contextual information, terminology (LR vs Bayes factor)

Likelihood ratio (LR)

- Current state
  - Successfully used for “single source” DNA
    - Underlying biology is understood
    - Biological theory provides a probability model
    - Data is available
    - Note that DNA mixtures remain challenging
  - Examples in other disciplines
    - Glass (Aitken and Lucy)
    - Bullet lead (Carriquiry, Daniels, Stern)
  - Pattern evidence has proven challenging
    - How to represent the evidence as quantitative data
    - Score-based approaches are often used
      (replace evidence E by score S)

Likelihood ratio (LR)

- Strengths and Weaknesses
  - Explicitly compares two (or more) relevant hypotheses/propositions
  - Provides a mapping from a specified set of assumptions to a quantitative summary of the evidence
    - Assumptions regarding probability distributions, manufacturing, transfer of evidence, etc.
    - Making such assumptions explicit has the potential to enhance the transparency of the evidence assessment process
    - But LR can be quite sensitive to the assumptions (Lund and Iyer, 2017)
  - Avoids arbitrary match/non-match decisions when faced with continuous data
  - Can potentially accommodate a wide range of factors (e.g., manufacturing, distribution, wear)
  - Very difficult to develop models for pattern evidence; score-based models have promise but also limitations
  - Challenging for people (especially non-quantitative people) to understand and interpret

Putting ideas together – LR & Expert Opinion

- Black box studies provide field-level data about error rates
- Can think about evidence E as being the expert opinion (not the prints, but the expert's opinion about the prints)
- LR would then tell us to find Pr(E | known match) and Pr(E | known non-match)
- From Ulery et al.
  - If E = ”ident”, then *LR* = (3663/5969) / (6/4083) = 418 in favor of same source
  - If E = ”exclude”, then LR = .085 in favor of same source or LR = 1/.085 = 12 in favor of different source
  - If E = “inconclusive”, then LR = 2.8 in favor of same source
- From the recent Monson et al. firearms (bullet) data
  - If E = ”ident”, then LR = 109 in favor of same source
  - If E = ”elimination”, then LR = .086 in favor of same source or LR = 1/.086 = 12 in favor of different source
  - If E = ”inconclusive-A”, then LR = 1 (not informative)
  - If E = ”inconclusive-B”, then LR = 3 in favor of different source
  - If E = ”inconclusive-C”, then LR = 10 in favor of different source

Putting ideas together – LR & Two-Stage

- Stage 1 of two-stage approach determines whether two evidence samples (e.g., glass) are ”indistinguishable”
- Can think about evidence E being ”observation that samples are indistinguishable”
- LR would then tell us to evaluate Pr(E | same source) and Pr(E | different source)
- Pr(E | same source) is usually very high (depends on statistical procedure used to determine whether we can distinguish), typically .95 or higher
- Stage 2 is our attempt to calculate Pr(E | different source)
- Stage 2 is key to understanding the value of the evidence

Conclusions

- Any approach to assessing the probative value of forensic evidence should:
  - Account for the two (or more) competing hypotheses about how the evidence (data) were generated
  - Be explicit about the reasoning and assumptions on which the assessment is based
  - Have relevant empirical support for the reasoning and assumptions
  - **Include an assessment of the level of uncertainty associated with the assessment**
- The language used in reports, testimony, opening/closing statements are critical.
- Contact: sternh@uci.edu

Communicating Forensic Findings:

Framing ~~the~~ Some Issues

Steve Lund

Evidential Statistics Focus Area Lead

Statistical Engineering Division

Acknowledgement

- Collaboration with colleagues Hari Iyer and Will Guthrie

Disclaimer

- Viewpoints expressed are our own and are not intended to reflect those of anyone else at NIST

Communicating Forensic Findings (CFF)

Dictionary

Definitions from Oxford Languages · Learn more

com·mu·ni·ca·tion

/kəˌmyo͞onəˈkāSH(ə)n/

noun

1. the imparting or exchanging of information or news.
   "at the moment I am **in communication with** London"

CFF: Experts imparting information to other parties in the judicial system.

What type of information?

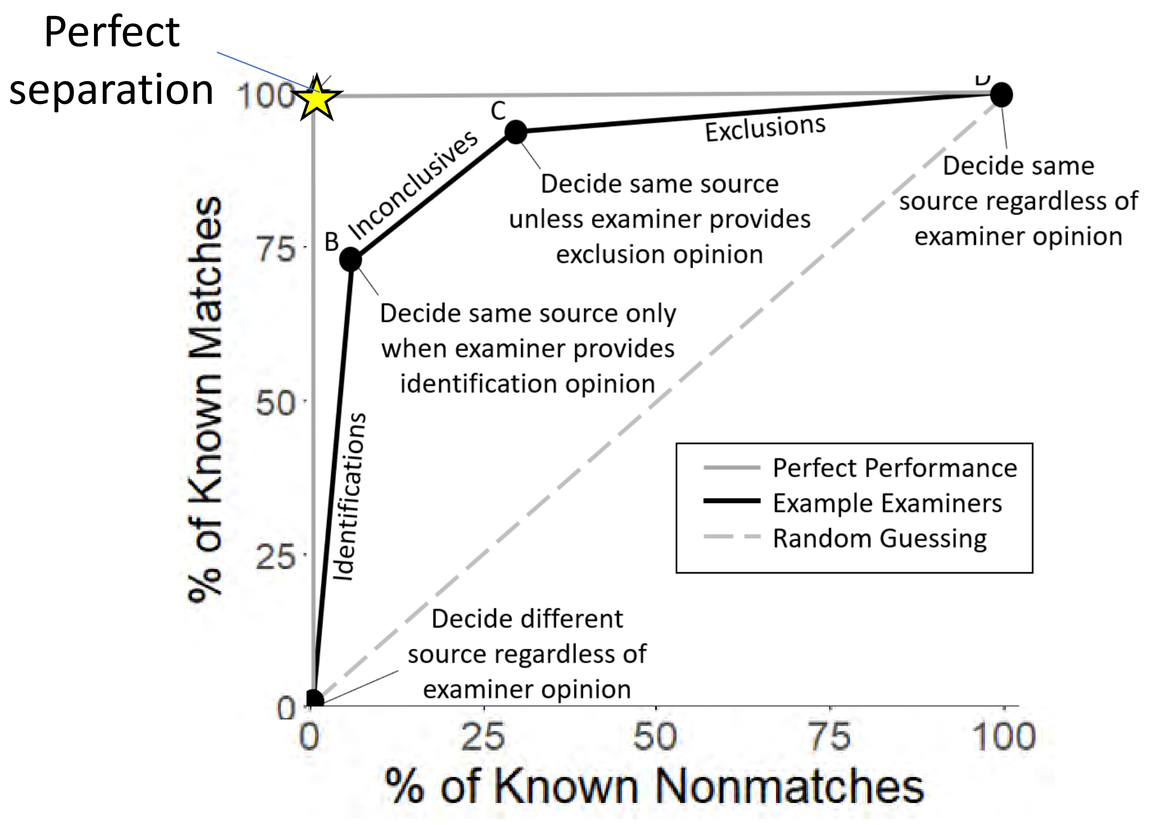
- **Observations about the evidence**
  - Descriptive, demonstrable
  - Often high-dimensional or complicated

- **Opinions of the expert(s)**
  - Interpretive, personal, some variability expected – “range of opinions”
  - Typically simpler than observations

CFF Goal: Help others make better decisions

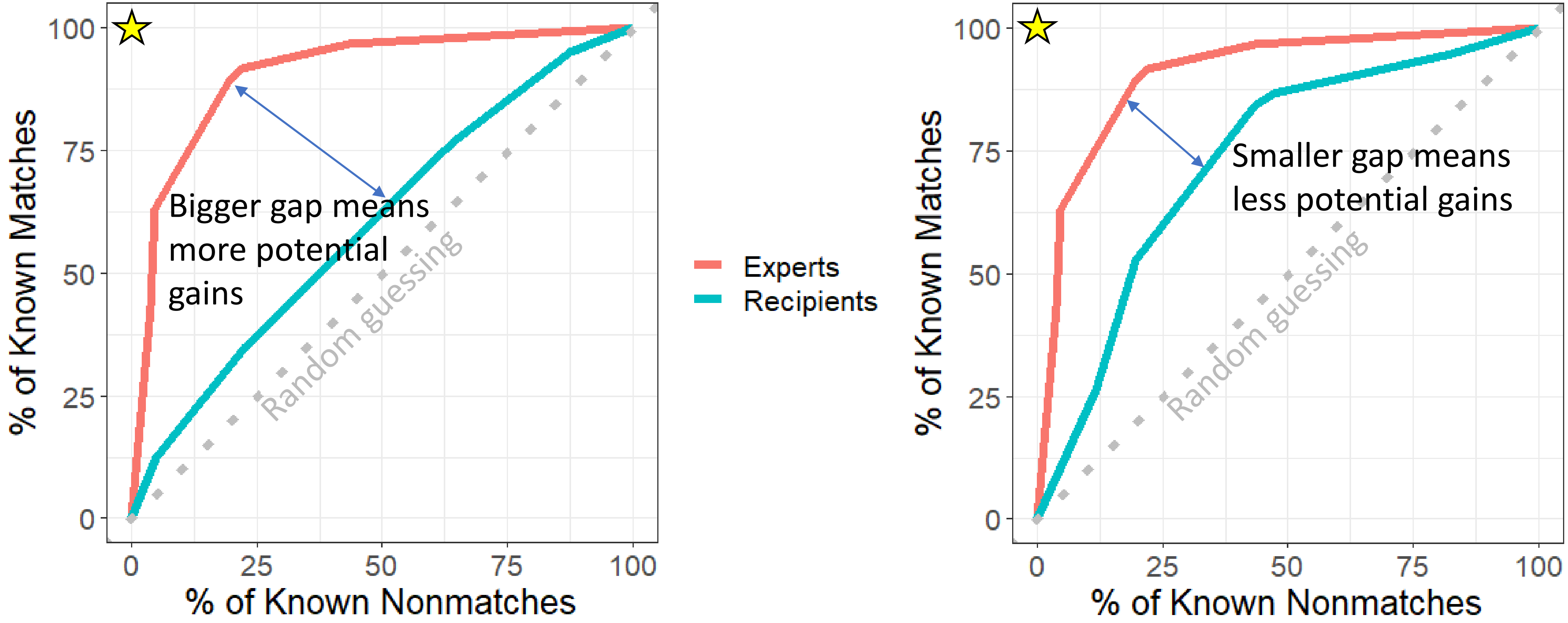
How to Measure/Grade CFF?

- What's the greatest potential gain from the expert?
  - Among ground-truth-known tests, could compare experts' ability to distinguish between propositions of interest with recipients' to measure the gap



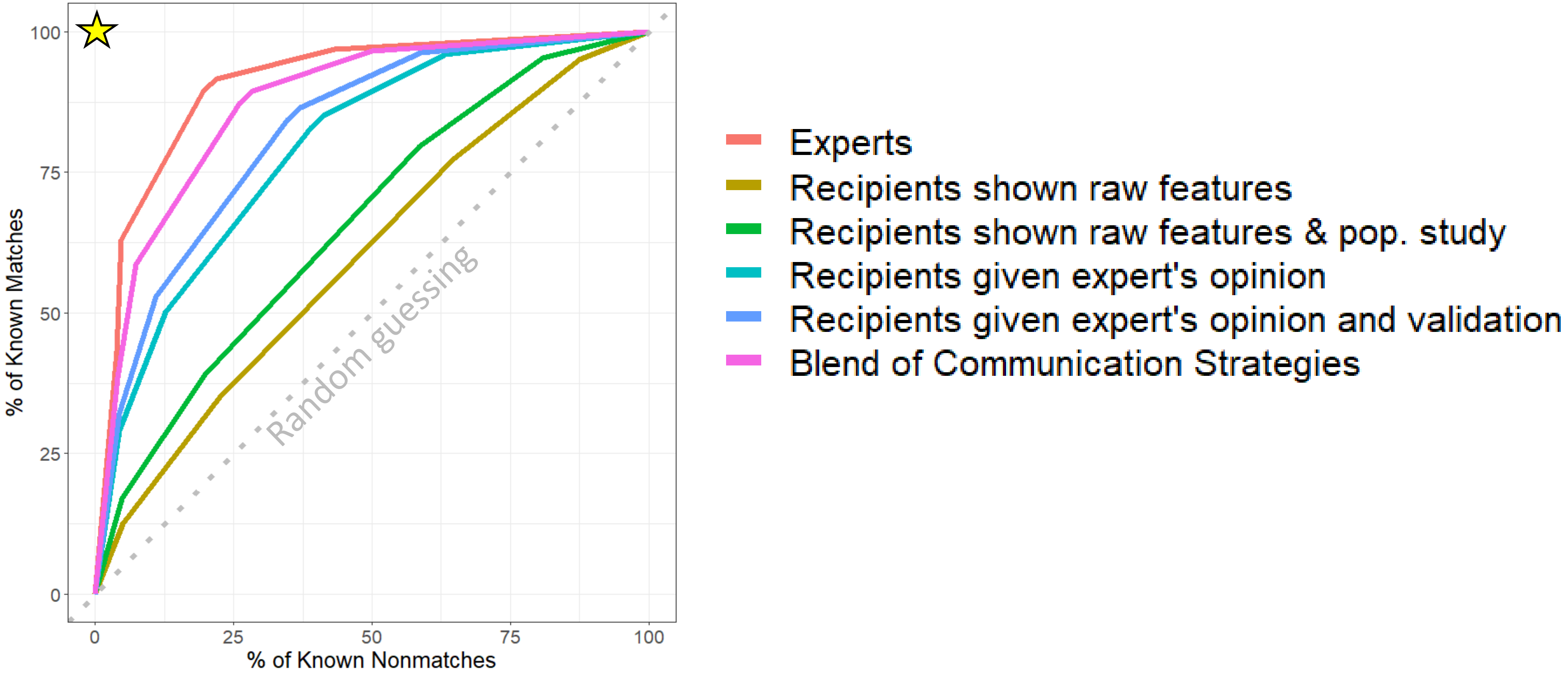
How to Measure/Grade CFF?

- What's the greatest potential gain from the expert?
  - Among ground-truth-known tests, could compare experts' ability to distinguish between propositions of interest with recipients' to measure the gap
    - No gap would imply no meaningful information to communicate. Typically expect a gap



See how well CFF approaches close the gap

- Consider multiple approaches:
  - Presenting observations vs presenting opinions
  - Accompanying supporting information (e.g., population study summaries, theoretical explanations)
  - Attempts to educate decision makers vs. attempts to instruct the decision makers



How to Measure/Grade CFF?

- This approach may encourage suggesting recipients adopt expert’s sentiment as their own (since then they’d have the same discrimination power as the expert)
  - Ignores range of opinions / treats personal and subjective interpretation as communal fact
  - What to do with disagreements among equally competent experts?
  - What about uncharted territory?
- Blindly accepting an expert’s opinion opens a doorway for junk science or pushing boundaries too far (extrapolation)
- Focusing on validation data could help close the door
  - Recognize overconfidence or unsupported claims
- Reliable communication is critical, including validation details

Important Caveat

- Judicial outcomes relying on forensic science provide less observable feedback than real world outcomes relying on other applications of science. E.g.,
  - Building remains standing or collapses (e.g., Champlain Towers South)
  - Side effects of drug released for public consumption (e.g., Vioxx with ~30,000 adverse cardiac events)
  - Most forensic casework applications are like rockets disappearing immediately after launch
- More difficult to recognize real world successes and failures for forensic applications
  - Allowed overconfident performance conjecture unsupported by empirical testing (e.g., to the exclusion of all other sources, error-free method, etc.)
  - Prior to DNA, no obvious signs of trouble means these claims largely avoided scientific scrutiny
  - Following public errors and work of the Innocence Project, legal and scientific communities increase demand for empirical studies

Validation…

- … is even more important to assessing reliability of forensic methods than it is for most applied sciences
- … has a critical role in…
  - Labs deciding whether to use a method in a particular case
  - Recipients deciding how much weight to give a method’s result in a particular case
    - High-stake decisions made by peers rather than specialists
- … is an important component in CFF

So how do we talk about validation?

- “Validated”

- “Error rate”

(Google Gemini result for “generate an image for the word unsatisfactory”)

“Validated”

- Falsely implies there’s a checklist that, once completed, renders uncertainty regarding method performance inconsequential
  - “How many samples do I need?”
  - Overlooks benefit to collecting additional validation data
- Suggests performance is one-size-fits-all
- Masks subjectivity of chosen validation criteria as consequence of statistics and science, making it harder to question

Error rates

- Biggest Positive: Brings attention to empirical performance studies
- Biggest Drawback: Requires oversimplifying to label each opinion/conclusion as either correct or incorrect
  - Most opinion/conclusion scales are on a more refined spectrum
  - Throws away relevant information
  - Leads to many proposals for handling inconclusive conclusions, some of which can be misleading


ELSEVIER

Forensic Science International: Synergy

Volume 8, 2024, 100472

FSI SYNERGY

Inconclusive decisions and error rates in forensic science

H. Swofford, S. Lund, H. Iyer, J. Butler, J. Soons, R. Thompson, V. Desiderio, J.P. Jones, R. Ramotowski

Example

Pauw-Vugts, P., Walters, A., Øren, L., & Pfoser, L. (2013). FAID2009: proficiency test and workshop. *AFTE Journal*, *45*(2).

Pauw-Vugts et al. -- FAID2009 125

Test Set	Castings of Bullets or Cartridge Cases	Number of times a conclusion was given						One or two firearms
		A	B	C	D	E	Z	
A	cartridge cases	63	1	0	0	0	0	1
B	bullets	1	1	0	7	58	0	2
C	cartridge cases	35	14	7	0	7	1	1
D	bullets	4	0	19	23	14	3	2
E	cartridge cases	0	1	2	14	46	1	2
F	bullets	61 (1*)	1	0	1	0	0	1
G	cartridge cases	14	0	9	17	22	0	2
H	bullets	3 (4*)	15 (2*)	34	1	2	3	1
J	cartridge cases	1	1	2	9	51	0	2
K	bullets	13 (1*)	16 (4*)	17	8	4	1	1

Known Matches

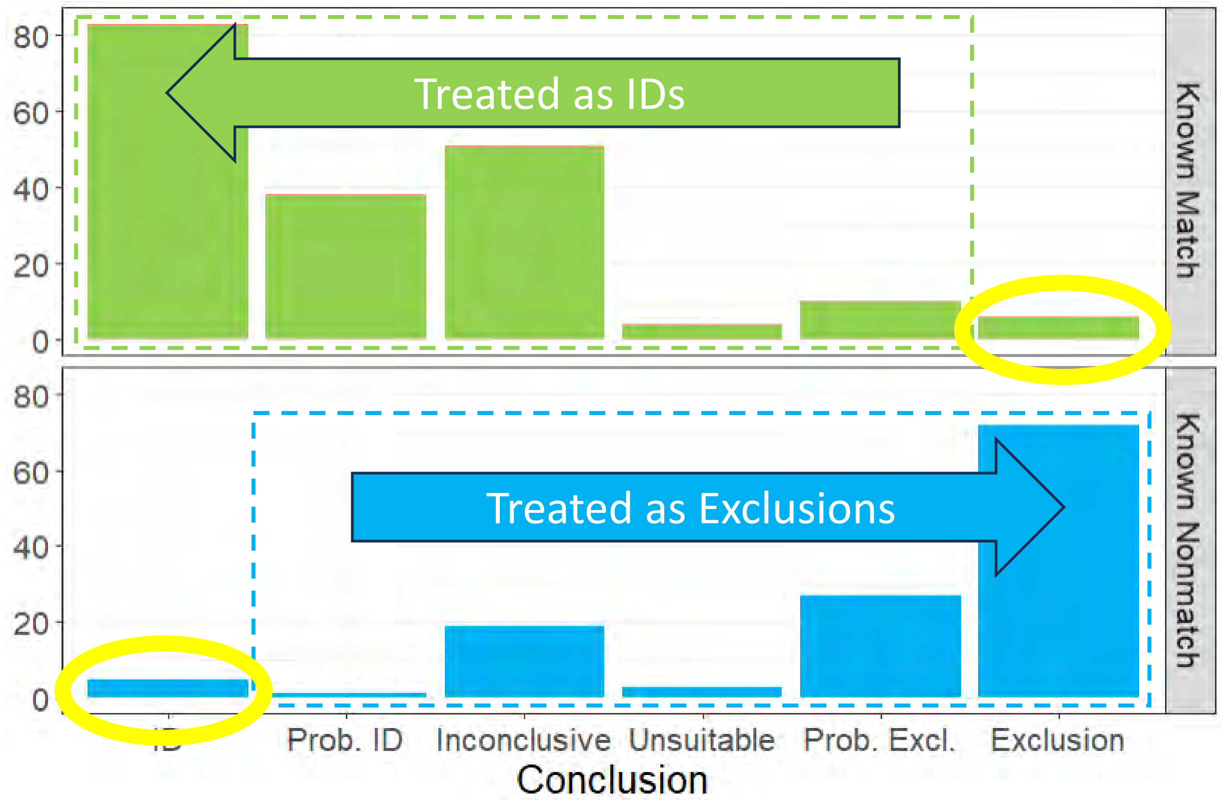
Known Nonmatches

	Identification	Probable ID	Inconclusive	Probable Ex	Exclusion	Unsuitable
Known Matches	83	38	51	10	6	4
Known Nonmatches	5	1	19	27	72	3

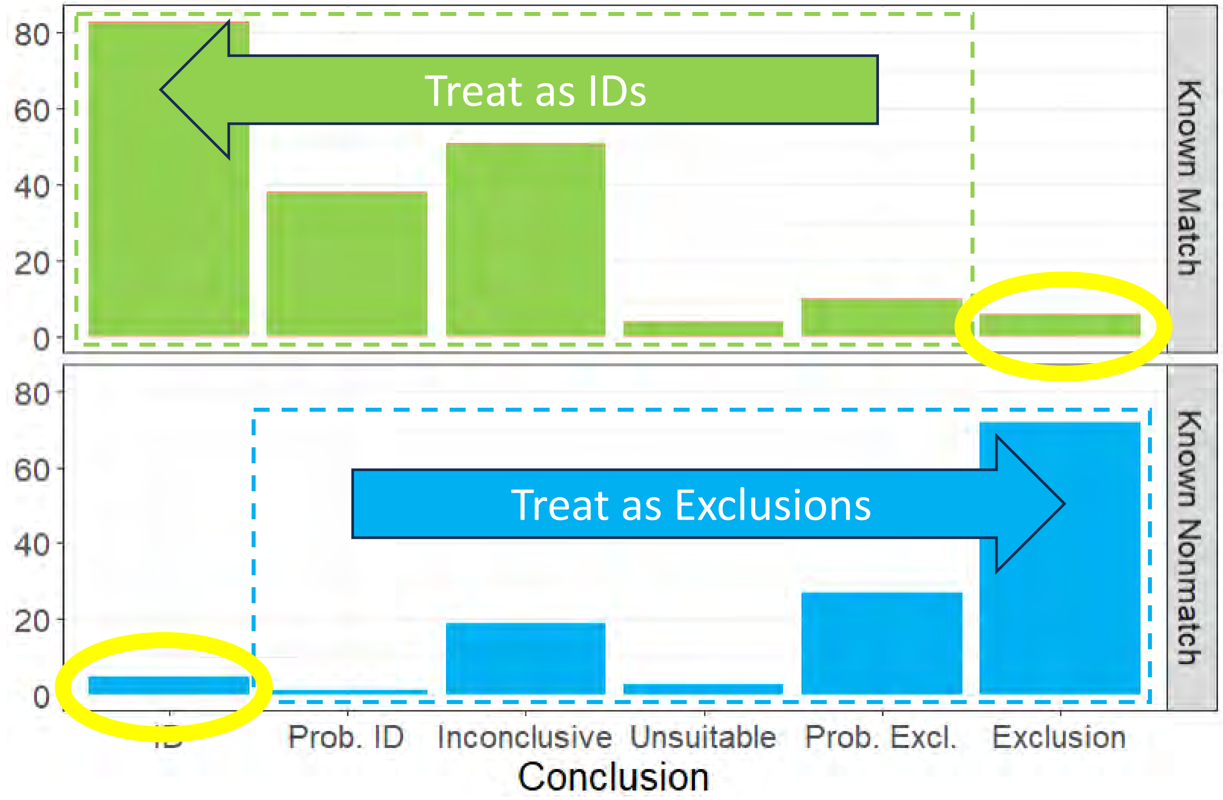
Example

Pauw-Vugts, P., Walters, A., Øren, L., & Pfoser, L. (2013). FAID2009: proficiency test and workshop. *AFTE Journal*, *45*(2).

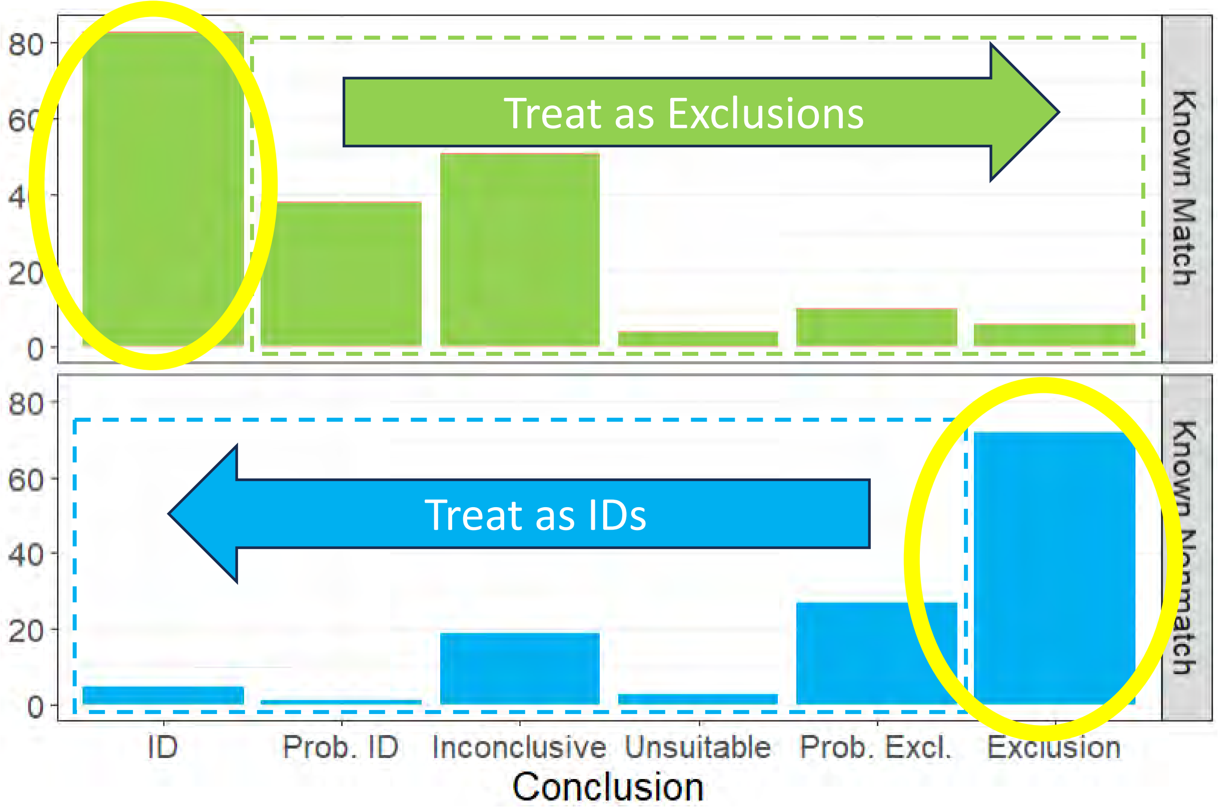
Quote Part 1: “Scientifically, an inconclusive result has to be automatically incorrect: a comparison is either from a same-source or a different-source. AFTE rules allow inconclusives to be counted as both identifications and eliminations, and therefore artificially decrease error rates.”



Quote Part 2: "If we focus on a correct source decisions only, the percentage of correct decisions can be as low as 49%, leaving at least 51% of the decisions as errors (correct source identification rate taken from bullet comparisons in Pauw-Vugts et al. (2013))."

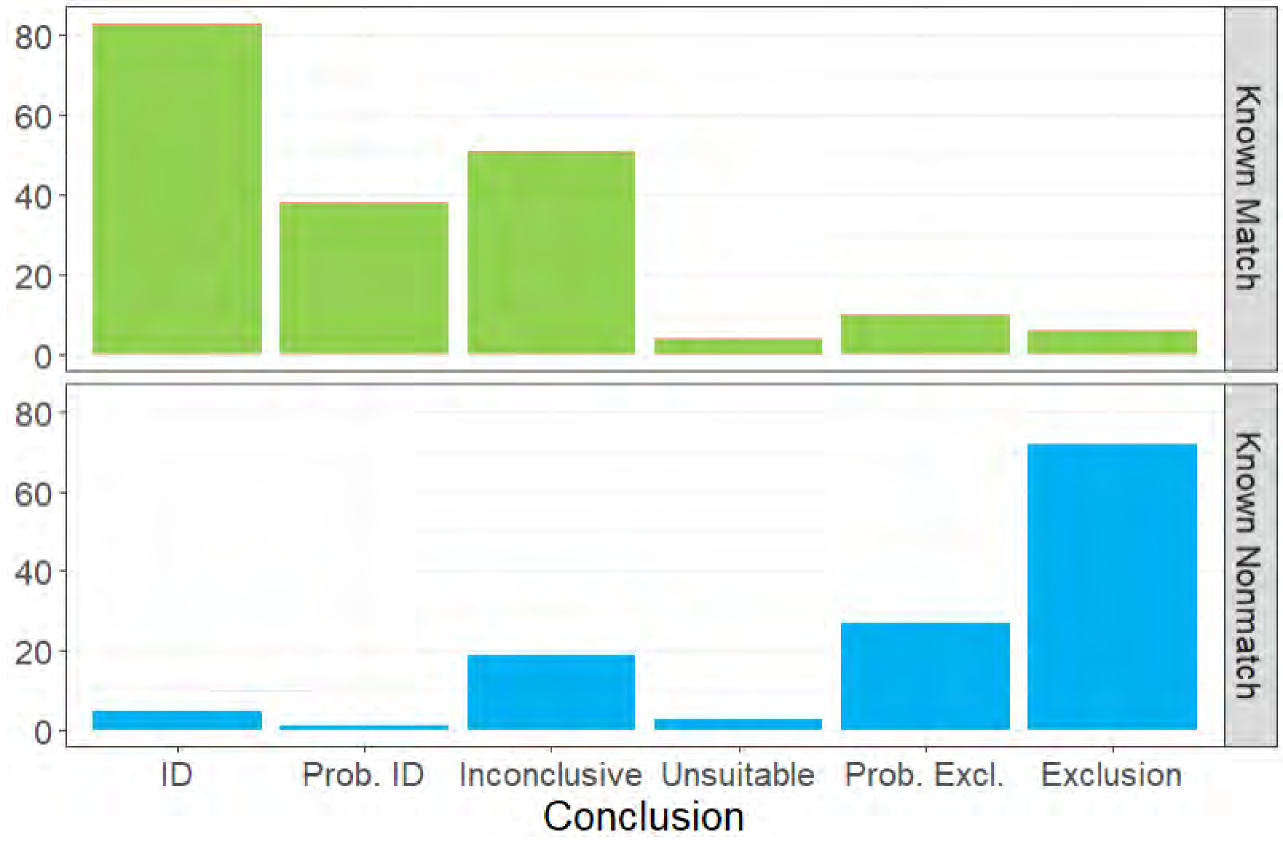


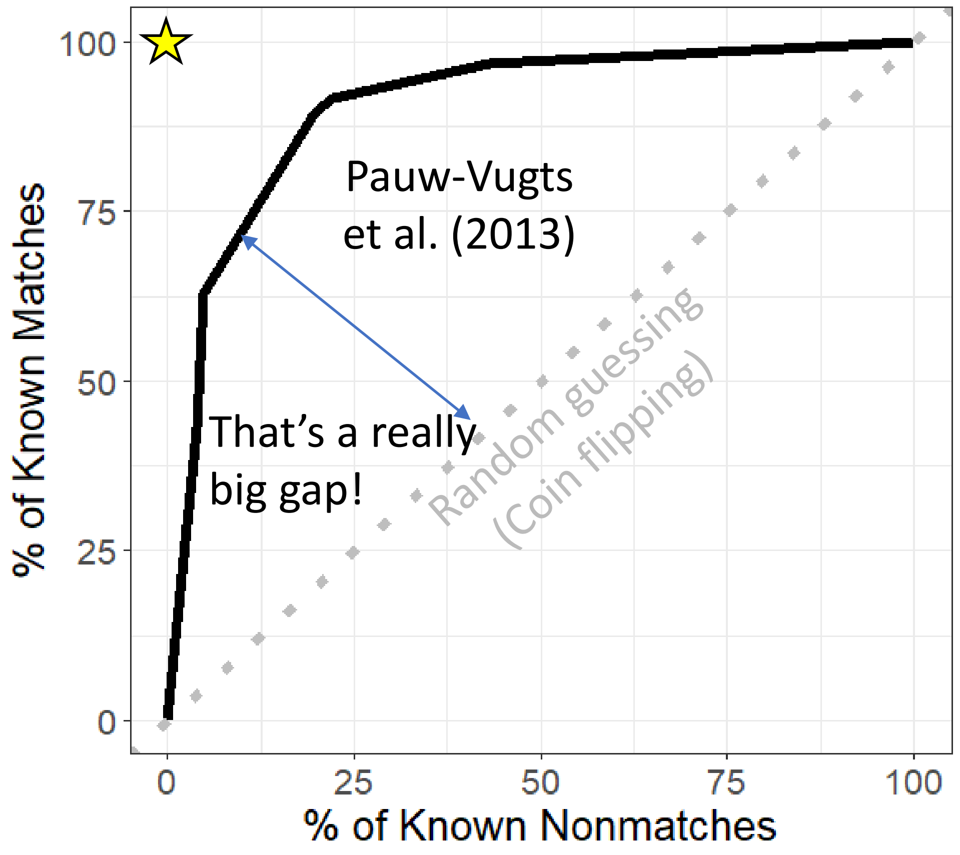
AFTE Treatment (Common)



Suggestion from Authors (Statisticians)

Full quote: “Scientifically, an inconclusive result has to be automatically incorrect: a comparison is either from a same-source or a different-source. AFTE rules allow inconclusives to be counted as both identifications and eliminations, and therefore artificially decrease error rates. If we focus on a correct source decisions only, the percentage of correct decisions can be as low as 49%, leaving at least 51% of the decisions as errors (correct source identification rate taken from bullet comparisons in Pauw-Vugts et al. (2013)). **This is statistically worse than random chance - that is, examiners would perform about as well if they were flipping a coin to make the decision!**”





Credit: https://craftbits.com/project/diy-collage-of-pages-bookcase/

Validation Nuances

- Attempt to assign weight to an opinion in a particular case
- Efficacy expected to vary across case types
  - E.g., expect mostly IDs and Exclusions when comparing two exemplars, expect mostly inconclusives for very low-quality questioned impressions
- Some factors describing case type may allow us to predict changes in examiner performance
  - What are these factors? What are their effects?

- Available data is not ideal
  - Fewer tests than we'd like (cost-benefit analysis)
  - Few, if any, tests match circumstances of current case (e.g., different quality sample(s), different lab or expert, awareness of being tested, etc.)
  - Departures from ideal statistical sampling approaches: volunteer participants, convenience sample materials, not all tests are answered
  - Important details that changes or adds uncertainty to the meaning of the data
- Despite limitations, available data can be (are) informative
  - E.g., demonstrate that some experts perform well in some scenarios (i.e., not coin-tossers)
  - How informative will depend on subjective reactions to limitations

- How to reasonably summarize or present available validation information?

Key Points

- Validation testing remains the primary means by which society can understand the efficacy of forensic science methods (more so than many other areas of science)

- Forensic science relies more on general population (e.g., fact finders) to carry out its mission than do other scientific applications
  - Don't take 12 random people to approve space shuttle launch or decide whether open heart surgery will be performed

- We need to improve how we communicate about validation
  - "Error rates" and "validated" oversimplify in potentially misleading ways

- Looking forward to hearing your thoughts and perspectives on these, and other, important CFF topics!

It's not you, it's me

Communicating Results in Forensic Reports and Testimony

Julie Burrill, PhD

Alan Alda Center for Communicating Science

Leverhulme Research Centre for Forensic Science

NIST Workshop: Communicating Forensic Findings

June 25th, 2024

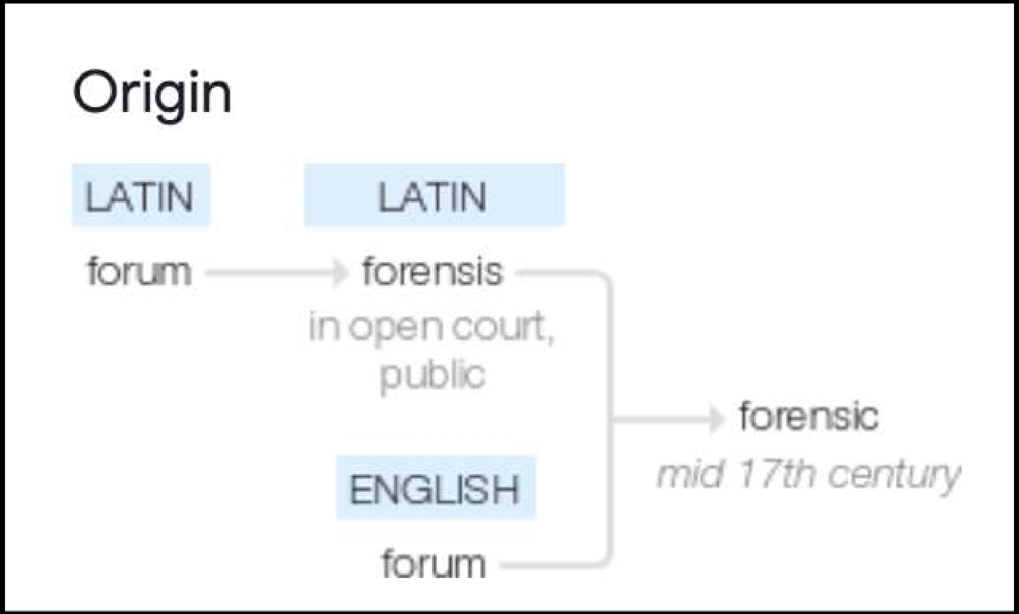
TLDR:

Science communication research explains WHY this work is so hard.

You need to set your audience up for success.

Let’s train scientists in the actual skills required to do that.

Why does this matter?



1) Relating to or denoting the application of scientific methods and techniques to investigation of crime.

2) Relating to courts of law.

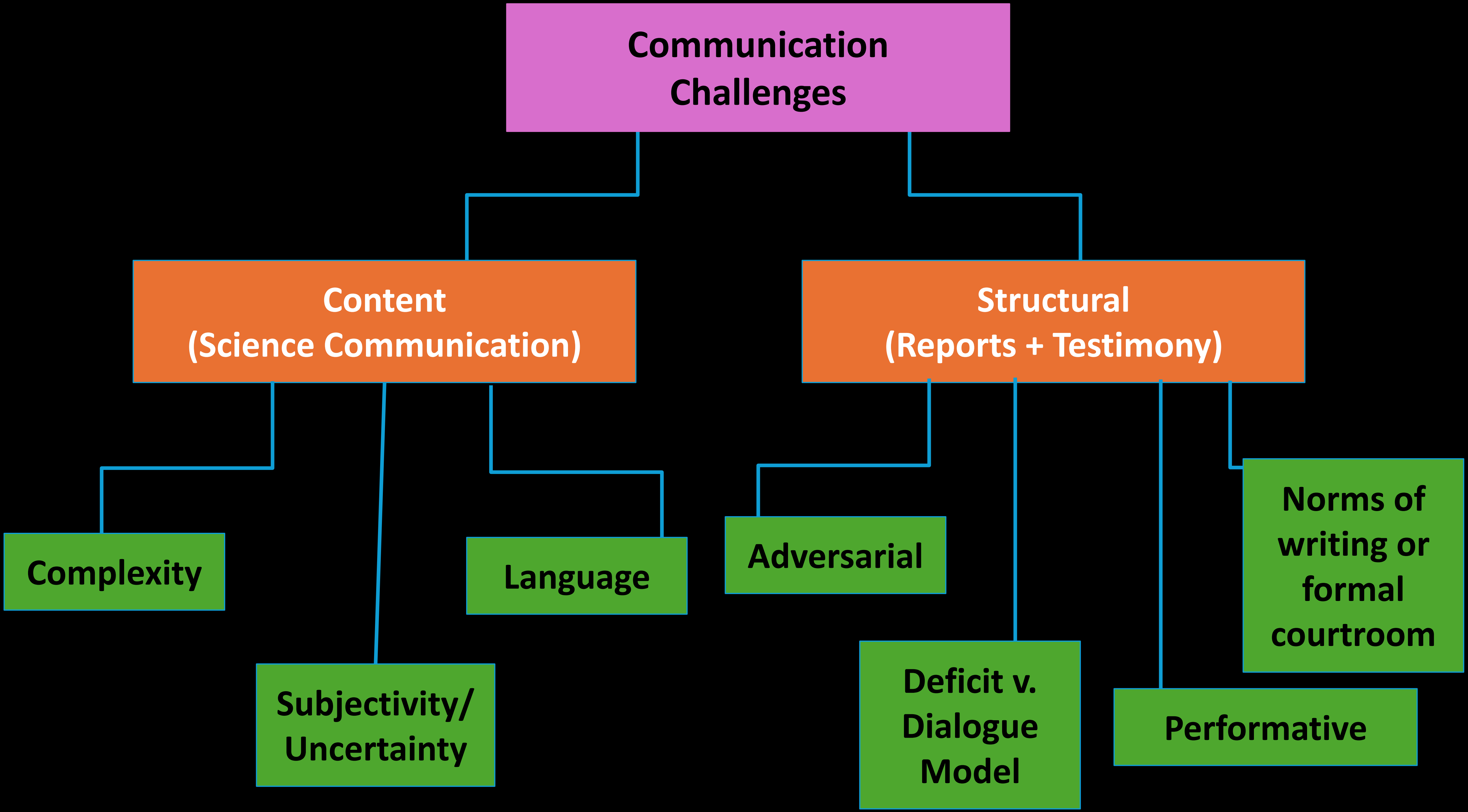
Mid 17th c. from Latin forensic “in open court, public,” from forum

Forensics exists to be communicated.

Now is our chance to get it right!

"...amidst the furor over the most "correct" or "accurate" way to present evidence, the perspective of the fact-finder is often lost. **Without comprehension, correctness is moot.**"

Eldridge, 2019, FSI: Synergy, Vol. 1, p 24-34.


Communication Challenges
Content (Science Communication)
Structural (Reports + Testimony)
Complexity
Subjectivity/ Uncertainty
Language
Adversarial
Deficit v. Dialogue Model
Performative
Norms of writing or formal courtroom

Norms of Scientific Writing

Following this PCR procedure, 1ul of fluorescently-tagged, amplified DNA product was added into the prepared CE mixture tube.

Following this PCR procedure, 1ul of fluorescently-tagged, amplified DNA product was added into the prepared CE mixture tube.

Passive voice

Transition phrase

***Following this PCR procedure,* 1ul of fluorescently-tagged, amplified DNA product *was added* into the prepared CE mixture tube.**

Passive voice

Transition phrase

Adjective chain

Following this PCR procedure, 1ul of fluorescently-tagged, amplified DNA product was added into the prepared CE mixture tube.

Passive voice

Transition phrase

Adjective chain

Following this PCR procedure, 1ul of fluorescently-tagged, amplified DNA product was added into the prepared CE mixture tube.

Noun stacking

Passive voice

Transition phrase

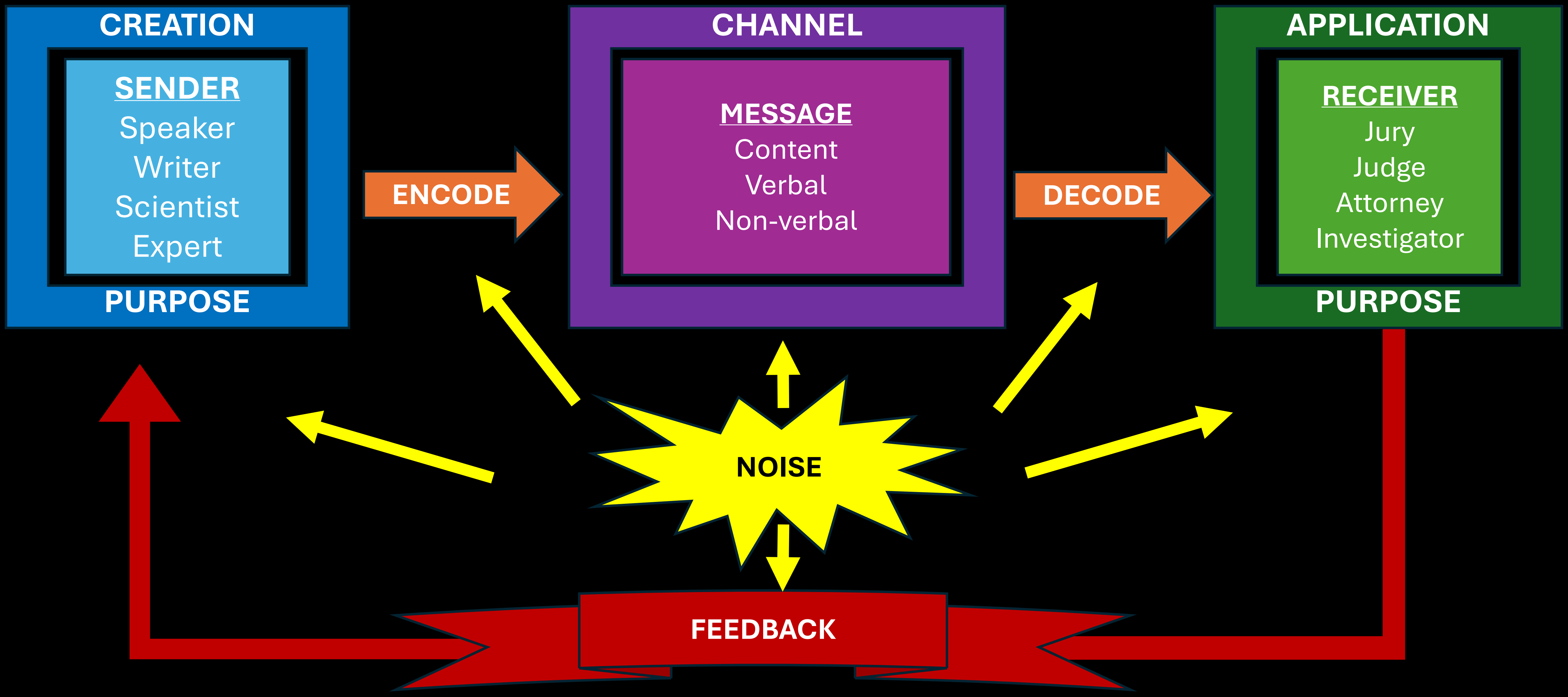
Adjective chain

Following this PCR procedure, 1ul of fluorescently-tagged, amplified DNA product was added into the prepared CE mixture tube.

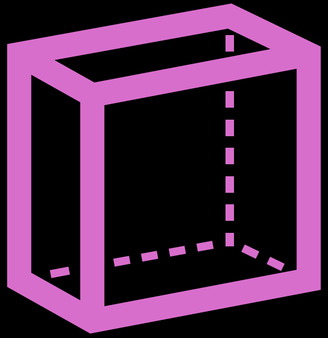
Noun stacking

Passive voice

Communication as a 2-way process



Research-driven Communication Theories

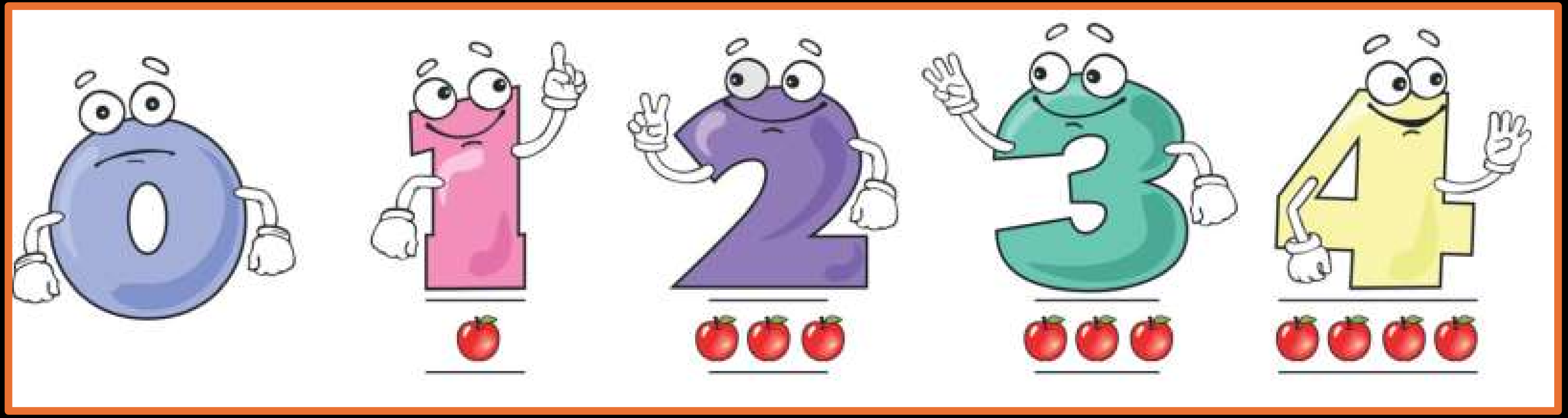


Framing Theory

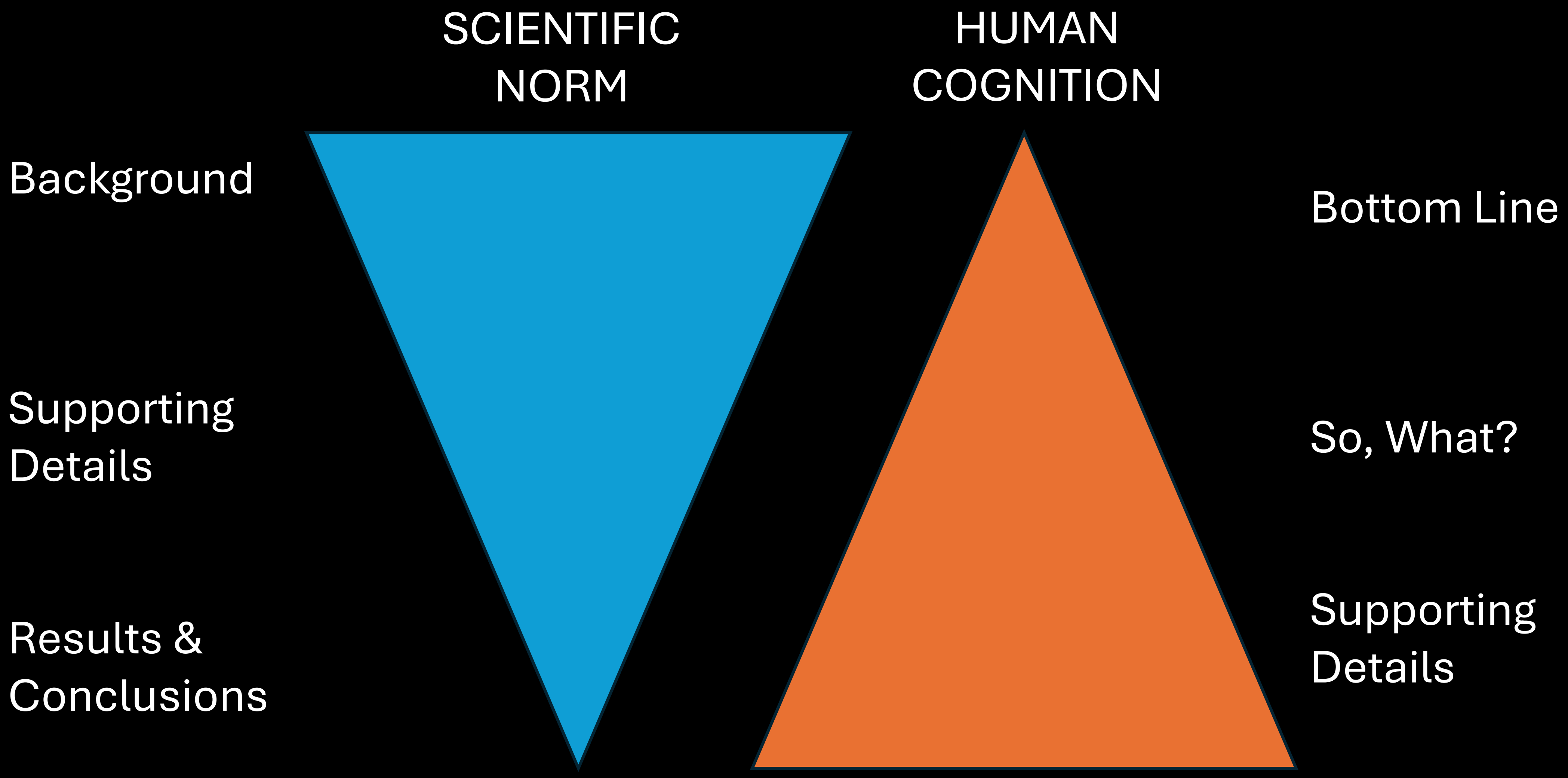
Narrative validity

Primacy, Recency,
Repetition Biases

Affect heuristic for risk assessment



Ordering Information



Use the world of Science Communication research!

- Frame of Reference/ Analogy shifting
- Active listening → Adaptability
- Constructive Empathy
- Building a narrative
- Language Recognition

Now is our chance to IMRPOVe FS communication!

Use the world of Science Communication research!

- Frame of Reference/ Analogy shifting
- Active listening → Adaptability
- Constructive Empathy
- Building a narrative
- Language Recognition

Thank You!

Alan Alda Center for Communicating Science

Leverhulme Centre for Forensic Science Research

julie.burrill@stonybrook.edu

julie.burrill@gmail.com

References

- Edmond, G., Towler, A., Growns, B., Ribeiro, G., Found, B., White, D., ... & Martire, K. (2017). Thinking forensics: Cognitive science for forensic practitioners. *Science & Justice*, *57*(2), 144-154.
- Jackson, G., Kaye, D. H., Neumann, C., Ranadive, A., & Reyna, V. F. (2015). Communicating the results of forensic science examinations.
- Roland, M. C. (2009). Quality and integrity in scientific writing: prerequisites for quality in science communication. *Journal of Science Communication*, *8*(2), A04.
- Steijaert, M. J., Schaap, G., & Riet, J. V. T. (2021). Two-sided science: Communicating scientific uncertainty increases trust in scientists and donation intention by decreasing attribution of communicator bias. *Communications*, *46*(2), 297-316.
- Davis, P. R., & Russ, R. S. (2015). Dynamic framing in the communication of scientific research: Texts and interactions. *Journal of Research in Science Teaching*, *52*(2), 221-252.
- Winter, R. J., & Greene, E. (2007). Juror decision-making. *Handbook of applied cognition*, *2*, 739-762.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 192–221). Cambridge University Press.
- Kahan, D. M., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of risk research*, *14*(2), 147-174.
- Lange, R. D., Chattoraj, A., Beck, J. M., Yates, J. L., & Haefner, R. M. (2021). A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLOS Computational Biology*, *17*(11), e1009517.
- Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current directions in psychological science*, *28*(3), 247-253.
- Peters, E. (2017). Overcoming innumeracy and the use of heuristics when communicating science. *The Oxford handbook of the science of science communication*, 389-398.
- Igou, E. R., & Bless, H. (2003). Inferring the importance of arguments: Order effects and conversational rules. *Journal of Experimental Social Psychology*, *39*(1), 91-99.
- Po¨ttker, H. (2003). News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies*, *4*(4), 501–511. https://doi.org/10.1080/1461670032000136596
- Somerville, R. C., & Hassol, S. J. (2011). Communicating the science of climate change. *Physics Today*, *64*(10), 48-53.

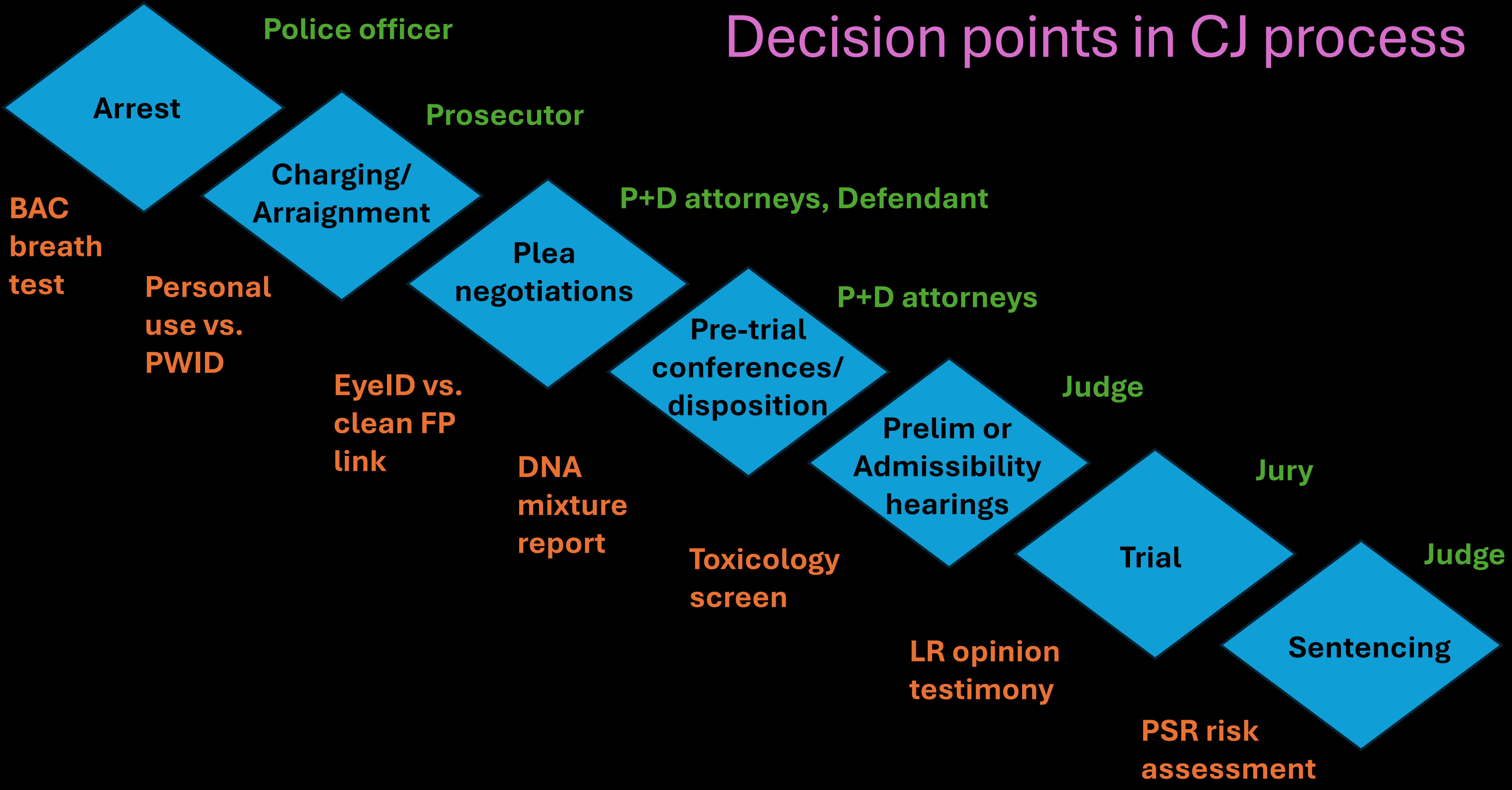
Core Competencies

- Frame of Reference/Analogy shifting
- Active listening → Adaptability
- Constructive Empathy
- Building a narrative
- Language Recognition

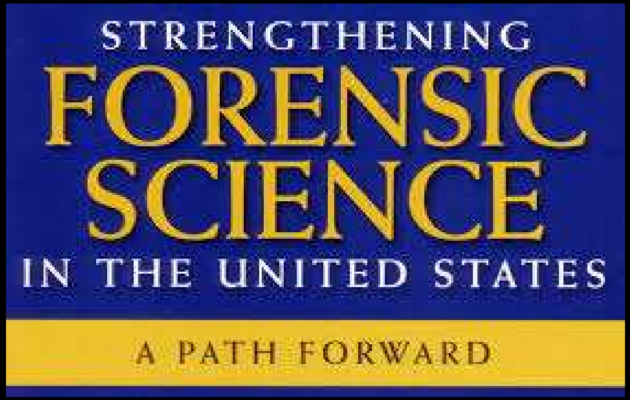
NORMS OF SCIENCE WRITING

- **Passive Voice** - When the action itself is more important than who performed the action, i.e. in methodology sections, the experimental process is the subject of the sentence.
- **Transitional Words v. Topic Sentence** – continually putting steps and inferences in context, indirectly justifying choices by putting basis in dependent clauses.
- **Compound Nouns and Adjective Chains** - combining nouns and adjective chains or multiple nouns to create lengthy multi-part nouns with greater precision and accuracy.

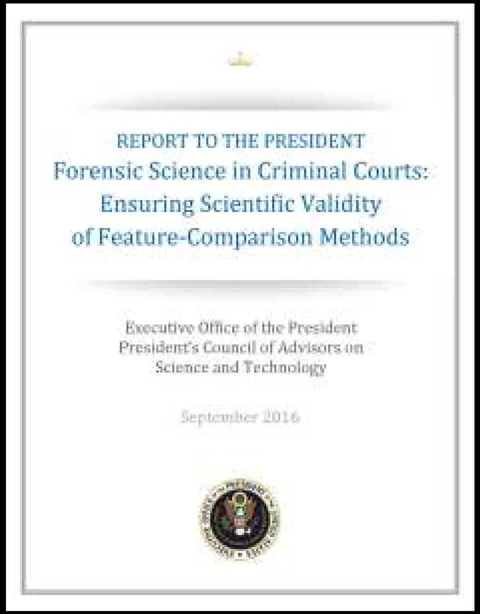
Decision points in CJ process

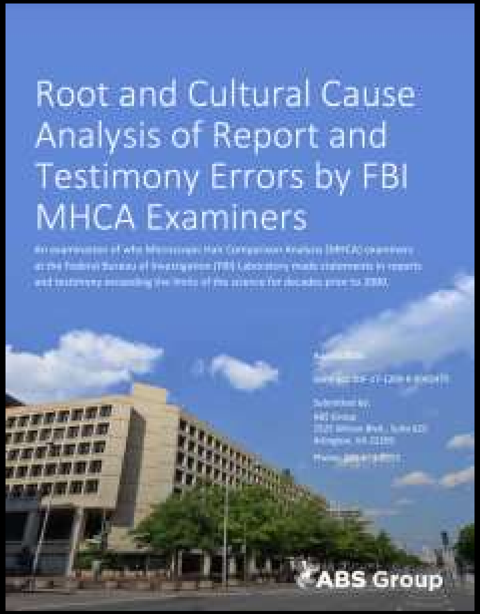


Some people agree!



Organization of
OSAC
Scientific Area Committees
for FORENSIC SCIENCE





Communication as a "core element" in cross-disciplinary continuing forensic education training

Communication as "required component" of forensic Training-to- Competency programs

Lack of standard language for testimony identified as a significant problem

Testimony monitoring and review adopted as corrective measure.

Good communication is becoming even more important.

Cellebrite

MOBILedit Forensic

STRmix. EMPOWERING FORENSIC SCIENCE.

Cybergenetics

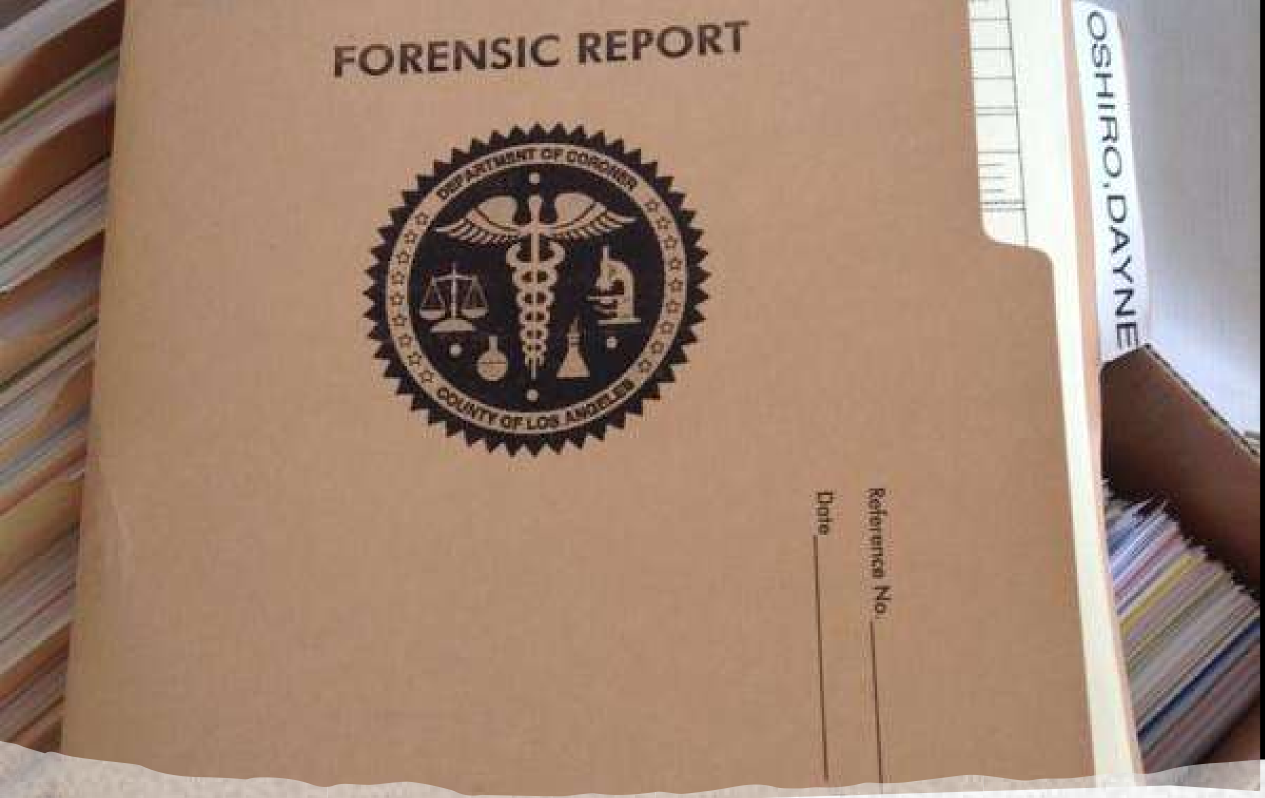
→ Evaluative/LR opinions in more disciplines.

→ Software and AI analytical bases for interpretations.

The Courtroom is a terrible model for communication!

Testimony is the Evidence.

Forensic Science findings acquire meaning in context.



Forensic Reports and Expert Testimony

You keep using that word…I do not think it means what you think it means:

Challenges in Communication Comprehension

Dr. Heidi Eldridge

George Washington University

June 25, 2024

Roadmap

- What do factfinders hear?
- How do they hear it?
- Are they even listening?
- Why can’t we just speak in plain English?
- What about when others misrepresent our words?
- Where do we go from here?

What do factfinders hear?

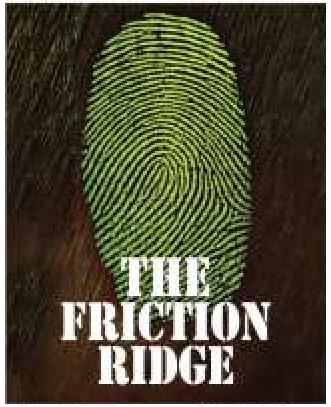
"Our words matter. Language is a powerful weapon. It can be used to inform, but it can also be used to persuade or mislead. We must remember that many of the phrases we use as scientists are a kind of shorthand for larger concepts that other scientists understand. But juries do not have that level of understanding. Juries accept them at face value."

"I am 100% certain of my conclusion." (But should the jury be certain?)

Written by Heidi Eldridge

FROM TIME OUT OF MIND, forensic scientists have testified to results with phrases like "one hundred percent certain," and felt completely comfortable doing so. After all, why would we testify under oath to something that we did not believe to be true? Then, in 2009, the National Academy of Sciences report on forensic science was released, and in the aftermath, forensic scientists began to be cautioned against using this phrase and others like it. Many embraced this change, while others continue to ask: *But why?*

Many arguments have been made addressing the lack of wisdom in using a phrase such as "100% certain". Here is how the most common argument goes: The assertion of one's certainty does not equate to a scientific stance. Nothing in science is ever 100% certain. The cornerstone of scientific exploration is the formation of a con-



bring to the courtroom that training, experience, and knowledge. And they look to us with a faith that, for some, borders on reverence. And because of this faith, we bear a huge burden of responsibility: *Clarity.*

a proven fact that every fingerprint is different."

Similarly, when we say, "I am 100% certain of my conclusion," we might mean that we have conducted a careful examination, reached the best conclusion possible with the data available, and that we would not have reported that conclusion unless we were confident that we had done our work well. But what does the jury hear? They hear, "I'm an expert, and I'm telling you that this conclusion is fact and cannot possibly be wrong."

But the truth of the matter is, sometimes we are wrong. And what we are stating for the jury is not fact; it is opinion. To be clear, the opinion is based on something—it is not just made up out of thin air. But it is still opinion. And to state it in terms that give it the veneer of fact is both over-stating and just plain misleading.

Remember your audience: The jury

Evidence Technology Magazine, March-April 2012

What do factfinders hear?

Forensic Science International: Synergy 4 (2022) 100199

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: www.sciencedirect.com/journal/forensic-science-international-synergy

Describing communication during a forensic investigation using the Pebbles on a Scale metaphor

Laura Carlson [a,*], Jarrah Kennedy [b], Kimberly A. Zeller [c], Thomas Busey [d]

[a] University of Notre Dame, The Graduate School, 110C Bond Hall, Notre Dame, IN, 46556, USA
[b] Kansas City Police Crime Laboratory, 2645 Brooklyn Avenue, Kansas City, MO, 64127, USA
[c] Houston Forensic Science Center, 500 Jefferson Street, 13th Floor, Houston, TX, 77002, USA
[d] Indiana University, Psychological and Brain Sciences, 1101 E. 10th Street, Bloomington, IN, 47405, USA

ARTICLE INFO

Keywords:
Forensic science
Communication
Cognitive psychology
Reporting results

ABSTRACT

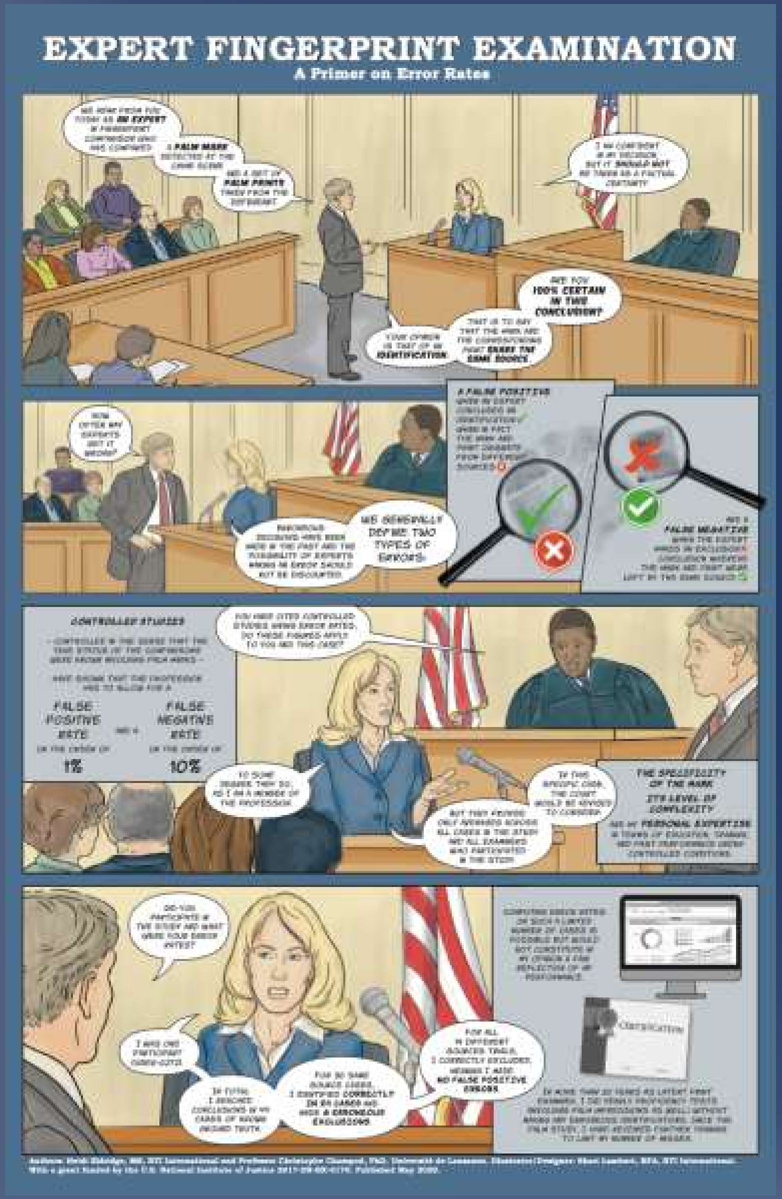
During the investigation of a crime, evidence is collected, analyzed, interpreted, and discussed by various stakeholders. This article examines the communication that may occur between two of these stakeholders: detectives and forensic analysts, and how their interaction influences the interpretation of evidence as the investigation proceeds and the theory of the case evolves. Such communication can be understood as sets of actions that are inter-dependent: for example, a request for a specific analysis by a detective leads to analyses and conclusions that the analyst shares with the detective, which leads to an assessment of these conclusions relative to the theory of the case, which leads to further analysis requests, and so forth. We present the Pebbles on a Scale metaphor, which describes how communication and the understanding of evidence takes place between the detective and analysts, and the different ways in which they consider the information as a function of their roles in the investigation. Using a hypothetical case for illustration, we discuss communicative challenges, the evolving theory of the case, the language that is used by analysts to discuss "yes", "no" and "I don't know" conclusions, and how those conclusions are used by detectives during the progression of the investigation.

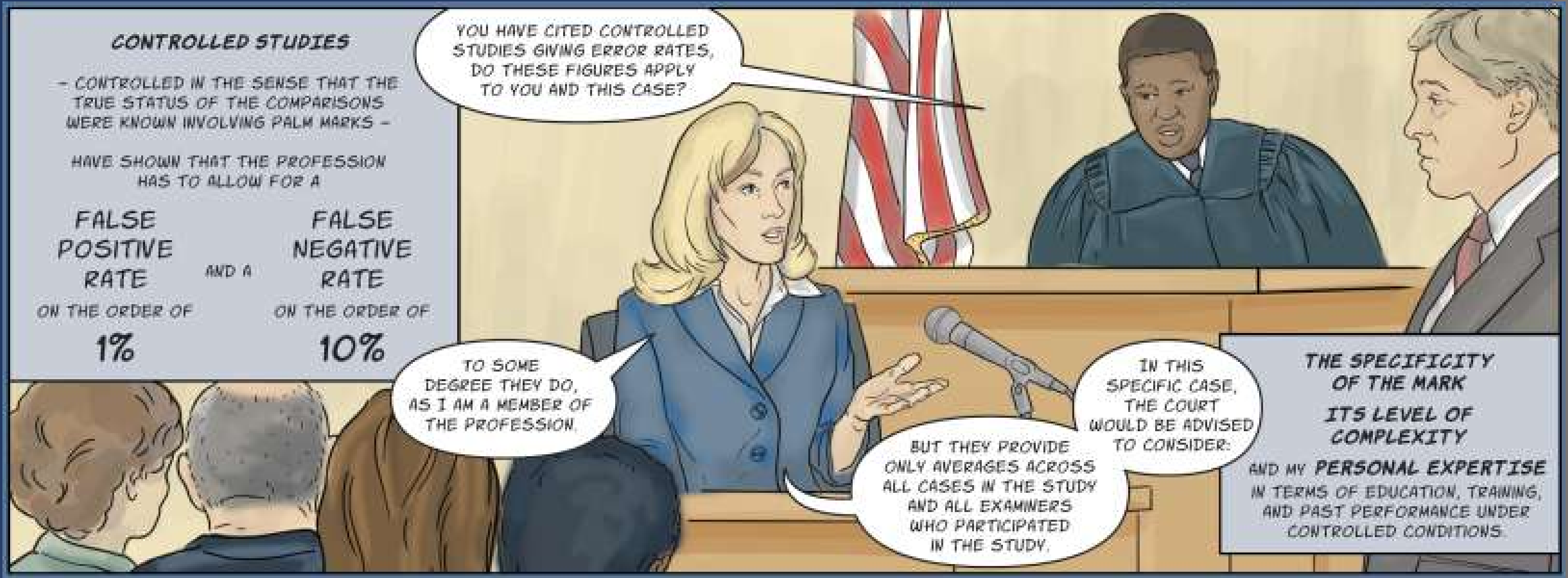
Table 1

The articulation language used to convey "yes,", "no," and "I don't know" for various forensic disciplines. This list is not meant to be exhaustive but rather to demonstrate the potential variation of terms between and within forensic disciplines.

	Results and articulation language		
Discipline	Yes	No	I don't know
Seized drugs	Present, confirmation, or determined to contain	Not present or does not contain	Inconclusive
DNA analysis (multiple rows = variation across laboratories)	Included	Excluded	Inconclusive or uninformative
	Included	Cannot be included	
	Cannot be excluded	Excluded	
Firearms	Identified	Eliminated	Inconclusive
Latent print (multiple rows = variation across laboratories)	Identified	Excluded	Inconclusive
	Associated	Excluded	
Bloodstain pattern analysis/pattern classification	Yes (could be)	No (eliminated)	Undetermined
Fire investigation (ignition and source)	Included	Excluded	Undetermined

What do factfinders hear?





https://zenodo.org/records/3734560

How do they hear it?

- REPORTS
  - Written
  - Frequently used for decisions
  - Freer format, but what is read?
  - Limited to no research
- TESTIMONY
  - Oral
  - Occasionally used for decisions
  - Constrained format
  - Much research but few solutions

Are they even listening?

- Central Processing
  - Engaged
  - Focus on appropriate cues
    - Data
    - Explanations
    - Experience

- Peripheral Processing
  - Bored / zoned out
  - Focus on inappropriate cues
    - Appearance
    - Likability
    - Background

Not only must we be understandable, we must be engaging!

Why can’t we just speak in plain English?

- Scientists value precision
- Clarity is *hard*

What about when others misrepresent our words?

- Interpretation Scales
  - 3-scale vs 5-scale
  - “I can’t say it’s him” (wink wink, nudge nudge)
  - Pushing the envelope with ID
  - Giving no useful information
  - Subjectivity
  - Fully continuous scale
  - So…what’s the effect?

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, July 2020, Vol. 65, No. 4
doi: 10.1111/1556-4029.14298
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

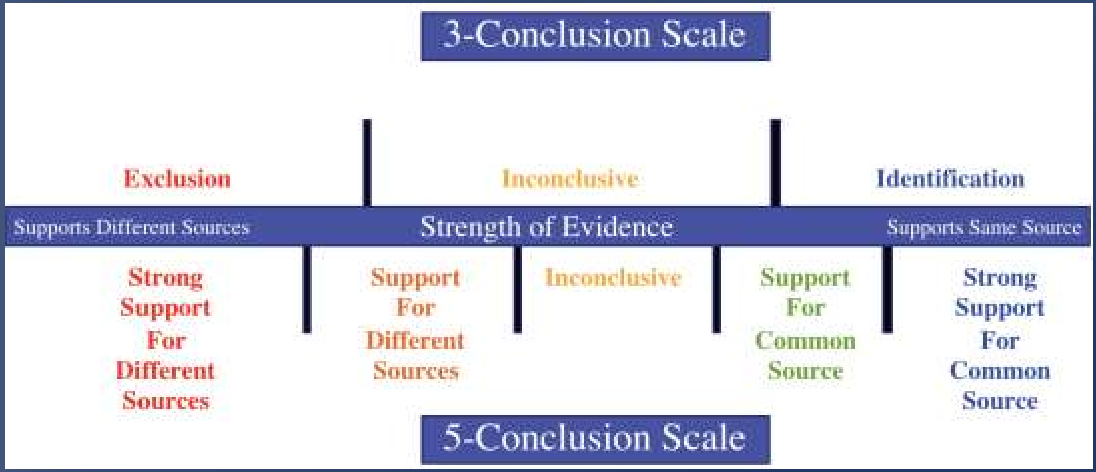
Kelly E. Carter,[1] B.A.; Macgregor D. Vogelsang,[1] B.S.; John Vanderkolk,[2] B.A.; and Thomas Busey,[1] Ph.D.

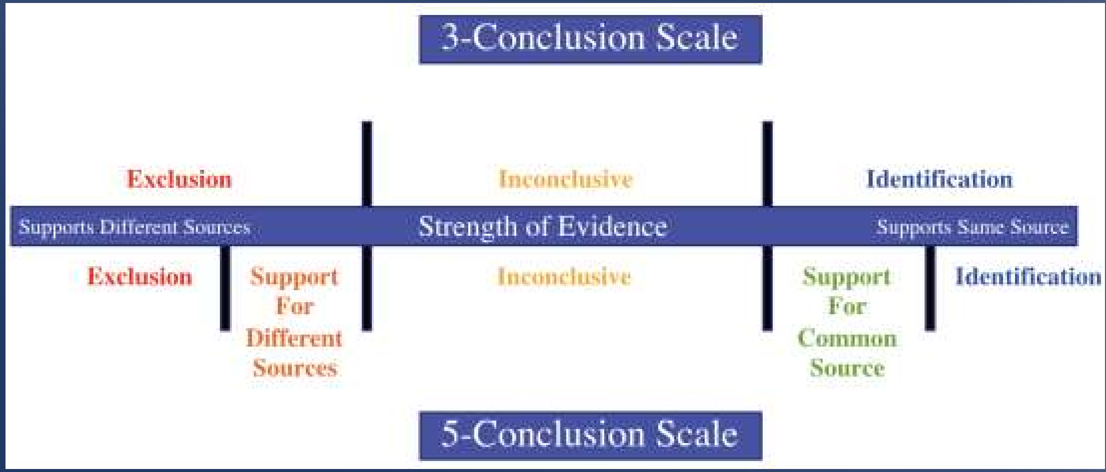
The Utility of Expanded Conclusion Scales During Latent Print Examinations

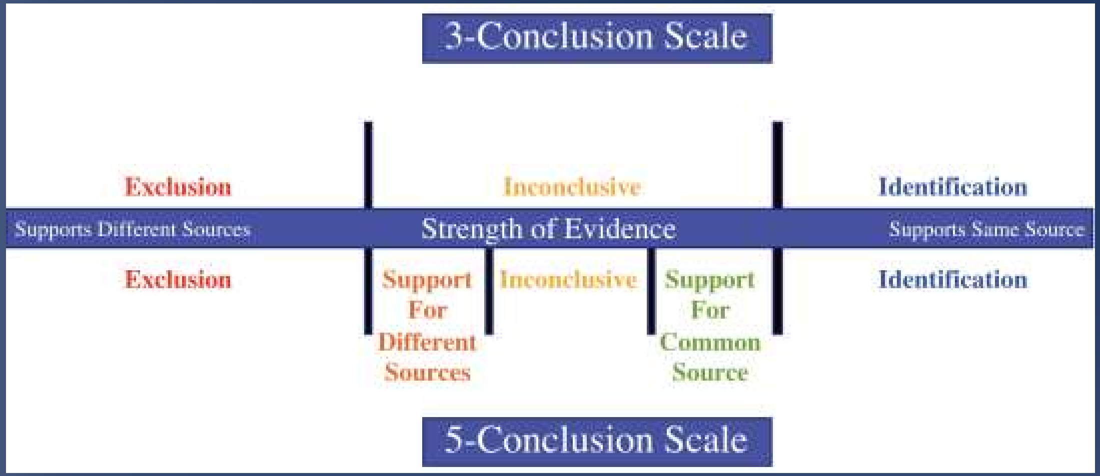
ABSTRACT: During fingerprint comparisons, a latent print examiner visually compares two impressions to determine whether or not they originated from the same source. They consider the amount of perceived detail in agreement or disagreement and accumulate evidence toward same source and different sources propositions. This evidence is then mapped to one of three conclusions: Identification, Inconclusive, or Exclusion. A limitation of this 3-conclusion scale is it can lose information when translating the conclusion from the internal strength-of-evidence value to one of only three possible conclusions. An alternative scale with two additional values, support for different sources and support for common sources, has been proposed by the Friction Ridge Subcommittee of OSAC. The expanded scale could lead to more investigative leads but could produce complex trade-offs in both correct and erroneous identifications. The aim of the present study was to determine the consequences of a shift to expanded conclusion scales in latent print comparisons. Latent print examiners each completed 60 comparisons using one of the two scales, and the resulting data were modeled using signal detection theory to measure whether the expanded scale changed the threshold for an “Identification” conclusion. When using the expanded scale, examiners became more risk-averse when making “Identification” decisions and tended to transition both the weaker Identification and stronger Inconclusive responses to the “Support for Common Source” statement. The results demonstrate the utility of an expanded conclusion scale and also provide guidance for the adoption of these or similar scales.

KEYWORDS: decision making, expanded conclusions, fingerprints, friction ridge, model comparison, identification

What about when others misrepresent our words?







- IDs of mated pairs 0.377 → 0.266
- Inc overall 0.569 → 0.351
- 17 ’erroneous SSS’ *but…*
- 97 correct SSS

“[W]e view it as important that consumers of investigative leads understand that these are not firm conclusions”

What about when others misrepresent our words?

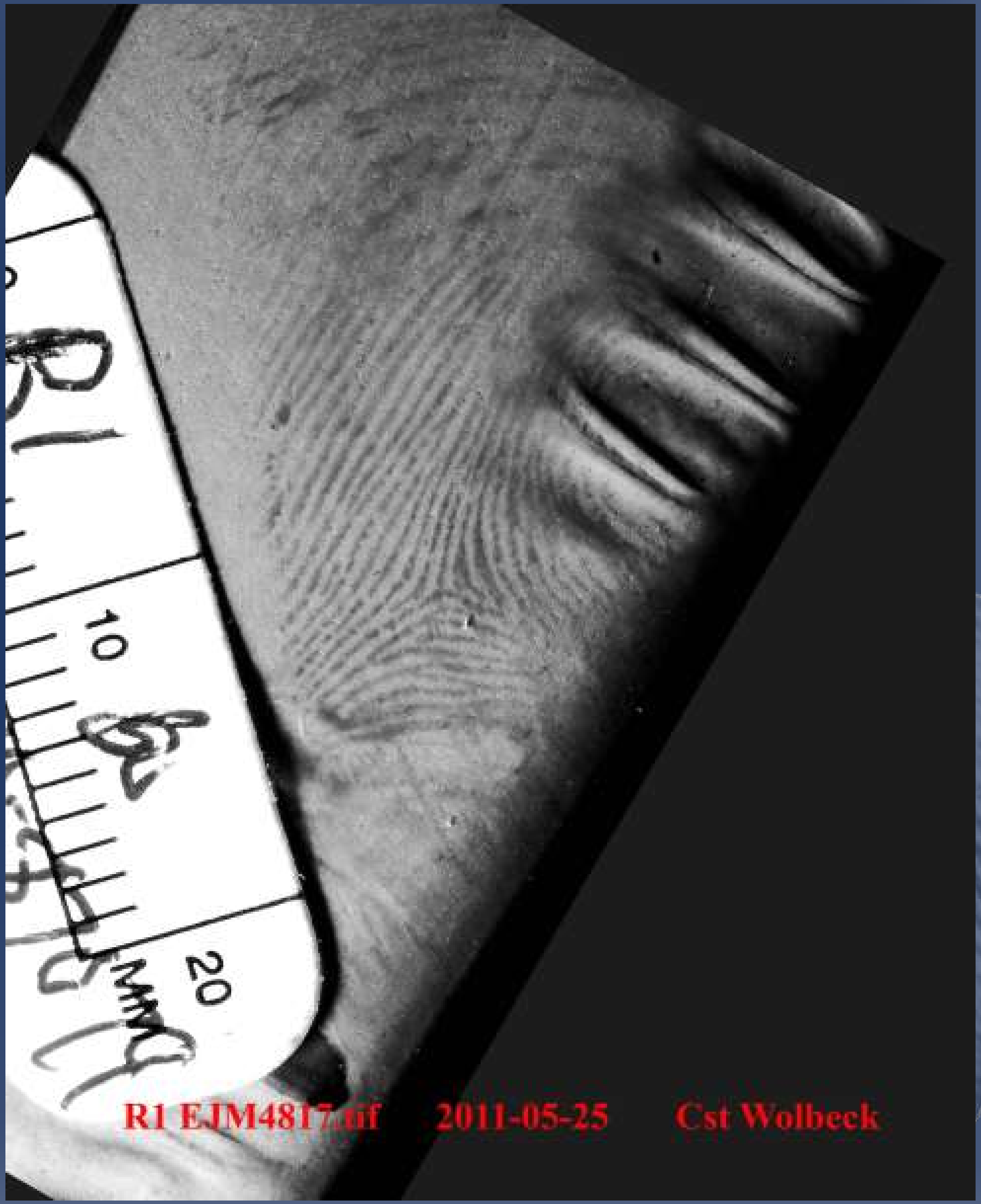
- Closing arguments
- Re-stating of our testimony
- Plea bargaining from reports
- Even judges on occasion…

Regina v Bornyk (2013, British Columbia)

- *"Following a day of legal argument I reserved judgment. During reserve, I became aware of further materials…"*
- *"[m]ost of the well-known errors have occurred in cases involving a single, distorted impression."*

--Eldridge, 2011

- Judge Funt acquitted because *"While the usable portion of the latent fingerprint and the known fingerprint are quite similar, I have more than a reasonable doubt that there is a match […]"*



Where do we go from here?

- Focus on development of *understandable* language
  - Cognitive psychologists—Linguistics
- Focus on development of ways to quickly and effectively communicate complex concepts
  - Cognitive psychologists—Learning
- Focus on development of effective visual aids
- Standardization of interpretation scales
- Research into efficacy of all above

Forensic Science International: Synergy 1 (2019) 24–34

Contents lists available at ScienceDirect

Forensic Science International: Synergy

journal homepage: https://www.journals.elsevier.com/forensic-science-international-synergy/

Juror comprehension of forensic expert testimony: A literature review and gap analysis

Heidi Eldridge

RTI International, 3040 E. Cornwallis Rd., Research Triangle Park, NC, 27709, USA

A R T I C L E I N F O

Article history:
Received 19 October 2018
Received in revised form
1 March 2019
Accepted 4 March 2019
Available online 9 March 2019

Keywords:
Expert testimony
Juror comprehension
Verbal scale
Likelihood ratio
Strength of evidence
Cognitive psychology

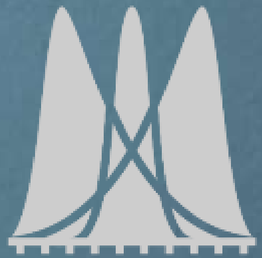
A B S T R A C T

Forensic scientists and commentators including academics and statisticians have been embroiled in a debate over the best way to present evidence in the courtroom. Various forms of evidence presentation, both quantitative and qualitative, have been championed, yet amidst the furor over the most “correct” or “accurate” way to present evidence, the perspective of the fact-finder is often lost. Without comprehension, correctness is moot. Unbeknownst to many forensic practitioners, there is a large, though incomplete, body of literature from the cognitive psychology domain that explores the question of what jurors understand when forensic scientists testify. This body of work has begun to test different proposed methods of testimony in an effort to understand which are most effective at communicating the strength of evidence that is intended by the expert. This article is a review of that literature that is intended for the forensic scientist community. Its aim is to educate that community on the findings of completed studies and to identify suggestions for further research that will inform changes in testimony delivery and ensure that any modifications can be implemented with confidence in their effectiveness.

Dr. Heidi Eldridge
George Washington University
heidi.eldridge@gwu.edu
202.994.3108

What is Successful Communication of Scientific Findings?

Professor Kristy Martire

Categorical conclusion

"…Suspect X's left shoe made the impression…"

Random-match probability

"…there is 1 chance in 1,000 of observing the evidence using a different shoe…"

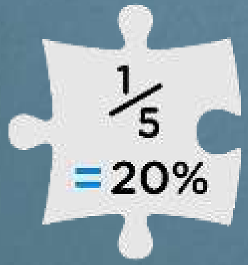
Verbal label

"…there is strong support for the proposition that Suspect X's left shoe made the impression…"

Likelihood ratio

"…the observed evidence is 1000 times more likely if Suspect X's left shoe made the impression…"

Bali, Martire, & Edmond, 2021; Martire, 2018; Martire & Edmond, 2020; National Research Council, 2009; Thompson, Grady, Lai, & Stern, 2018;
Icons created by Priyanka, Annamarie Kosto, Toli for Noun Project

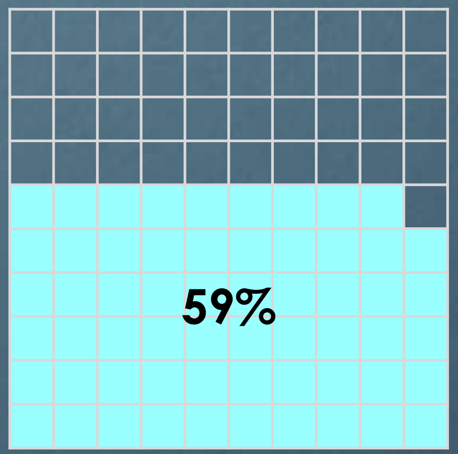


CONSISTENCY

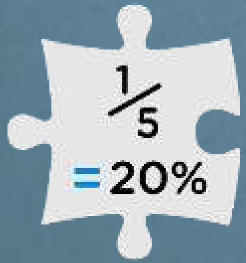
To give equal weight to evidence of equal strength

“1 in 1 million” Vs “0.0001%”

Lindsey, Hertwig & Gigerenzer, 2003



Bali et al., 2021

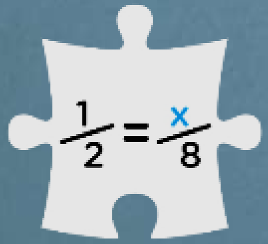


CONSISTENCY
To give equal weight to evidence of equal strength

Evidence that mathematical equivalence often does not guarantee psychological equivalence.

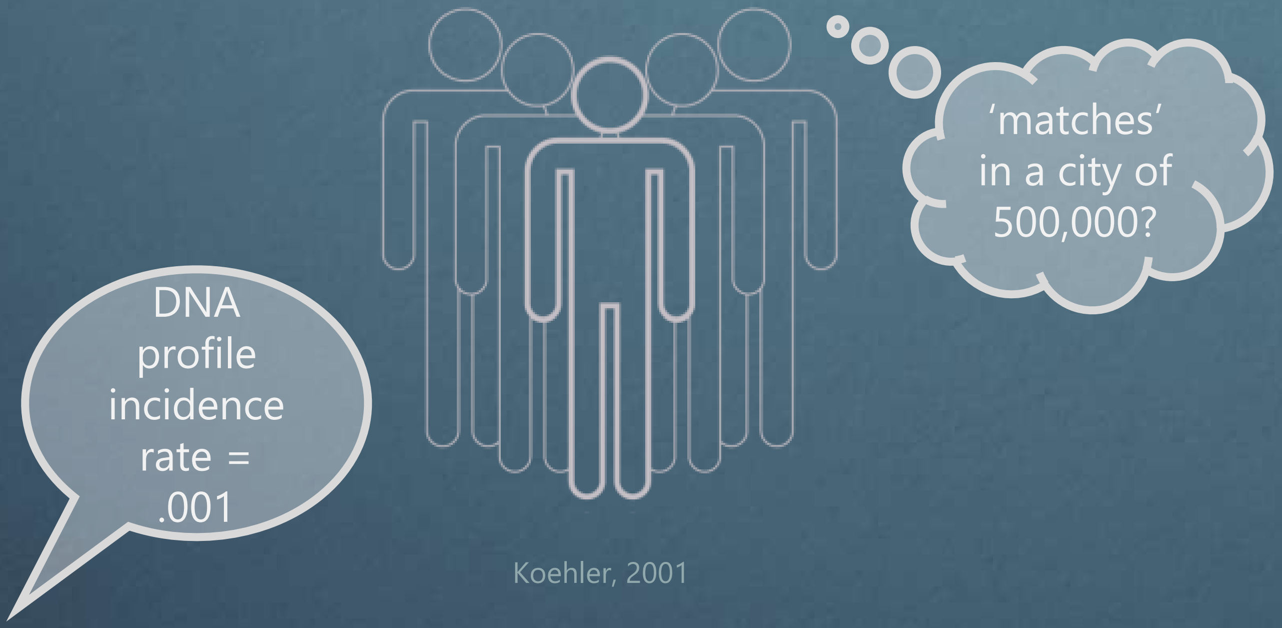
Martire & Edmond, 2020

Goodman, 1992
Lindsey et al, 2003
Koehler, 1996
Koehler, 2001
Martire et al, 2013
Martire et al, 2014
McQuiston-Surrett & Saks, 2009
Nance & Morris, 2002
Nance & Morris, 2005
Thompson & Schuman, 1987
Thompson & Newman, 2015
Wells, 1992

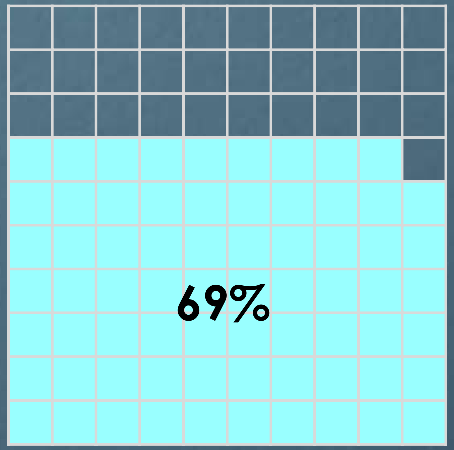


ABILITY

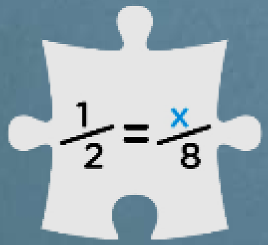
To be able to infer new information from the evidence



Koehler, 2001



Bali et al., 2021



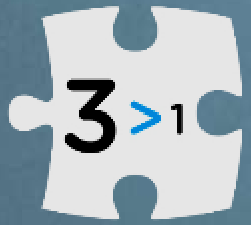
ABILITY

To be able to infer new information from the evidence

Evidence is limited and inconsistent

Goodman, 1992
Lindsey et al, 2003
Kaye et al, 2007
Koehler, 2001
McQuiston-Surrett & Saks, 2009

Martire & Edmond, 2020

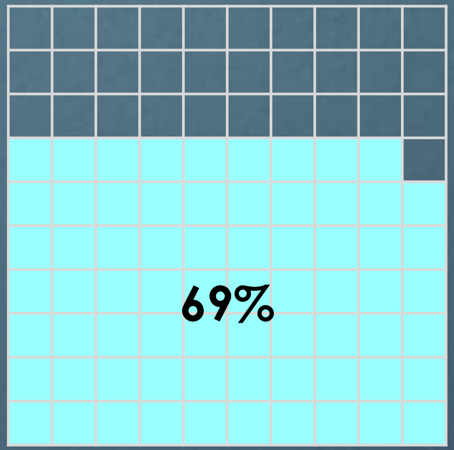


SENSITIVITY

To give more/less weight to evidence of greater/lesser strength

“5.5 times more likely” Vs “5500 times more likely”

Martire, Kemp, Sayle & Newell, 2014



Bali et al., 2021

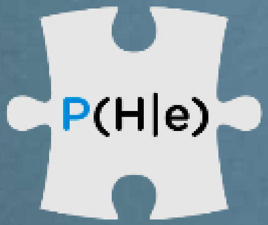
SENSITIVITY

To give more/less weight to evidence of greater/lesser strength

Evidence of broad (rather than precise) sensitivity to evidence strength

Martire & Edmond, 2020

De Keijser et al, 2016
Faigman & Baglioni, 1988
Goodman, 1992
Kaasa et al, 2007
Koehler, 1996
Koehler, 2001
Martire et al 2013
Martire et al 2014
Nance & Morris, 2002
Nance & Morris, 2005
Scurich & John, 2013
Smith et al, 1996
Thompson et al, 2013
Thompson & Newman, 2015

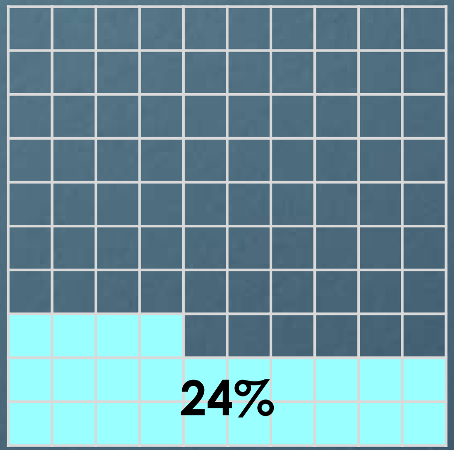


ORTHODOXY

To update beliefs in line with (Bayesian) normative expectations

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(E)}$$

Bayes Theorem



Bali et al., 2021

$P(H|e)$

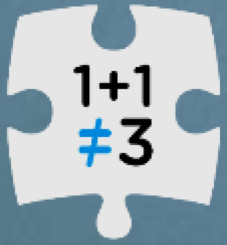
ORTHODOXY

To update beliefs in line with (Bayesian) normative expectations

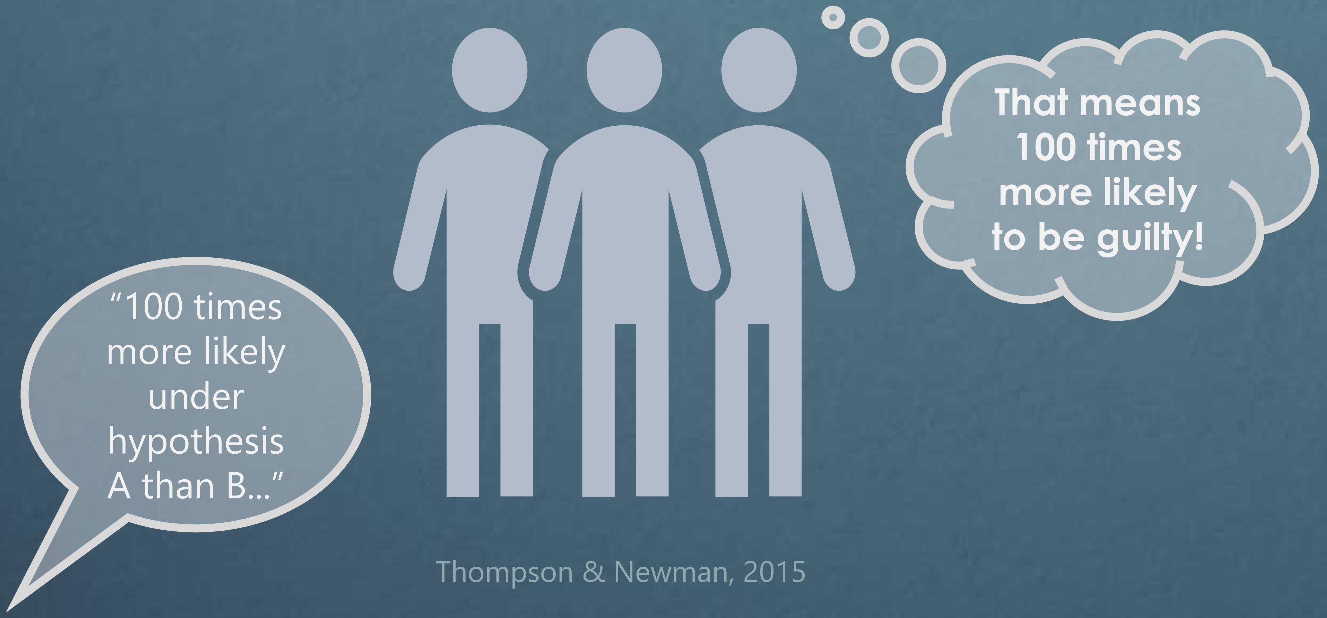
Evidence is mixed

Martire & Edmond, 2020

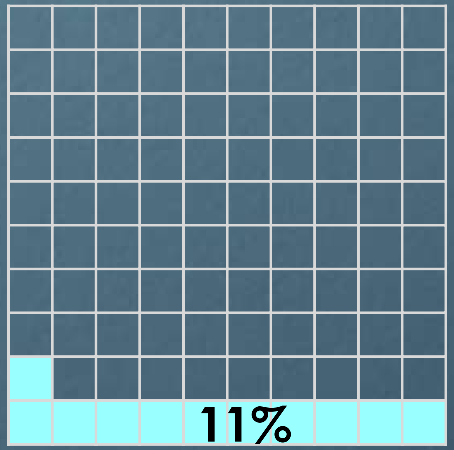
Goodman, 1992
Martire et al, 2013
Martire et al, 2014
Nance & Morris, 2002
Nance & Morris, 2005
Schklar & Diamond, 1999
Smith et al, 1996
Thompson & Schuman, 1987
Thompson et al, 2013
Thompson & Newman, 2015



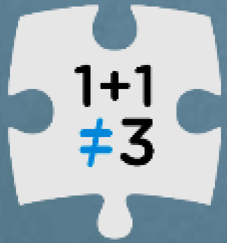
COHERENCE
To treat evidence in a logical and rational manner



Thompson & Newman, 2015



Bali et al., 2021



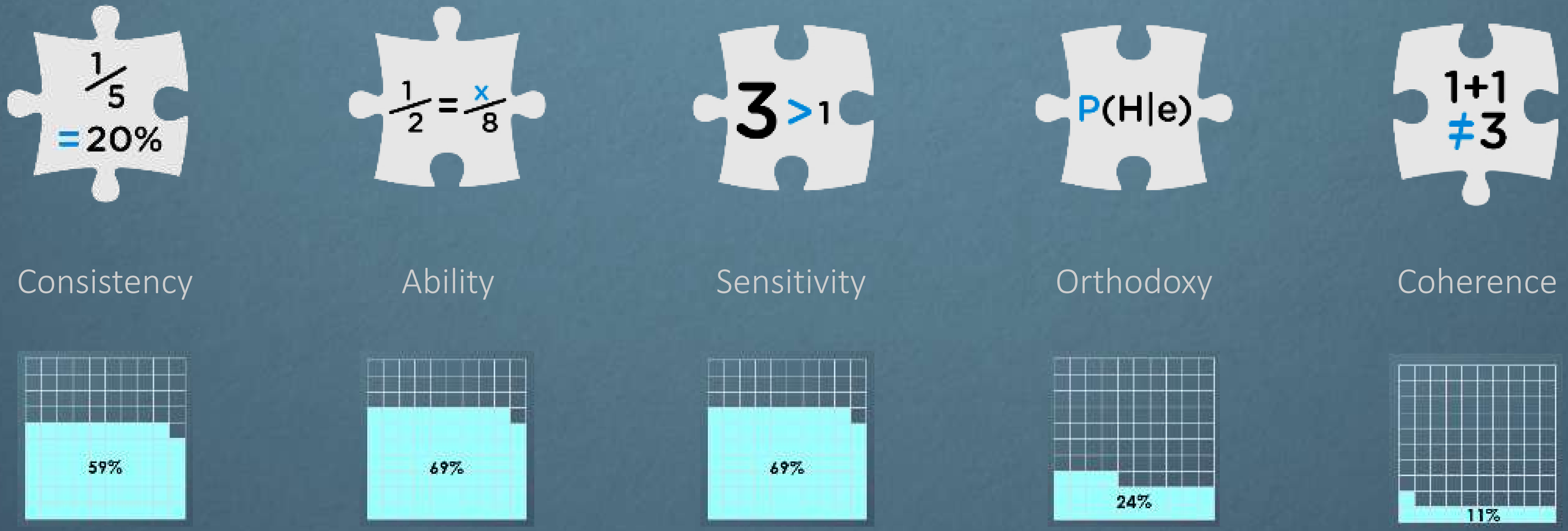
COHERENCE

To treat evidence in a logical and rational manner

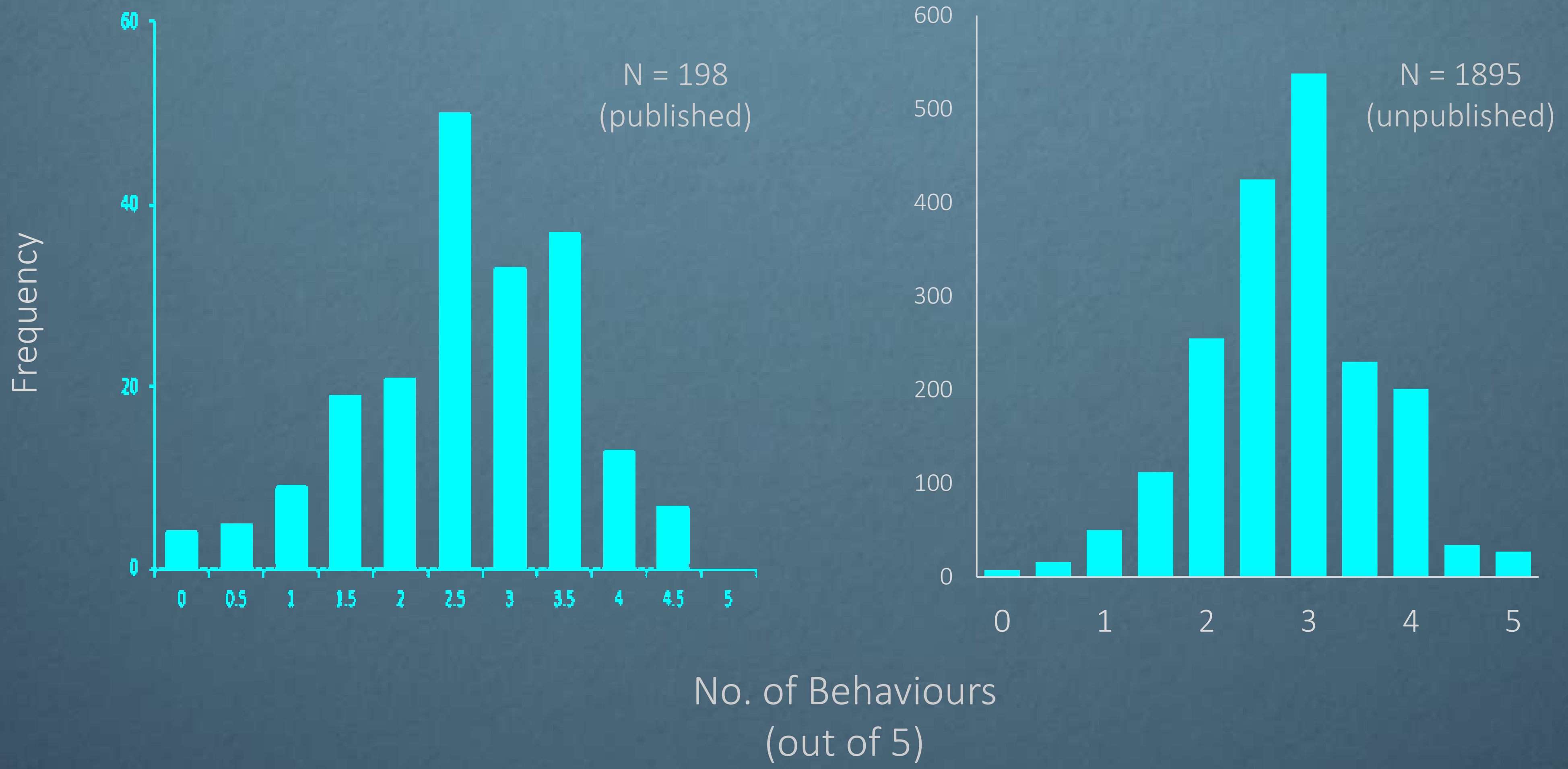
Clear evidence of aggregation errors and fallacious reasoning (e.g., defense attorney's fallacy)

Martire & Edmond, 2020

Goodman, 1992
Kaye et al, 2007
Koehler et al, 1995
Martire et al, 2013
Martire et al, 2014
Nance & Morris, 2002
Nance & Morris, 2005
Schklar & Diamond, 1999
Smith et al, 1996
Thompson & Schuman, 1987
Thompson et al, 2013
Thompson & Newman, 2015



Is this what successful communication of scientific findings looks like?



Bali et al., 2021, Bali Thesis

	Consistency	Ability	Sensitivity	Orthodoxy	Coherence
Random Match Probability					
Likelihood Ratio					
Verbal Label					

Bali et al., 2021

Qualifications

Evidence of training, study or certification directly relevant to the opinion

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

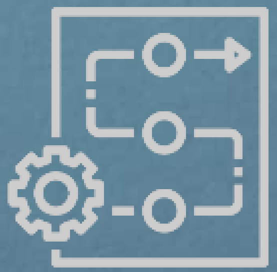
Proficiency

Proven track record of completing competent analyses and accurate opinions

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency



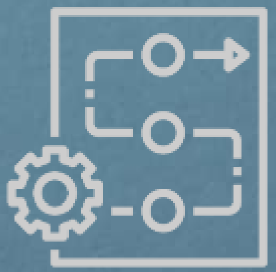
Procedure

What analyses were completed and in what way

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency



Procedure

Assumptions

What did/does the practitioner assume to be true when forming their opinion

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency

Procedure

Assumptions

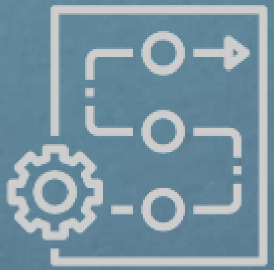
Validity

Evidence of the accuracy and reliability of the methods and procedures used

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency



Procedure

Assumptions

Validity

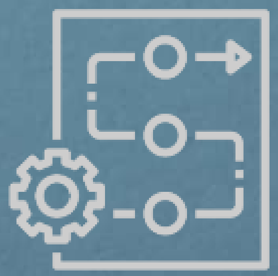
Human Factors

Information about who knew what when and how any potential for bias was managed

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency



Procedure

Assumptions

Validity

Human Factors

Limitations

Disclosures about evidence quality, contamination, non-conformities, peer disagreement etc.

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency

Procedure

Assumptions

Validity

Human Factors

Limitations

Conflict

Information about significant controversy's or disagreements relevant to the opinions provided

Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

Qualifications

Proficiency

Procedure

Assumptions

Validity

Human Factors

Limitations

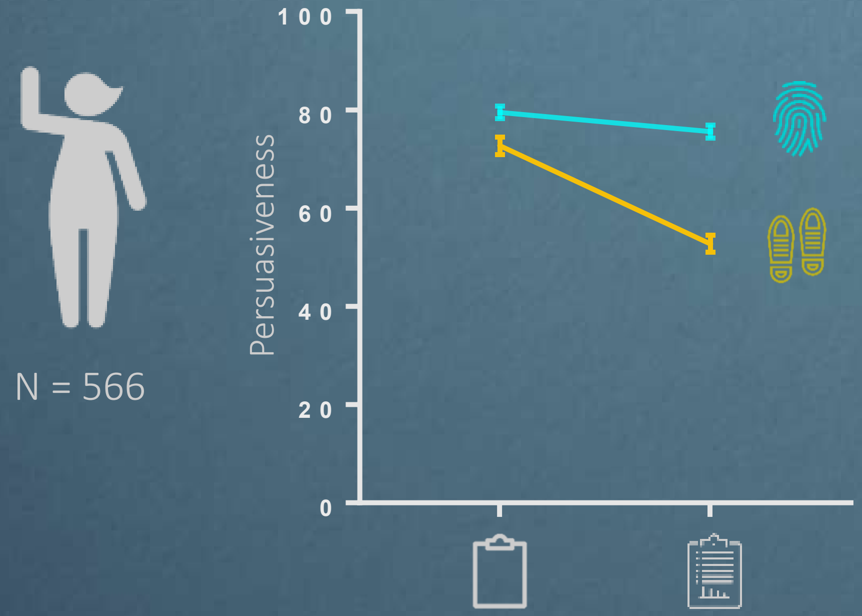
Conflict

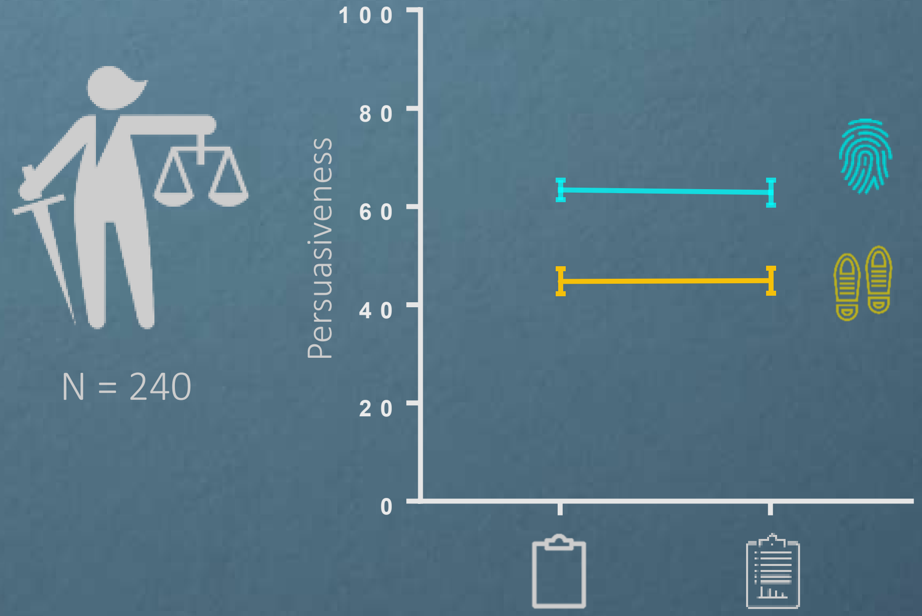
Summersby, Edmond, Kemp, Ballantyne & Martire, 2024; Icons by Saeful Muslim, dDara, Nithinan Tatah, Victoruler, Gregor Cresnar, Eucalyp, mikicon for Noun Project

"Only two properly designed studies...have been conducted...found false positive rates... that could be as high as 1 in 306 in one study and 1 in 18 in the other study."

"No properly designed studies... have been conducted, so we cannot give an accurate estimate of error rates."

Edmond et al., 2017; PCAST, 2016





Summersby et al *in prep*; Icons by mikicon, yoyon Pujiyono, Luis Prado, prakruti, icon 54, eucalypt, Creative Stall for Noun Project

What would it look like for someone to genuinely understand my scientific findings?

Thank you

k.martire@unsw.edu.au

Please share your feedback about this talk https://goo.gl/EUiOE9

- Bali, A. S., Martire, K. A., & Edmond, G. (2021). Lay comprehension of statistical evidence: A novel measurement approach. *Law and Human Behavior, 45(4), 370 390.* https://doi.org/10.1037/lhb0000457
- Koehler JJ. (2001). When are people persuaded by DNA match statistics? *Law and Human Behavior* 25(5):493 513.
- Lindsey S, Hertwig R, Gigerenzer G. (2002) Communicating statistical DNA evidence. *Jurimetrics.* 43:147 163.
- Martire, K. A. (2018). Clear communication through clear purpose: understanding statistical statements made by forensic scientists. *Australian Journal of Forensic Sciences, 50(6), 619 627.* https://doi.org/10.1080/00450618.2018.1439101
- Martire, K. A., & Edmond, G. (2020). How well do lay people comprehend statistical statements from forensic scientists. *Handbook of Forensic Statistics, 201 224.* https://osf.io/preprints/osf/67fgp
- Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International,* 240, 61 68. https://doi.org/https://doi.org/10.1016/j.forsciint.2014.04.005
- Summersby, S., Edmond, G., Kemp, R. I., Ballantyne, K. N., & Martire, K. A. (2024). The effect of following best practice reporting recommendations on legal and community evaluations of forensic examiners reports. *Forensic Science International*, 359, 112034. https://doi.org/10.1016/j.forsciint.2024.112034
- Thompson, W. C., Grady, R.H., Lai, E., & Stern, H (2018) Perceived strength of forensic scientists' reporting statements about source conclusions, *Law, Probability and Risk, 17(2), 133 155.* https://doi.org/10.1093/lpr/mgy012
- Thompson, W. C., & Newman, E. J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law and human behavior,* 39(4), 332.
- National Research Council (2009). *Strengthening forensic science in the United States: a path forward.* National Academies Press.

Communicating Forensic Findings Workshop: Current Practices and Future Directions

Session 2: Communicating Results in Forensic Reports and Testimony

Clinton Hughes – Brooklyn Defender Services

Question 1 – **Fantasy Island**

If we took all your advice and suggestions today on how to communicate results effectively – what would that look like tomorrow in lab reports or in court testimony?

Question 4 – **Culture Club**

Can you discuss more of the *cultural divide* between scientists and their use of language vs. what those reading the reports or hearing the testimony want/need? (*scientific language and precision vs layman's terminology*)

Fantasy Island – Example

Assumption – the greatest danger of assigning false support for a non-contributor in Forensic DNA Mixture Analysis occurs when relative(s) of the person mixture are contributors to a crime scene sample.

"**Likelihood Ratio** (LR): A measure of the relative strength of support that particular findings give to one proposition against a stated alternative. . ." (xix)

"We define **error** as the failure of a system to achieve its intended goal or outcome." (16)

"**Outcome Error:** An error in the final opinion or result." (xix)

Q. So your testimony is that the only false positives with STRmix come from misidentifying the number of contributors.

A. In addition to limit artefacts. Meaning when we're doing edits and removing artefacts, we didn't remove the artefacts. And multiple artefacts are actually aligning with alleles with that person of interest that we are comparing with the mixture. That may lead to false positives.

Q. False inclusion of values can also occur because of allele sharing between true contributors and noncontributors, correct?

A. Not necessarily.

Q. That's not my question, right?

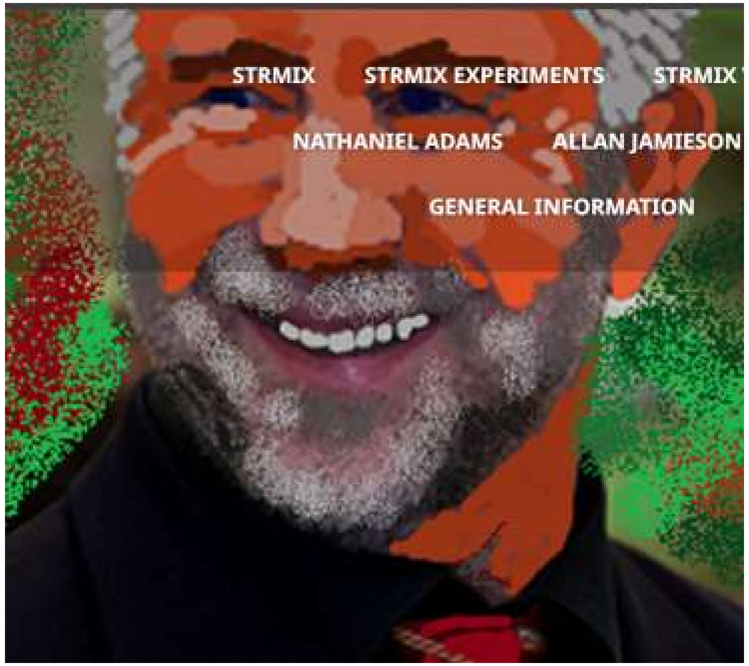
False inclusions can occur because allele sharing between true contributors and noncontributors, correct?

A. The answer is no.

Workshop Description – Day 1

"From the presentations and discussions, we are looking to examine . . . *any knowledge gaps that may impact an end user's understanding of the findings*."


MY THREE SONS


STRMIX
STRMIX EXPERIMENTS
NATHANIEL ADAMS
ALLAN JAMIESON
GENERAL INFORMATION

February 2021

AMERICAN ACADEMY OF FORENSIC SCIENCES

73RD AAFS ANNUAL SCIENTIFIC MEETING

ONE ACADEMY

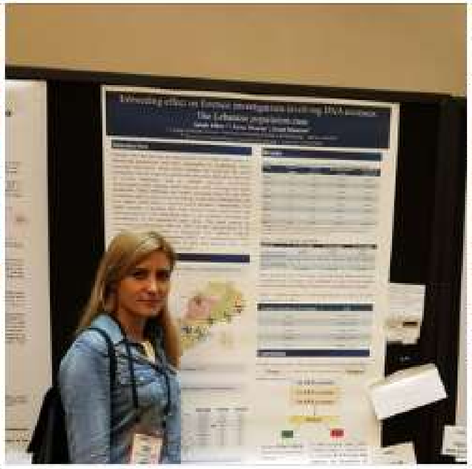
PURSUING JUSTICE THROUGH TRUTH IN EVIDENCE

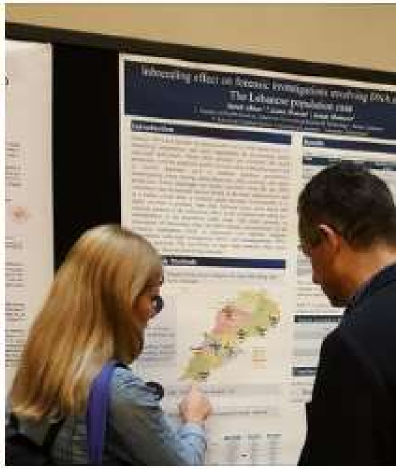
CRIMINALISTICS

The False Inclusion of Non-Contributors in DNA Mixtures Cases

Marie Semaan, MS; Sarah Abbas, MS; Issam Mansour, PhD*

(FSF Emerging Forensic Scientist Award Oral Presentation)





106

9 12

AUST - American University of Science and Technology is with **Marie S. Semaan** and **Sarah Abbas**.
September 18, 2017 ·

The American University of Science and Technology attended the 21st triennial meeting of the International Association of Forensic Sciences IAFS 2017 that was held in Toronto, Canada; on August 21-25 2017.
Two graduate students from the Faculty of Health Sciences, Ms. Marie Semaan and Mrs. Sarah Abbas, presented part of the research work conducted at AUST in the field of Forensic Science (DNA Analysis), under the supervision of Dr. Issam Mansour.
Ms. Semaan presented her Master's research in an oral presentation entitled "DNA Mixture Analysis in inbred Lebanese communities. Assessment of expert DNA mixture software". Whereas, Mrs. Abbas' communication entitled "Inbreeding effect on forensic investigations involving DNA mixtures: The Lebanese population case" was presented as part of her PhD research work conducted in collaboration with the University of Lausanne - Switzerland.
Forensic Science experts, professional organizations and delegates from around the world were available in this meeting to share information, new practices and advancements in the field.
Among the audience were Professor Frederick Bieber, Professor Pierre Margot, Dr. Michael Pollanen, world leaders in the field of Forensic Science and decision makers in forensic investigative strategies.
The meeting served as a unique opportunity for the AUST community to reassure their pivotal role in the advancement of forensic sciences in the Middle-East and North Africa region, introduce themselves to

Journal of Forensic Research
ISSN: 2157-7145
Case Report

A Mixed DNA Profile Controversy

Marie Semaan[1], **Sarah Abbas**[1,2] **and Issam Mansour**[1*]

[1]*Department of Laboratory Science and Technology, Faculty of Health Sciences, American University of Sci*

[2]*School of Criminal Justice, University of Lausanne, Switzerland*

***Corresponding author:** Issam Mansour, Department of Laboratory Science and Technology, Faculty
Technology, Beirut, Lebanon, Tel: +961218716230, E-mail: fhs@aust.edu.lb

Received date: February 01, 2020; **Accepted date:** March 10, 2020; **Published date:** March 20, 2020

9.5 Novemvigintillion (9.50E+90)

Journal of Forensic Research
ISSN: 2157-7145

Case Report

A Mixed DNA Profile Controversy

Marie Semaan[1], **Sarah Abbas**[1,2] **and Issam Mansour**[1*]

[1]*Department of Laboratory Science and Technology, Faculty of Health Sciences, American University of Sci*

[2]*School of Criminal Justice, University of Lausanne, Switzerland*

***Corresponding author:** Issam Mansour, Department of Laboratory Science and Technology, Faculty
Technology, Beirut, Lebanon, Tel: +961218716230, E-mail: fhs@aust.edu.lb

Received date: February 01, 2020; **Accepted date:** March 10, 2020; **Published date:** March 20, 2020

JOURNAL OF FORENSIC SCIENCES

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

PAPER

Criminalistics

JOURNAL OF FORENSIC SCIENCES

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

2.86 Quintillion

(2.86E+18)

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
$LR_{iu/uu}$								
STRmix™ (GlobalFiler 21 loci)								
1:1	0	0	1.65×10^{19}	0	0	2.84×10^{18}	2.86×10^{18}	0
3:1	0	0	6.40×10^{27}	0	0	1.07×10^{27}	0	0
$LR_{ij/ju}$								
STRmix™ (GlobalFiler 21 loci)								
1:1	0	0	1.43×10^{28} $LR_{3,6/6,U}$	0	0	2.43×10^{27} $LR_{3,6/3,U}$	0	0
3:1	0	0	1.53×10^{28} $LR_{3,6/6,U}$	0	0	2.33×10^{27} $LR_{3,6/3,U}$	0	0

JOURNAL OF FORENSIC SCIENCES AAFS

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

TABLE 6 Unconditioned LRs for Experiment 4 (4:1 low-level mixture) using STRmix™ and data from El Andari et al. (6) and $\theta = 0.01$

i	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
$LR_{iu/uu}$	1.46×10^{10}	9.14×10^{11}	5.86×10^{23}	1.58×10^{11}	7.40×10^{8}	4.61×10^{11}	1.80×10^{10}	4.07×10^{12}

Note: All eight references give values that support the first proposition compared to the alternative when no conditioning profiles are used.

4.07 Trillion

(2.86E+12)

JOURNAL OF FORENSIC SCIENCES

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"Most software deal with dyadic relationships, that is relationships between two individuals. . ."

PAPER

Criminalistics

JOURNAL OF FORENSIC SCIENCES AAFS

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"*Neither* STRmix™ *nor* LRmix deal with *triadic situations* or higher, although DBLR™ does [22]."

JOURNAL OF FORENSIC SCIENCES

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"Neither LRmix *nor any other software* or interpretation method can claim that the rate of false support is *zero*."

JOURNAL OF FORENSIC SCIENCES

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"... .there will always be uncertainty about the source of the DNA, as we cannot know who left the DNA trace."

PAPER

Criminalistics

JOURNAL OF FORENSIC SCIENCES AAFS

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"… this explains why DNA (or any evidence) should not be solely relied upon to reach a conclusion, but instead must be considered in combination with the other elements of the case."

JOURNAL OF FORENSIC SCIENCES

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"Empirical work has previously been reported assessing the risk of false support to a non-donor who is related to the true donor(s) **(see for example [4])**."

PAPER

Criminalistics

A mixed DNA profile controversy revisited

Tim Kalafut PhD[1] | Simone Pugh MS[2] | Peter Gill PhD[3,4] | Sarah Abbas MSc[5,6] | Marie Semaan MSc[5] | Issam Mansour PhD[5] | James Curran PhD[7] | Jo-Anne Bright PhD[8] | Tacha Hicks PhD[9,10] | Richard Wivell BSc (hons)[8] | John Buckleton DSc[7,8]

"Empirical work has previously been reported assessing the risk of false support to a non-donor who is related to the true donor(s) (see for example [4])."

Colorado-Bureau-Investigation-2018-STRmix-Validation_Summary.pdf. 2018. **https://indefenseof.us/uploads/ColoradoBureau-Investigation-2018-STRmix-Validation_Summary.pdf.** Accessed 30 Jul 2021.

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Hari Iyer
Rich Press
Melissa K. Taylor
Peter M. Vallone
Sheila Willis*
*International Associate under contract; retired director of Forensic Science Ireland

"An important **missing element** from many validation studies is the **degree of allele** sharing that has been tested." (86)

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

John M. Butler
Hari Iyer
Rich Press
Melissa K. Taylor
Peter M. Vallone
Sheila Willis*
*International Associate under contract; retired director of Forensic Science Ireland

"If validation studies are conducted using mixtures that do not explore the complexity induced by allele sharing, **the user may inadvertently extrapolate validation results** and **apply methods beyond the limits of the validation studies** conducted." (89)

Investigation into the effect of mixtures comprising related people on non-donor likelihood ratios, and potential practises to mitigate providing misleading opinions

Tim Kalafut [a], Jo-Anne Bright [b], Duncan Taylor [c,d], John Buckleton [b,e,*]

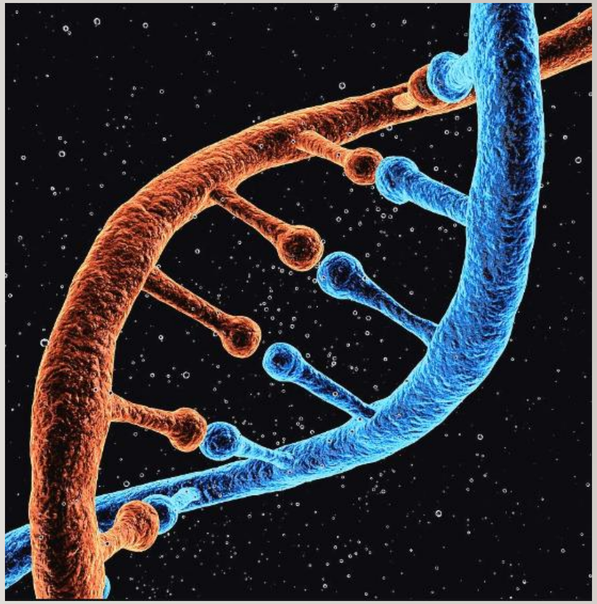
"The analysis of the in vitro and in silico mixtures assuming NoC = 3 with no use of a conditioning profile or *with the use of a conditioning profile* but without informed priors on the mixture proportions (Mx priors) was *ineffective*."

Workshop Description – Day 1

"From the presentations and discussions, we are looking to examine . . . *any knowledge gaps that may impact an end user's understanding of the findings*."

Thank you –
Dr. Sandra Koch,
Donna Ramkissoon,
and all the other folks at NIST!

COMMUNICATING FORENSIC BIOLOGY FINDINGS



JARRAH R. KENNEDY

ASST. DNA SUPERVISOR, KANSAS CITY POLICE CRIME LABORATORY

CHAIR – HUMAN FORENSIC BIOLOGY SUBCOMMITTEE (OSAC)

*Disclaimer- most, if not all, of the content of this presentation are my own opinions!

CFF – BIOLOGY POINTS TO COVER:

1) HIERARCHY OF PROPOSITIONS: what is the issue and can we help?
2) COMMUNICATING BIOLOGICAL RESULTS IN THE US
   - Serology /biological screening (is it blood? semen?)
   - DNA results/comparisons (is POI a contributor or not?)
   - Numbers? Words? Both?
3) STRATEGIES FOR COMMUNICATION

TYPICAL QUESTIONS THAT FORENSIC BIOLOGY MAY BE ABLE TO ***HELP*** WITH

1) What is the nature of this material?
   - Blood?
   - Semen?
   - Saliva?
   - Limited: feces/urine
2) Is there detectable DNA on an item?
3) Is the POI contributing DNA to the item or not?

We often cannot directly answer these questions

HIERARCHY OF PROPOSITIONS (LATE 90s)

LEVEL	QUESTION/ISSUE	RESULTS	EXAMPLE PROPOSITIONS
Offense	Is the POI the offender?		• POI sexually assaulted Ms. A • POI had nothing to do with the assault of Ms. A
Activity	Did the POI perform the activity?	• Presence/absence of DNA • Quantity/quality of DNA • DNA profile comparison (does not always have to be uncontested) • Presumptive tests • Multiple traces	• POI digitally penetrated the vagina of Ms. A • POI and Ms. A only had social interactions
Source	Is the POI the source of the biological material (such as blood or semen)?	• DNA profiling comparison • Obvious nature of the material (large bloodstains, millions sperm heads, bone)	• The blood came from POI • The blood came from some other man

Citation: Cook et al. A hierarchy of propositions: deciding which level to address in casework. Science & Justice 1998; 38: 231–239.

HIERARCHY OF PROPOSITIONS – EXPANDED "SOURCE"

LEVEL	QUESTION/ISSUE	RESULTS	EXAMPLE PROPOSITIONS
Source	Is the POI the source of the biological material (such as blood, semen)?	• DNA profile comparison	• POI is source of the bloodstain • An unknown, unrelated person is the source of the bloodstain
Sub-Source	Is the POI the source of the DNA?		• POI is source of the DNA • An unknown, unrelated person is the source of the DNA
Sub-sub-source	Is the POI the source of a component of the DNA?		• POI is the major contributor of the DNA mixture • An unknown, unrelated person is the major contributor of the DNA mixture

Why? – Increased sensitivity of DNA kits (+ more loci, highly discriminating)

No longer just testing sources where biological source is obvious – we are assessing ***DNA TRACES***

Should we then consider transfer, persistence, and background?

Citation: Hicks, T.; Buckleton, J.; Castella, V.; Evett, I.; Jackson, G. A Logical Framework for Forensic DNA Interpretation. Genes 2022, 13, 957. https://doi.org/10.3390/genes13060957

FORMULATING PROPOSITIONS (typically sub-source)

- Task-relevant case information (framework of circumstances)
  - Where is the sample from?
  - Are there any sources of expected DNA (intimate, prevalence of owner/user DNA)
  - Are there potential relatives that could have expected DNA
    - Can they be typed? Conditioning can help….
  - Is a relative of the POI an alleged alternative POI?!
    - Will we know this information?

- Ideally this is done **prior to** assessing evidence to avoid findings-led propositions

CURRENT STATE OF COMMUNICATING BIOLOGICAL SCREENING RESULTS IN THE US

- Tests that give information about the nature of the material (blood? semen?)
- Some tests call themselves 'confirmatory' – but this is a misnomer
- These tests are variably reported such as positive, weak positive, negative
- Sometimes as "identification" of the tested biological material - appears as "facts"
  - Instead of communicating the value of the findings (by say, considering false positives or negatives, appearance of the stain, etc)
- The factfinder is often left to correlate these findings with the DNA results

CURRENT STATE OF COMMUNICATING DNA RESULTS IN THE US: PGS/LRs

The focus of a LR is on the results (NOT on the propositions!)

- Software assisted method to discern DNA profiles and provide value to the comparison
- Two primary software used in the US* that help separate DNA mixtures
  - STRmix™ being used by more than 80 organizations in the US (as of 1/11/2024)[1]
  - TrueAllele® - 10 user labs[2]
- Software help discern profiles (mixtures) and also assign Likelihood Ratio values for DNA comparisons
  - Differ on user inputs (NOC, thresholds) and some modelling
  - Differ on how an LR is communicated

But…what about "exclusions"?

*Approximately 400 publicly funded laboratories https://bjs.ojp.gov/funding/awards/15pbjs-23-gk-00836-bjsb - estimated ~220 U.S. DNA labs https://le.fbi.gov/science-and-lab/biometrics-and-fingerprints/codis/codis-ndis-statistics

1: https://www.strmix.com/news/strmix-has-produced-dna-evidence-in-more-than-530000-criminal-cases-worldwide/

2: Defeating opposition experts: winning with science (cybgen.com)

CURRENT STATE OF COMMUNICATING DNA RESULTS IN THE US: MANUAL METHODS

- Use of terms to convey the similarity between profiles from evidence and people
  1. Similar? “match” “included” “cannot be excluded”
  2. Not sure? “inconclusive”
  3. Not similar? “excluded” “cannot be included”
- Use of statistics is typically provided if “similarity” observed. Only expresses the rarity of the unknown evidentiary profile is (CPI/RMP etc)
- Sometimes conclusions about “source” were/are drawn within a ‘reasonable degree of scientific certainty.’

Lacks the balance of LR framework

COMMUNICATING THE VALUE OF DNA COMPARISONS TO THE END-USERS / FACTFINDER…

- Complex scientific topics can be challenging to convey - lots of jargon and can be difficult to describe our process and testing in a way that is easily digestible in a short period of time
- ***Numbers are hard***
  - Any statistical concept is going to be difficult to communicate to lay persons
  - This misunderstanding did not improve with the transition to likelihood ratio framework
  - Common fallacies associated with expressing evidence value numerically
    - By the speaker and the end-user
    - Attaching a probability to a proposition (whether RMP/LRs)
    - Source attributions (rarity does not equal unique)

COMMUNICATING LRs TO THE END-USERS / FACTFINDER… TACKLING NUMERACY ISSUES #1: **CLEARLY STATE THE LIMITATIONS!!!**

What DNA CANNOT help with!

1. Difficult to attribute a profile (or portion) to a biological material ('cellular source')
2. DNA profile comparisons only ***help*** to address WHOSE DNA may or may not be detected.
   - No conclusions about "identity" or "source attribution" can be supported by an evaluation or the magnitude of the LR
   - DNA comparison results should be viewed as one part of the puzzle
3. It is crucial that it is understood that a LR assigned for a DNA comparison cannot be carried up to a question about an activity.
   - Actions? Timing? Motives?
   - The value of the evidence given sub-source level propositions has no meaning in that context (recall the hierarchy).

HUMAN FACTORS REPORT: TESTIMONY CHAPTER

Recommendation 6.2: When explaining the nature of DNA analysis during testimony, the DNA expert should address common misconceptions and state the limitations of the analysis. At a minimum, the DNA expert should address the following main points:

- The DNA results are only part of the overall case.
- Errors can occur in any human process, including DNA analysis.
- The evaluation of the DNA comparison cannot conclusively identify an individual as the source of the DNA.
- DNA analysts cannot provide any information on how or when DNA was deposited in a particular case, based on a report considering only the source of the DNA.

Expert Working Group on Human Factors in Forensic DNA Interpretation (2024). Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach. National Institute of Standards and Technology, Gaithersburg, MD, NIST IR 8503. https://doi.org/10.6028/NIST.IR.8503

COMMUNICATING TO THE END-USERS / FACTFINDER… TACKLING NUMERACY ISSUES #2: **CONSIDER CAPPING THE LR?**

- Can a cap to the LR prevent some common misconceptions or cognitive fallacies associated with extremely large numbers?
  - Something smaller than the world's population for communication/comprehension
  - Impact of hearing numbers never heard before (octillions?) - will it overshadow other evidence?
  - Do we need such large values to *adequately* convey the strength of the comparison?
- No recent studies on larger 20+ locus kits- but the studies done support **1 billion** as cap (if considering an individual unrelated to the POI)
  - There are many different "caps" though … so it seems to be a matter of preference and policy as well
    - UK, Swiss (1 billion); Australia (100 billion); Denmark (1 million)

COMMUNICATING TO THE END-USERS / FACTFINDER… TACKLING NUMERACY ISSUES #3: **USE VERBAL EQUIVALENTS ??**

- There are many scales - which one is merely a matter of choice, convention, or consensus.

WORDS ARE HARD, TOO!! They mean different things to different people.

- No special "DNA" scale - should work across disciplines (LR=1,000 same whether DNA or glass)
- Verbal qualifiers are only applied after the numerical LR value is assigned. These terms should not stand alone or replace the communication of the LR value.
- Verbal qualifiers should reference both propositions by conveying the support the DNA results provide for one proposition versus the other.

- Many verbal scales stop at an LR of 1 million. Once the LR exceeds (or goes below if the results provide support for H2 vs H1) it becomes difficult to find more words that convey additional meaning [Extremely very strong support ???!!!]

COMMUNICATING TO THE END-USERS / FACTFINDER… TACKLING NUMERACY ISSUES #3: **EXAMPLE VERBAL EQUIVALENTS**

LR for H_p Support and 1/LR for H_d Support	Verbal Qualifier
1	Uninformative
2 – 99	Limited Support
100 – 9,999	Moderate Support
10,000 – 999,999	Strong Support
≥ 1,000,000	Very Strong Support

Scientific Working Group on DNA Analysis Methods (SWGDAM). Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios. 2018. https://www.swgdam.org/_files/ugd/4344b0_dd5221694d1448588dcd0937738c9e46.pdf

LR	Verbal Communication
> 10,000	This support is qualified as *extremely strong.*
> 1000 – 10,000	This support is qualified as *very strong.*
> 100 – 1000	This support is qualified as *strong.*
> 10 – 100	This support is qualified as *moderate.*
> 1 – 10	This support is qualified as *weak* or *limited.*[250]
1	The results support neither proposition. This support is qualified as *null.*

Marquis R, Biedermann A, Cadola L, Champod C, Gueissaz L, Massonnet G, Mazzella WD, Taroni F, Hicks T. Discussion on How to Implement a Verbal Scale in a Forensic Laboratory: Benefits, Pitfalls and Suggestions to Avoid Misunderstandings. Science & Justice. 2016; 56(5):364-70. doi:10.1016/j.scijus.2016.05.009

COMMUNICATING TO THE END-USERS / FACTFINDER… SOLUTIONS?

- EDUCATION: How do we/can we educate our end-users
- This differs on setting:
  - Investigators, lawyers
    - Our results are being used without us often (plea deals) – are they properly understood?
  - Factfinders
- Can we find better methods to convey our results in court
  - Starting with the end? (thank you Julie Burrill)
- How we speak matters – but how much? Can we study this in a real setting?
  - Jargon, confidence, trust
  - Speaking like a real person

THE END

CRITERIA FOR REASONING WHEN THERE IS UNCERTAINTY …

- **Balance:** Assessing the evidence in light of clearly defined competing views.
- **Transparency:** Clear delineation of the assumptions and data relied on.
- **Robust:** Can the reasoning stand up to scrutiny? Is the evaluation repeatable?
- **Logic:** Is the reasoning coherent / does it make sense? Are the inferences based upon the results?

PRINCIPLES OF INTERPRETATION

What is the issue?

1. Interpretations take place in a *framework of circumstances*
2. At least two competing propositions must be considered
3. Analysts must assign the probability of the (results/findings) – NOT of the alleged fact (e.g., who the source of the DNA is, what activity happened)
4. The value of the findings is expressed by the ratio of the probability of the findings given the case information and the propositions considered (LR)

Balance!

LIKELIHOOD RATIOS - DNA COMPARISONS

STAYING IN OUR LANE = TALK ABOUT RESULTS

- How we assess the similarity/differences between a reference and unknown profile
- Provides value to the DNA comparison
- Balanced and flexible – can evaluate the DNA results given two scenarios
  - The scenarios state the sources of DNA (POI vs. unknown)
  - May need to consider an individual related to the defendant?
  - May need to consider varying propositions if there are multiple defendants?

Balance = 2 propositions!

- When communicating the LR – the value of the DNA results must be stated in respect to the scenarios considered:
  - The DNA profile is 1 million times more likely to be observed if (scenario 1) than if (scenario 2).

EXTRA SLIDES – TPPR/ACTIVITY LEVEL PROPOSITIONS

EVOLUTION OF DNA TECHNOLOGY: SENSITIVITY

- The ability to detect genetic information has evolved significantly since inception
- Used to be very large visible bloodstains or semen from sexual assaults
- Now – we regularly test firearms, steering wheels, and spent shell casings because we can detect very small amounts ("trace") of DNA.
- Naturally, this leads to questions that other trace disciplines have had to deal with for a long time – with material this small and transient – the "***relevance***" or how or when the DNA got there is more important.

IT'S NOT JUST "TRANSFER"

- Implied in "transfer" is often whether it was "direct" or "indirect"
    - These statements imply an action. Is this DNA there because POI touched it – or because they hung out with a friend (who touched it)?
- Persistence: Expected loss of DNA over time (considerations of timing, nature of material)
- Prevalence: Expected DNA from individuals (your DNA on your own steering wheel)
- Recovery: Methods used to collect and detect DNA profiles (swabs/tape lift, old vs. new kits)
- Background: DNA present from unknown persons for unknown reasons
- Contamination: DNA from POI not from alleged actions but from scene/lab problems

LIMITATION REMINDER

- The DNA comparisons we perform at KCPCL – and their corresponding value – do not address these transfer questions.
- Once the questions move from "who" to "how or when" – it can be challenging to ensure that the (large?) LR value doesn't get carried up.
  - If these questions are being asked, typically everyone is agreeing that the DNA is from the PC
- Recall hierarchy slide – there are more factors than the DNA profiles to consider:
  - Biological screening tests (blood, semen, saliva)
  - Quantity/quality of profiles (how much and the contribution of DNA)
  - Published or in house data about probability of recovering DNA given *specific* actions
    - All the criteria on previous slide

Nearly impossible to do during testimony

WHY DON'T WE DO THIS EVALUATION THEN?

- In depth education, training, competency, building an entire QA program to support **laboratory-based** interpretations and reporting for disclosure.
  - Very limited training opportunities– as in nothing sufficient in US. University of Lausanne
- This is challenging and new to the U.S.'s thinking (STRmix 10+ years ago).
- Consideration of resources – not every case needs this, but the training needed is considerable.
- Due to the challenge- many people are still resorting to saying statements that are not appropriate given our current understanding:
  - Direct transfer is more likely than indirect
  - Sure, that is possible or one of many possibilities

This is NOT science!

WHY WE DON'T TESTIFY TO "POSSIBILITIES" AND EXPLANATIONS

possible ≠ probable

- There is an important distinction between "possible" and "probable"
- You'll notice with DNA comparisons we don't state it's possible …. We provide a value of the comparison (this is based on probabilities)
- Typically we are only asked about possibilities when it pertains to questions about how or when the DNA may have been transferred to the item in question
- Possible is not the same as probable (fair coin heads vs. lottery)

WHY WE DON'T TESTIFY TO "POSSIBILITIES" AND EXPLANATIONS

Sufficient facts or data during testimony?

Reliably applied during testimony (no quality process)?

- It is a direct comment on what happened (we do not know this…)
- Possible = speculative, has no inherent value
- It is very difficult to justify "possible" with facts or data – if you could- then you should be doing an evaluation that considers "how possible" – <u>with probabilities</u>
- Explanations = justifying your results once you know them (this is not great logic – anything will work here!)
- May mislead the jury about the strength of the DNA results (over or under value)
- We have aimed to **<u>stay in our lane</u>** of expertise, these questions require a different level of training and authorization

Communicating Forensic Findings: Current Practices and Future Directions Workshop

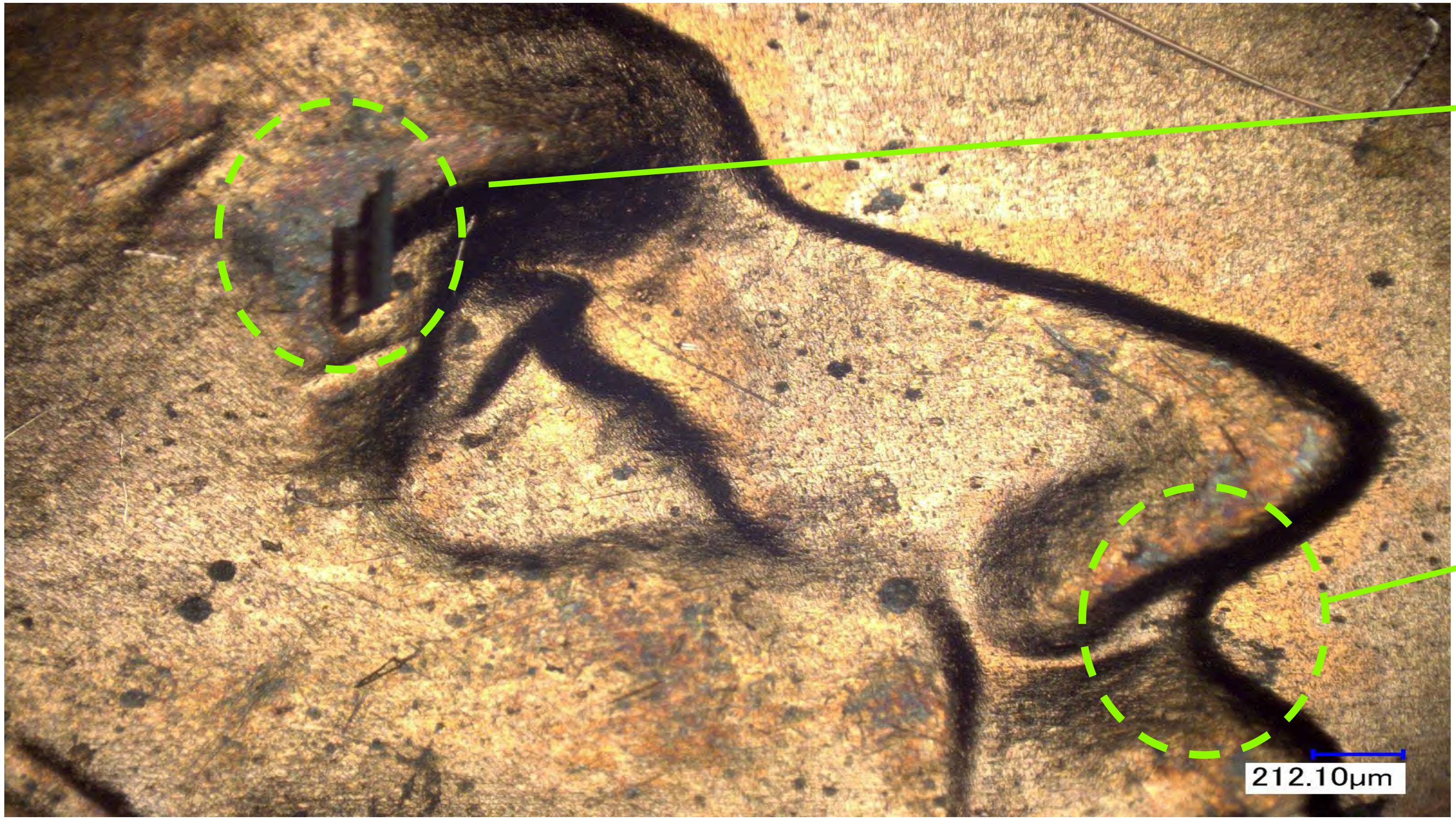
Trace Evidence Perspective on Interpretation Scales

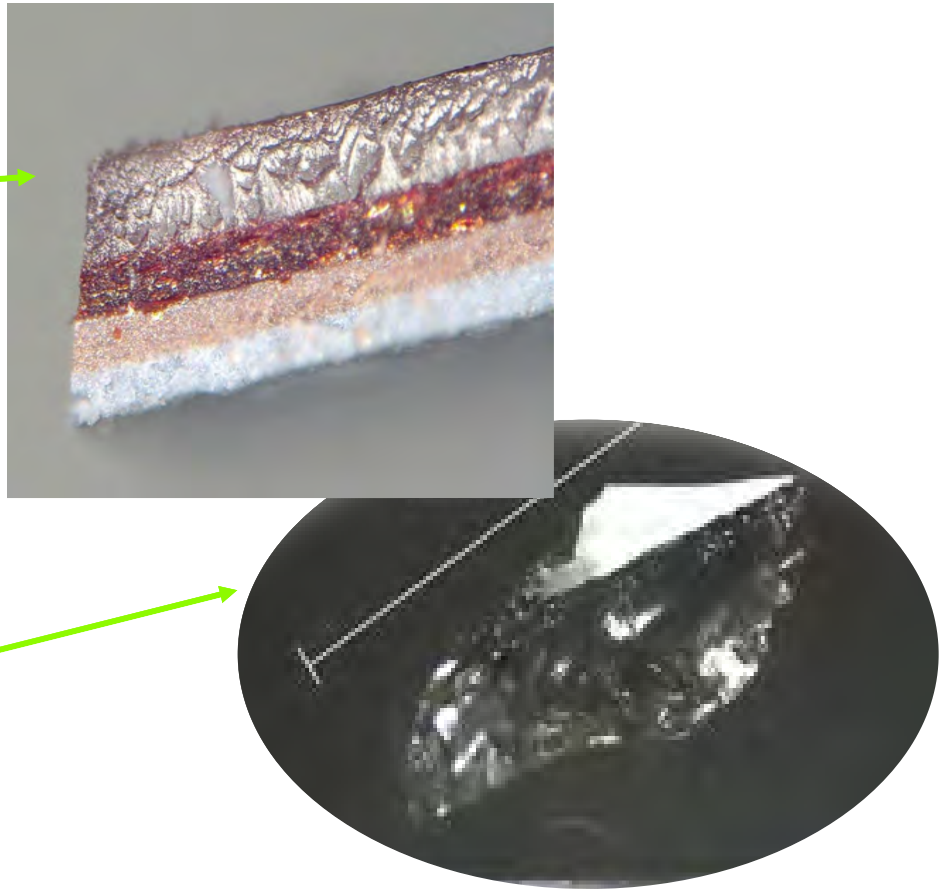
Tatiana Trejos

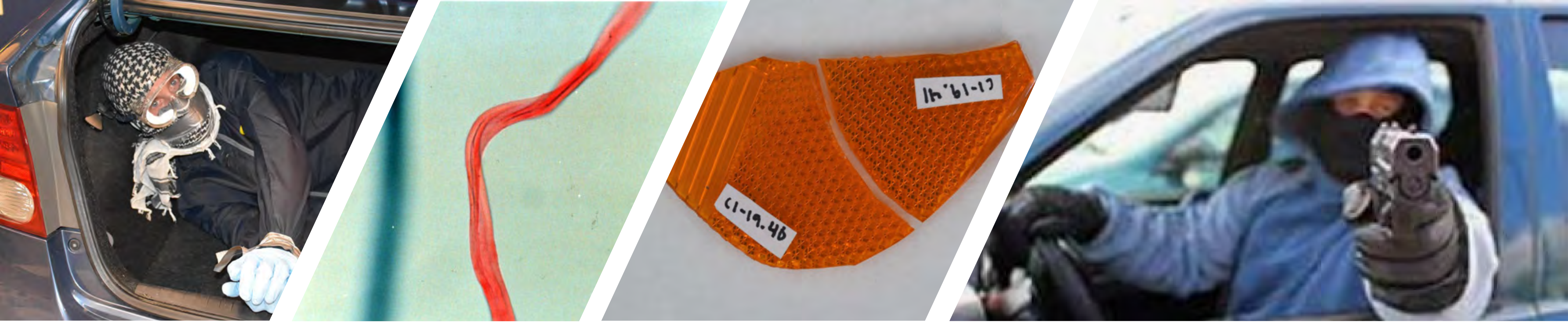
West Virginia University, Department of Forensic and Investigative Science

June 25-26, 2024; Rockville, Maryland

TRACE evidence: invisible clues that tell a story…





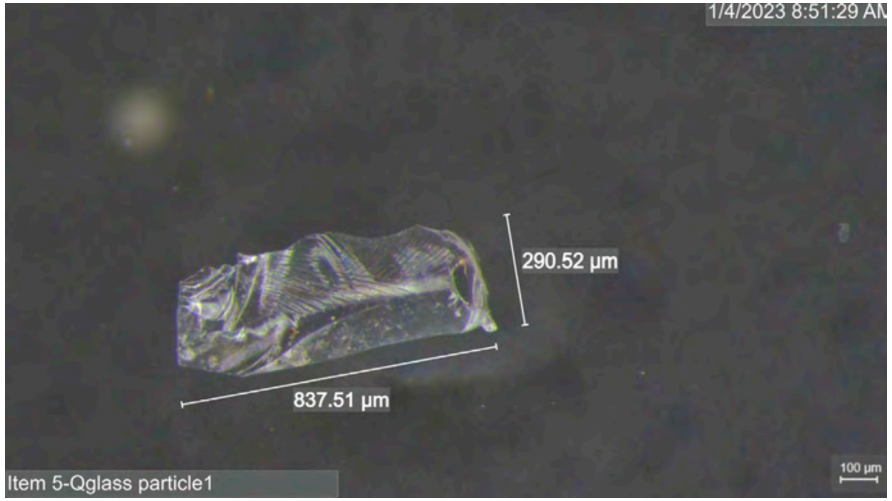


Links between objects can answer questions about when, how, where?

Images:

- https://depositphotos.com/stock-photos/car-crash.html
- https://www.la-criminaldefense.com/drive-by-shooting-murder-gang-defense-in-california/

Hit and run case example



Courtesy of Troy Ernst, Trace Evidence Unit, Forensic Science Division, Michigan State Police

Not too long ago…

“The glass fragment recovered from the jacket (item #5) **could have come from** the glass submitted as reference (mirror item #6)”

Initiatives in Trace Evidence: Interpretation Scale Development

1

2006

ORIGIN: Michigan State Police, Chris Bommarito

2

2008-2014

SWGMAT
Some versions implemented at forensic laboratories (as seen in proficiency test results or surveys)

3

2014-to date

NIST-OSAC Trace Subcommittee, Interpretation task group

OSAC 2022-S-0029 Standard Guide for Interpretation and Reporting in Forensic Comparisons of Trace Materials

4

Standard Guide: SUBMITTED TO ASTM

Interpretation & Reporting Guide

Interpretation Task Group Members, past and present

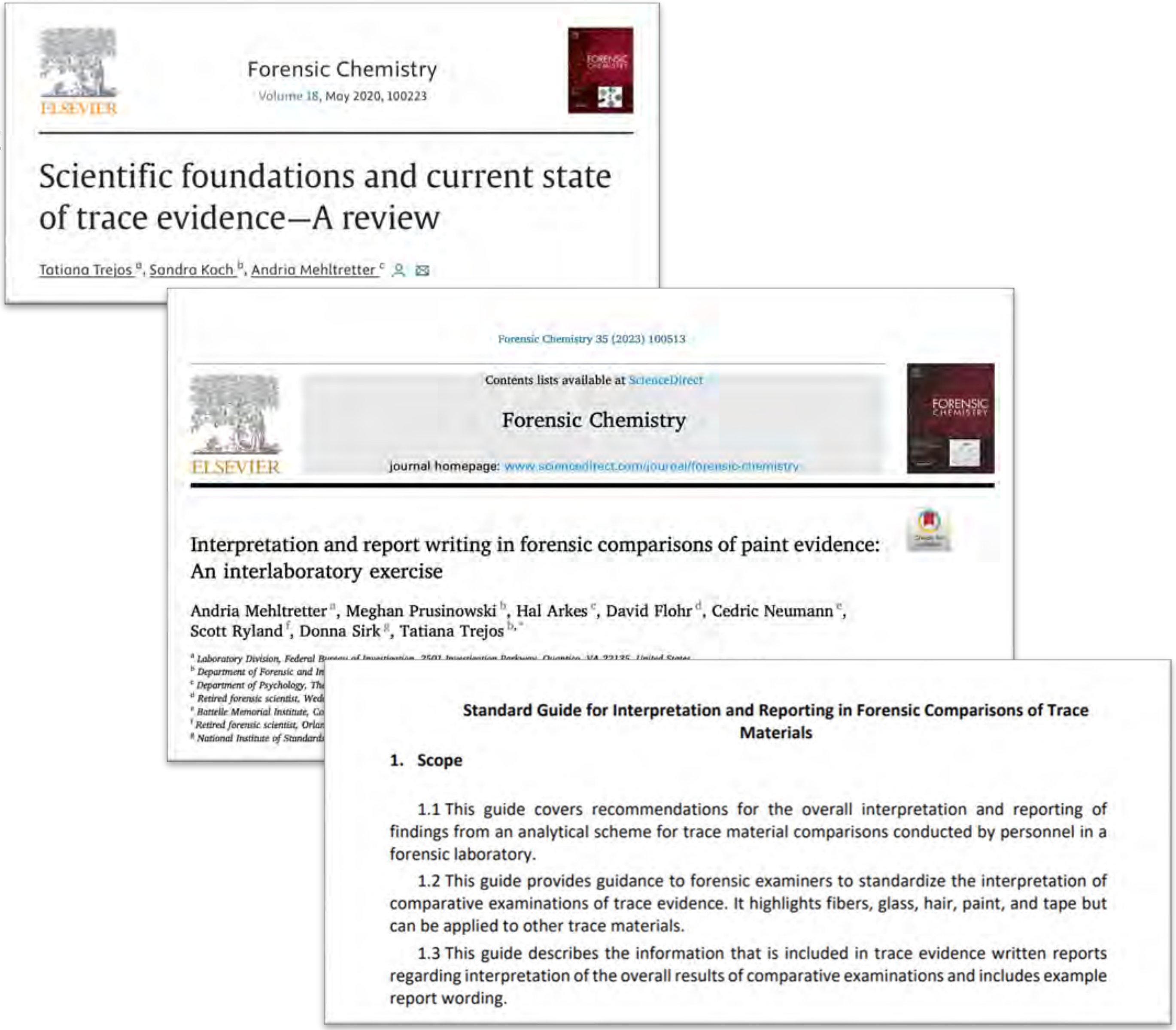
- Cathy Brown, Collaborative Testing Services (current chair)
- Mary Eng, New York City Police Crime Lab
- David Green, Lake County Crime Lab (Ohio)
- Susan Gross, Minnesota Bureau of Criminal Apprehension, ATF
- Tammy Jergovich, Georgia Bureau of Investigation
- Cheryl Lozen, Michigan State Police, retired
- Andria Mehltretter, FBI (past chair)
- Tatiana Trejos, West Virginia University

Statisticias and human factors (Hal Arkes, Cedric Newmann, Madeline Ausdemore , Shirly Montero

NIST-OSAC Trace/Materials Subcommittee

Revisions at Subcommittee, STRP panel, Legal & Human Factors, FSSB

Hundreds of reviewers: OSAC and public



Forensic Chemistry
Volume 18, May 2020, 100223

Scientific foundations and current state of trace evidence—A review

Tatiana Trejos [a], Sandra Koch [b], Andria Mehltretter [c]

Forensic Chemistry 35 (2023) 100513

Contents lists available at ScienceDirect

Forensic Chemistry

journal homepage: www.sciencedirect.com/journal/forensic-chemistry

Interpretation and report writing in forensic comparisons of paint evidence: An interlaboratory exercise

Andria Mehltretter [a], Meghan Prusinowski [b], Hal Arkes [c], David Flohr [d], Cedric Neumann [e], Scott Ryland [f], Donna Sirk [g], Tatiana Trejos [b,*]

Standard Guide for Interpretation and Reporting in Forensic Comparisons of Trace Materials

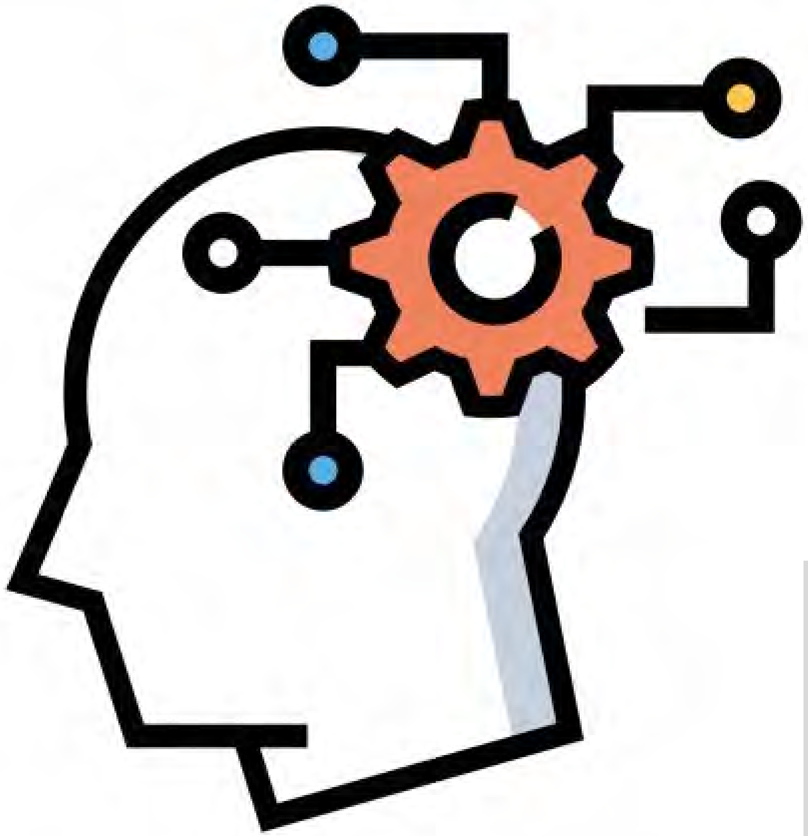
1. Scope

1.1 This guide covers recommendations for the overall interpretation and reporting of findings from an analytical scheme for trace material comparisons conducted by personnel in a forensic laboratory.

1.2 This guide provides guidance to forensic examiners to standardize the interpretation of comparative examinations of trace evidence. It highlights fibers, glass, hair, paint, and tape but can be applied to other trace materials.

1.3 This guide describes the information that is included in trace evidence written reports regarding interpretation of the overall results of comparative examinations and includes example report wording.

Interpretation Guide



Simple Interpretation Process for Comparative Examinations

- 5 main subdisciplines: fibers, hair, glass, paint, and tape.
- Can the compared items be discriminated?
- Evaluation of the results on a source level. Common source? Significance of the finding?
- Evaluation of these results considering various factors
- Also compatible with LR

✓ Uses a scale to assess and report the significance of the findings

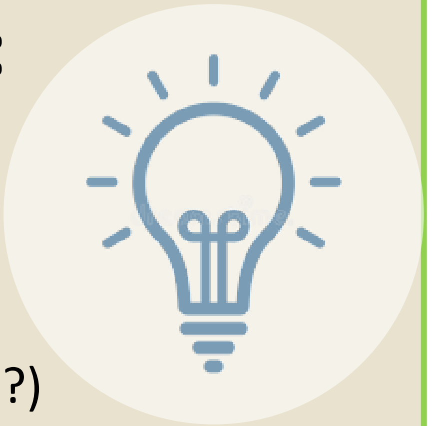
✓ Provides material-specific interpretation criteria and casework-based reporting examples

✓ Universal and flexible platform can use scientifically sound qualitative or quantitative inputs for decision-making

The Core of the Guide:
Interpretation categories based on systematic approaches and consensus criteria

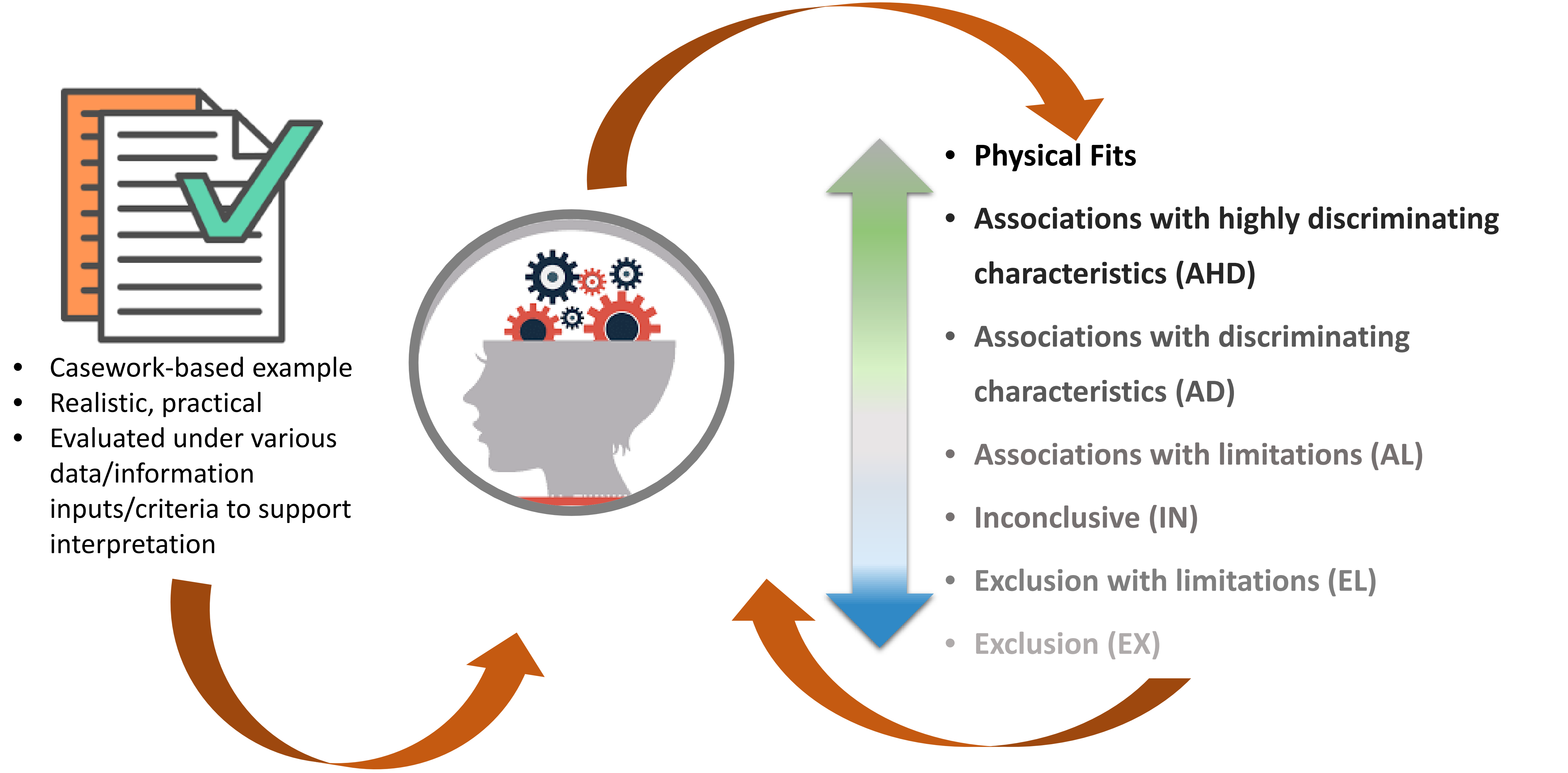
SYSTEMATIC CRITERIA

Interpretation based on:

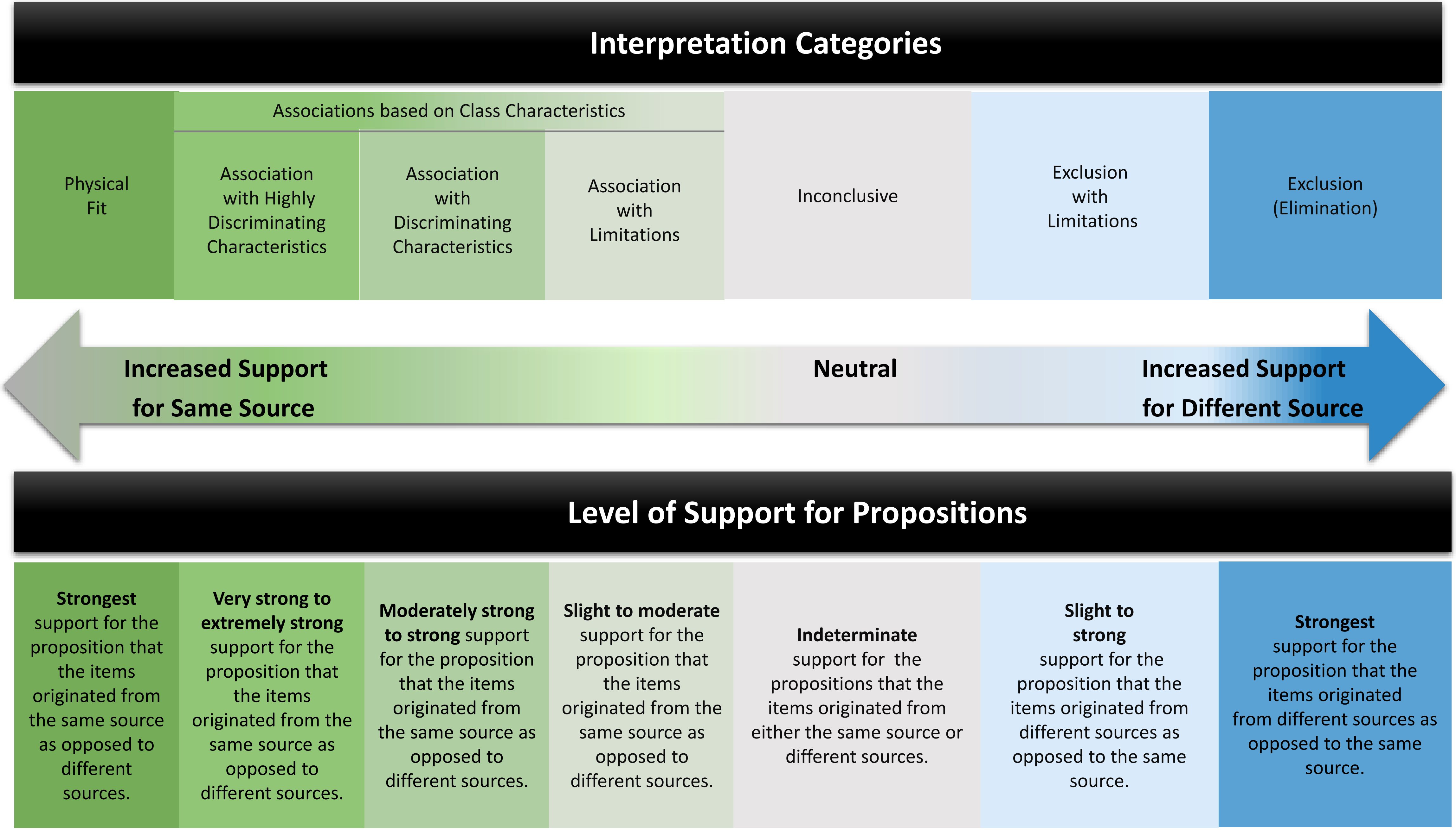


- ✓ Scientific foundations
- ✓ Analysis and Data Interpretation
- ✓ Rarity assessment
- ✓ Contextual relevance
- ✓ Population studies (what is out there?)
- ✓ Manufacturing and distribution information
- ✓ Discrimination studies
- ✓ Practitioner training and experience
- ✓ Case studies
- ✓ Databases and collections

Criteria developed by material supported by hundreds of scientific literature.



Casework-based example
Realistic, practical
Evaluated under various data/information inputs/criteria to support interpretation
Physical Fits
Associations with highly discriminating characteristics (AHD)
Associations with discriminating characteristics (AD)
Associations with limitations (AL)
Inconclusive (IN)
Exclusion with limitations (EL)
Exclusion (EX)


Interpretation Categories
Associations based on Class Characteristics
Physical Fit
Association with Highly Discriminating Characteristics
Association with Discriminating Characteristics
Association with Limitations
Inconclusive
Exclusion with Limitations
Exclusion (Elimination)
Increased Support for Same Source
Neutral
Increased Support for Different Source
Level of Support for Propositions
Strongest support for the proposition that the items originated from the same source as opposed to different sources.
Very strong to extremely strong support for the proposition that the items originated from the same source as opposed to different sources.
Moderately strong to strong support for the proposition that the items originated from the same source as opposed to different sources.
Slight to moderate support for the proposition that the items originated from the same source as opposed to different sources.
Indeterminate support for the propositions that the items originated from either the same source or different sources.
Slight to strong support for the proposition that the items originated from different sources as opposed to the same source.
Strongest support for the proposition that the items originated from different sources as opposed to the same source.

Exclusion

- **Exclusion (Elimination)** – The items exhibit differences that provide the strongest support that the items originated from different sources as opposed to the same source.

  Exclusion with Limitations – The item exhibits differences from the comparison sample that support that it did not originate from the source, as represented by the comparison sample; however, limiting factors prevented an Exclusion (Elimination) from being reached. This provides slight to strong support for the proposition that the items originated from different sources as opposed to the same source.

Arson- Molotov cocktail fabric: Exclusion

Courtesy of Troy Ernst, MSP.

Physical fit

Physical Fit is the highest degree of association between items. It is the opinion that the observations provide the strongest support for the proposition that the items were once joined together to form a single object as opposed to originating from different sources. *Physical Fit* is reached when the items that have been broken, torn, or separated exhibit physical features that correspond or re-align in a manner that is not expected to be replicated. A Physical Fit is not currently based upon a statistical evaluation of data; it is also not based upon exhaustive comparisons to all potential sources.

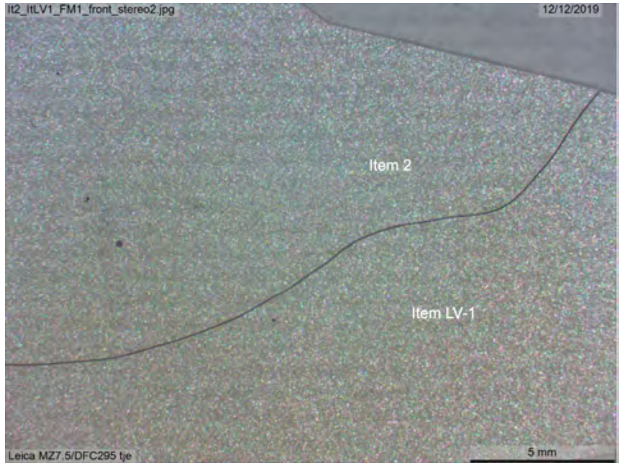
Physical Fit: Fatal hit and run case- MSP

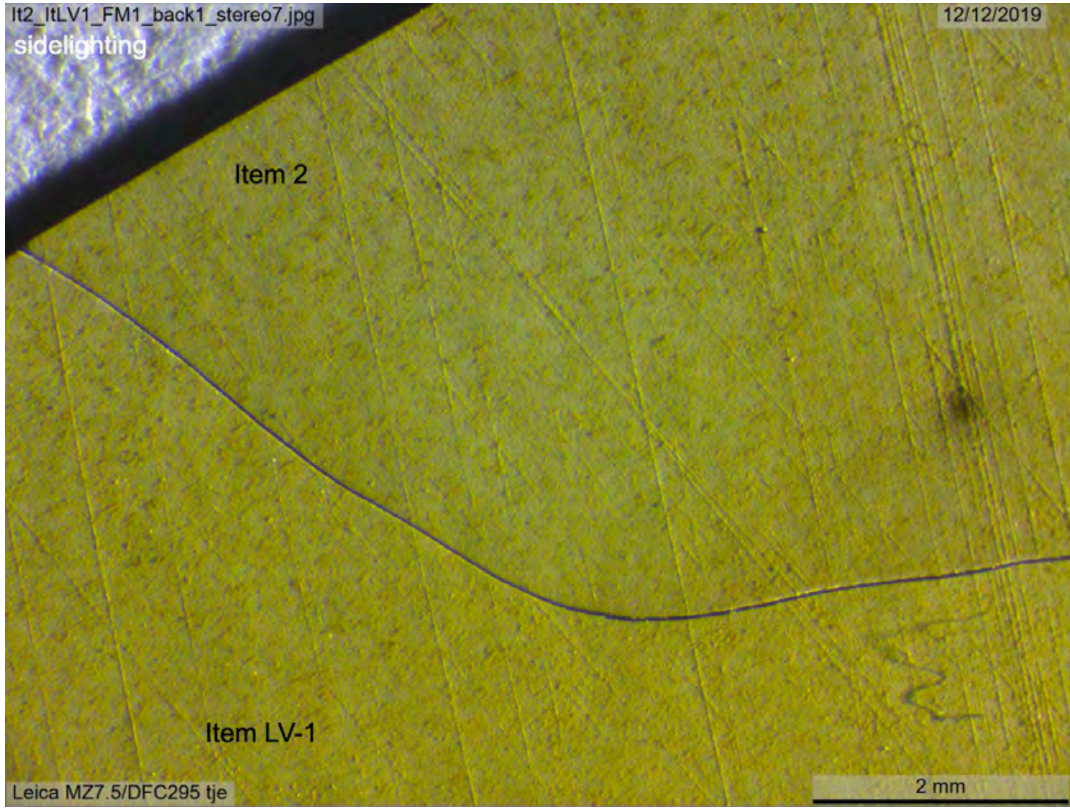
Police find vehicle parts near Norton Shores hit-and-run victim

By: FOX 17 News

NORTON SHORES, Mich. — Police are trying to find the driver responsible for a hit-and-run that left a 53-year-old woman dead on the side of a road.

The incident happened on Seaway Dr. between W. Norton Ave. and Seminole Rd. at about 12:44 a.m. Sunday morning, according to Norton Shores Police Department. Medical crews determined the Muskegon





Disclaimer: Courtesy of Troy Ernst, MSP. Conclusions are redacted to illustrate the scale and not verbatim of his report

Associations of Evidence based on Class Characteristics

Class characteristics are physical, optical, or chemical properties that establish membership in a group. Associations based on class characteristics do not establish that the items came from the same source. Class associations can have varying degrees of significance. In general, the smaller the size of the group relative to the relevant population, the more significant the association. These types of associations are categorized as follows:

- Association with Highly Discriminating Characteristics
- Association with Discriminating Characteristics
- Association with Limitations

Association- glass example

Association with Highly Discriminating Characteristics

The items share unusual characteristics that would rarely occur in the relevant population.

- Association of glass fragments characterized by elemental analysis using **ICP-based** methods.
- Association of glass fragments characterized by **RI and elemental analysis using µXRF when Sr, Zr, or an element that is less commonly or rarely detected** in glass by XRF is used in element intensity ratio comparisons.
- Association of glass fragments for which the estimated **random match probability of the measured properties is very small (i.e., smaller than 0.2%)**
- Association of glass fragments for which the estimated calibrated likelihood ratios (LR) provide very strong to extremely strong support for the same-source hypothesis over the different-source hypothesis (e.g., **LR greater than 1000**)

Association with Discriminating Characteristics

Other items have been manufactured or could occur in nature that would also be indistinguishable from the submitted items and could be encountered in the relevant population

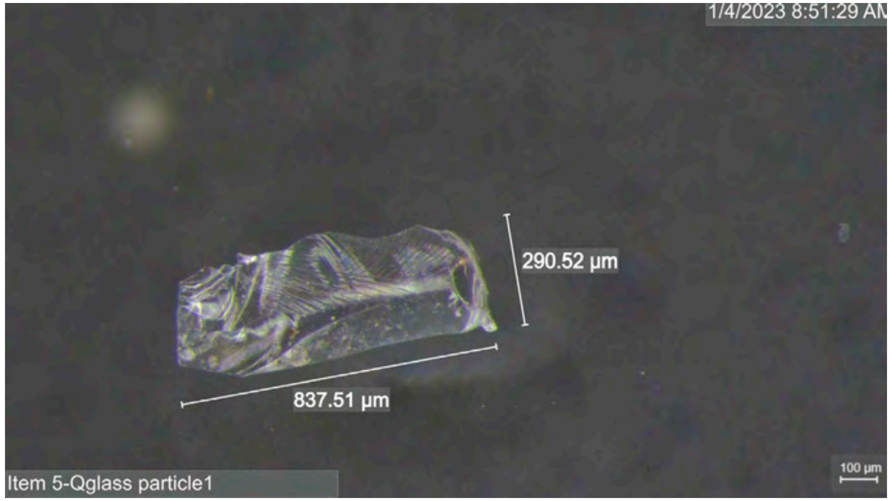
Association of glass fragments characterized by **elemental analysis using µXRF alone**, when Sr, Zr, or an element that is less commonly or rarely detected in glass by is used in element intensity ratio comparisons.

Association of glass fragments characterized by elemental analysis using RI and µXRF, when Sr, Zr, and all elements that are less commonly or rarely detected in glass by XRF are below the limit of quantitation.

Association of glass fragments for which the estimated **random match probability of the measured properties is small (e.g., between 0.2% and 2%)**

Association of glass fragments for which the estimated calibrated likelihood ratios (LR) provide moderately strong to strong support for the same-source hypothesis over the different-source hypothesis (e.g., **LR between 100 and 1000**)

GLASS: Fatal Hit and Run Case Example

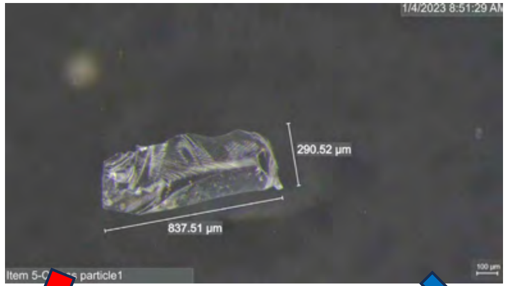


Courtesy of Troy Ernst, Trace Evidence Unit, Forensic Science Division, Michigan State Police

Hit and Run Reporting Example

- **Association with Discriminating characteristics**
- The questioned glass fragment (Item 5-Qglass) and the known glass from the large mirror of Item 2 corresponded in general appearance, refractive index, and elemental composition by µXRF.
- In the opinion of the examiner, Item 5-Qglass originated either from the large mirror of Item 2 or from another broken glass source with indistinguishable properties.(**Association with Discriminating Characteristics**). This type of association was reached because coincidental associations of glass originating from different sources could occur but are expected to be unusual.

- **Exclusion/Elimination**
- The questioned glass fragment (Item 5-Qglass) differed in elemental composition from the known glass from the small mirror of Item 2. In the opinion of the examiner, Item 5-Qglass did not originate from the small mirror of Item 2 (Elimination)



Association

Elimination

Disclaimer: Courtesy of Troy Ernst, MSPD. Conclusions are redacted to illustrate the interpretation scale and not verbatim of his report

Association- paint example

Association with Highly Discriminating Characteristics

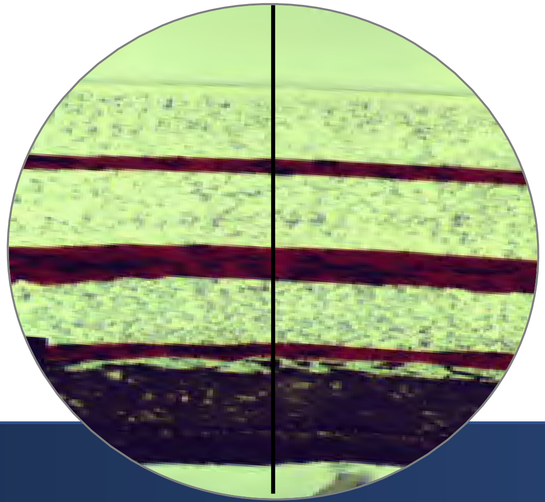
The items share unusual characteristics that would rarely occur in the relevant population.

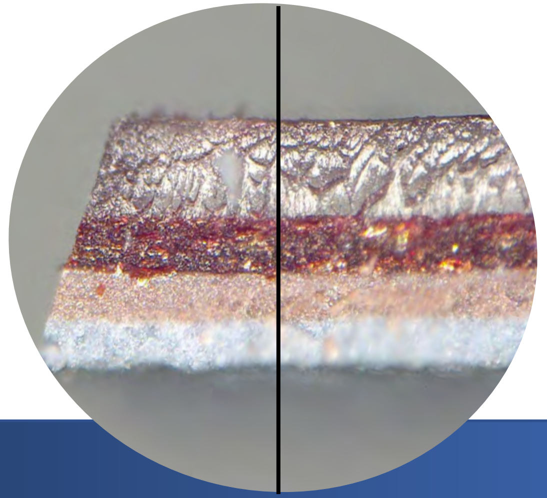
- OEM automotive system with **at least one aftermarket** basecoat or primer layer above the original clear coat.
- OEM automotive system with **two or more factory repairs** (i.e., three or more total basecoat-clearcoat sequences).
- **Architectur**al paint system with **two or more different layers**.
- Automotive system with architectural paint present.

Association with Discriminating Characteristics

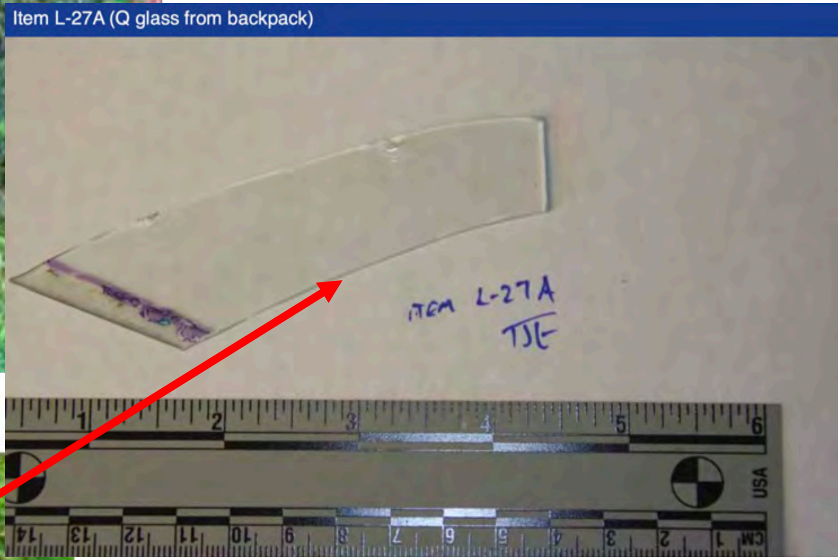
Other items have been manufactured or could occur in nature that would also be indistinguishable from the submitted items and could be encountered in the relevant population

- Association of paint in which the typical analysis scheme was performed on mass- produced materials that have numerous features for evaluation (e.g., four-layered OEM automotive paint).
- **OEM automotive paint** system with one factory repair of the same basecoat color and layer sequence (i.e., two total OEM basecoat-clearcoat sequences).
- Single-layered **paint for which there is knowledge of substantial discrimination power** (e.g., red architectural paint) or product manufacturing distribution information that reduces the potential sources.
- **Aftermarket refinish clearcoat and basecoat**

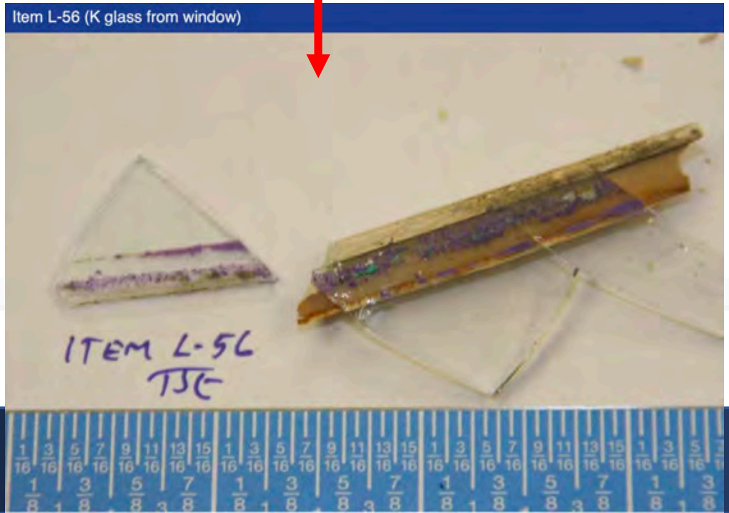




GLASS and PAINT Example - Homicide

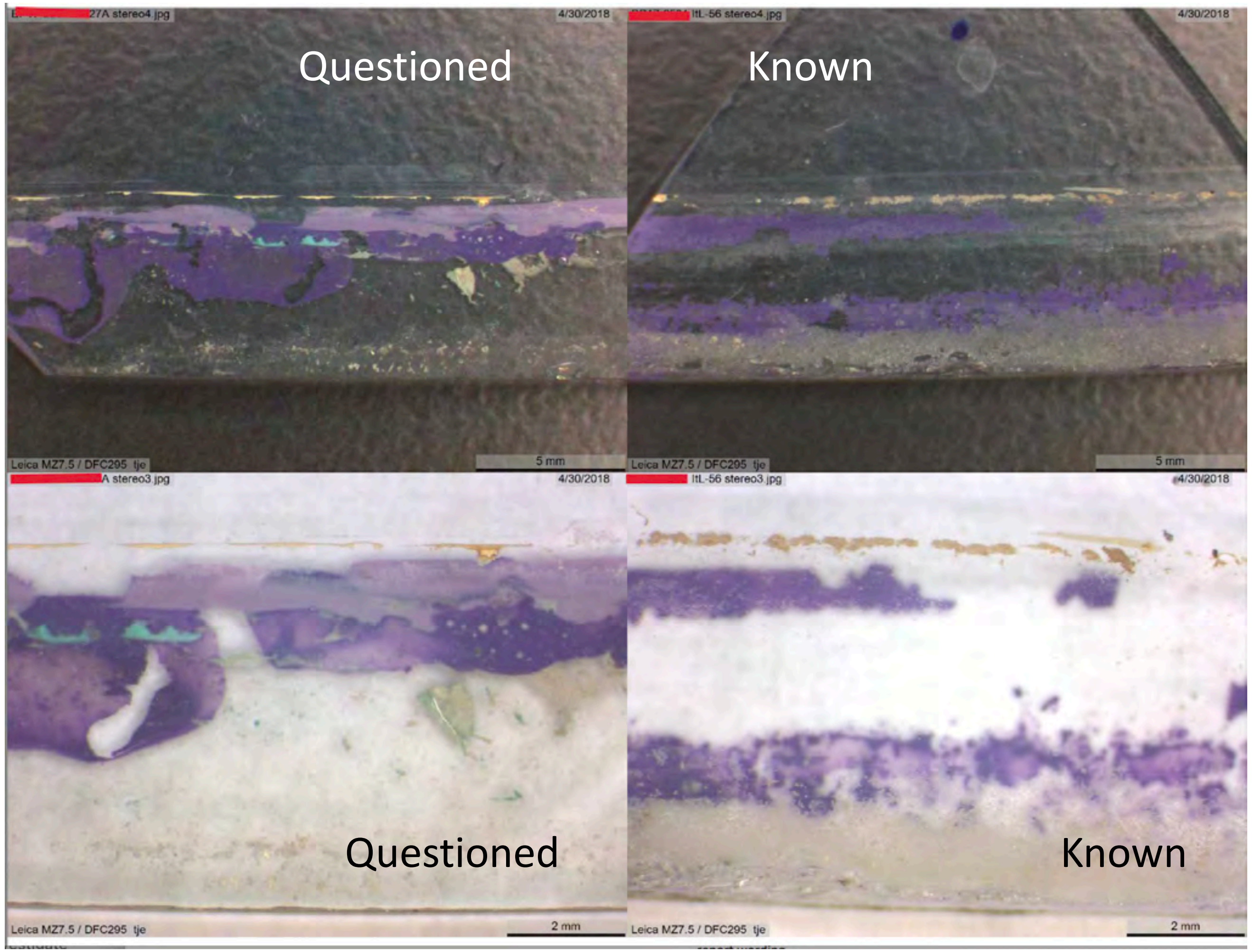


Known K
L-56

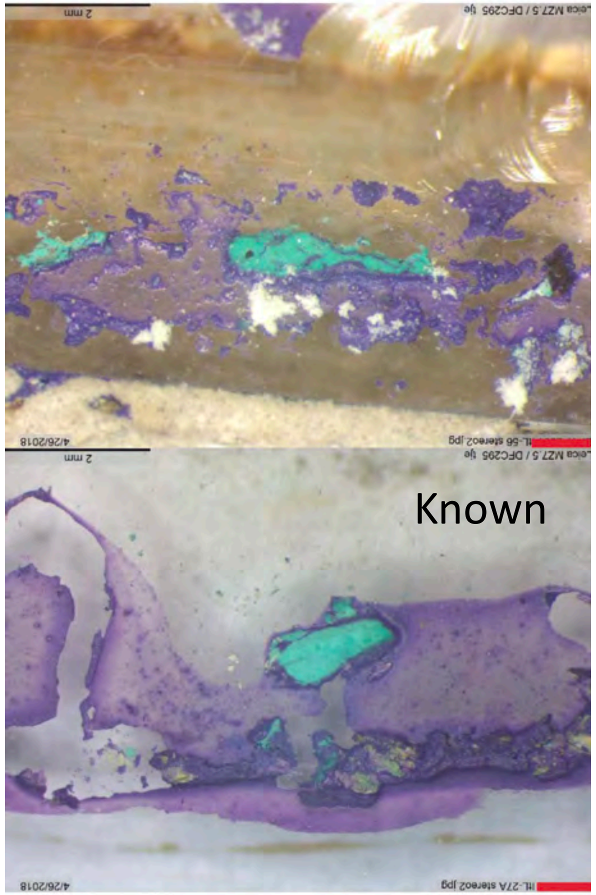


Questioned Q
L-527A

Courtesy of Troy Ernst, Trace Evidence Unit, Forensic Science Division, Michigan State Police



Questioned



Glass and Paint Reporting Example

- **Glass Association with Highly Discriminating Characteristics**
- The submitted questioned glass fragment (Item BP17-2581-L27A) and known glass fragments (Item L-56) were compared using physical characteristics, refractive index measurements, and elemental analysis by x-ray fluorescence (XRF).
- The questioned glass fragment was similar in color, thickness, type (float, non-tempered), refractive index, and elemental composition to the known glass. Additionally, there were corresponding colors (purple and turquoise) and location of apparent paint on the surfaces of the questioned and known glass samples, and corresponding color and location of apparent caulk on the surfaces opposite the paint of the questioned and known glass samples.
- **Therefore, the questioned glass originated from the broken window as represented by the known sample or another source of broken glass indistinguishable in the measured properties (Association with Highly Discriminating Characteristics). This type of association was reached because coincidental associations of glass originating from different sources could occur but are expected to be highly unusual.**
- **The presence of corresponding multiple colors of paint and of apparent caulk on both items greatly increases the significance of this association.**

- **Paint Association with Highly Discriminating Characteristics**
- Examination and comparison of the questioned paint with known sample, revealed they are consistent with respect to their observed and measured physical and chemical properties (e.g., architectural paint with two colors, purple and turquoise paint). It is therefore concluded that the Item questioned paint recovered from the glass fragment in the bag pack corresponds to the known item paint and therefore originated either from that window or from another window with architectural paint having the same distinct characteristics (Association with Highly Discriminating Characteristics).
- **The presence of corresponding multiple colors of paint and of apparent caulk on both broken glass items greatly increases the significance of this association.**

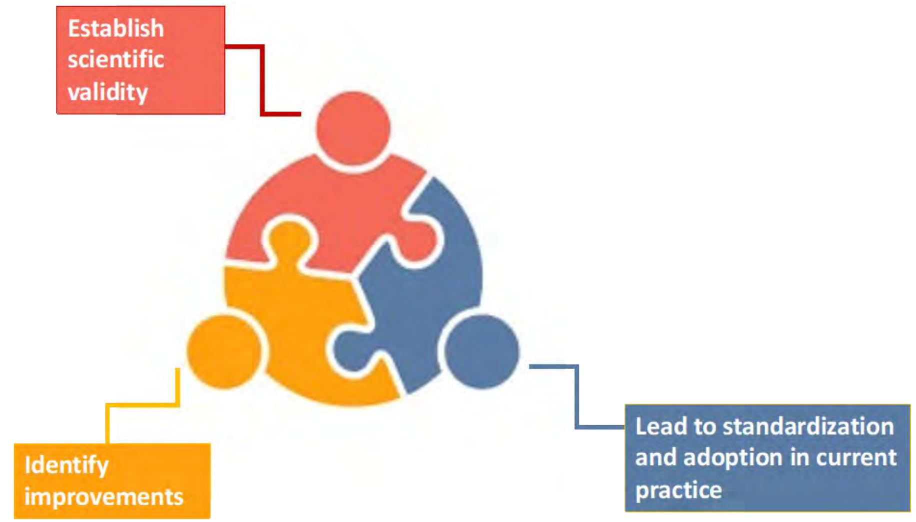
Disclaimer: Courtesy of Troy Ernst, MSP. Conclusions are redacted to illustrate the scale and not verbatim of his report

How trace evidence strengthen links!

Courtesy of Troy Ernst, Trace Evidence Unit, Forensic Science Division, Michigan State Police

Interlaboratory Study

- **Five possible levels of interpretation**
- Two difficulty levels
- 80 scenarios independently developed and evaluated by “**Subject Matter Expert Panel (SMEP**)” best 30 chosen for study
- 30 “SMEP consensus ground truth” scenarios
  - 15 scenarios randomly distributed to each participant
  - Overall, designed to have same number of total responses per scenario
  - **85 participants, 1267 responses**



Lessons learned from Paint Interpretation ILS

- These findings demonstrate that a **high level of agreement was achieved** among practitioners regarding the significance of results in comparative examinations when using the proposed guide.
- High **agreement between consensus reached by SMEP and within participants** (**93%** of the case scenarios, 28 out of 30)
- This exercise provided a **tangible means to assess the thinking process** of the participants in interpreting the results. The scale, criteria, and examples in the document aid in standardizing the interpretation process.

Acknowledgements

- Cedric Neumann, statistical support
- Hal Arkes, Emeritus Professor, Department of Psychology, The Ohio State University
- Donna Sirk, NIST-OSAC Program Office
- Scott Ryland, retired forensic scientist
- David Flohr, retired forensic scientist
- Meghan Prusinowski, Korina Menking Hoggatt, former graduate students, West Virginia University
- NIST-OSAC Trace/Materials Subcommittee Interpretation and Physical Fit Task Group Members
- Interlaboratory study participants

Questions?

Tatiana.Trejos@mail.wvu.edu

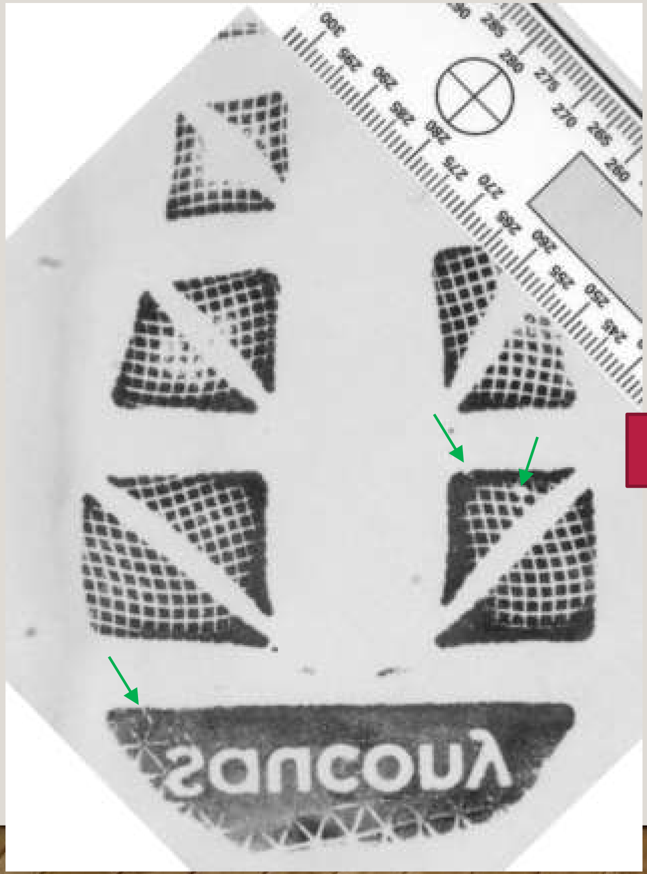
SCIENCE COMMUNICATION WORKSHOP
JUNE 2024

PERSPECTIVES FROM THE FOOTWEAR PRACTITIONER COMMUNITY

DAVID KANARIS
CHIEF, ALASKA SCIENTIFIC CRIME DETECTION LABORATORY
CHAIR, OSAC FOOTWEAR AND TIRE SUBCOMMITTEE

WHAT DOES A FOOTWEAR COMPARISON ENTAIL?

Crime Scene Impression



Superimposed Composite Image

Known Shoe Test Impression

CURRENT CONCLUSION SCALE

ARTICULATING OPINIONS INTO THE FUTURE

- Goals were:
  - To move away from categorical conclusions
  - Decide on a framework for articulating opinions to the stakeholders that were transparent, more justifiable, and didn't lose anything in terms of how understandable they were
- Also wanted to provide something that was usable to practitioners and adoptable by laboratories. To do this we had to answer two questions

1. Are the new opinions being understood?
2. Are the new opinions perceived as much "weaker" than previous associations?

QUESTION 1 – HOW TO IDENTIFY THE BEST FRAMEWORK?

- Weight of evidence type scale
  - Where the practitioner must consider the two propositions during the evaluation

QUESTION 2 – HOW MANY BOXES IS TOO MANY BOXES?

QUESTION 3 – HOW IMPORTANT IS SYMMETRY IN A SCALE?

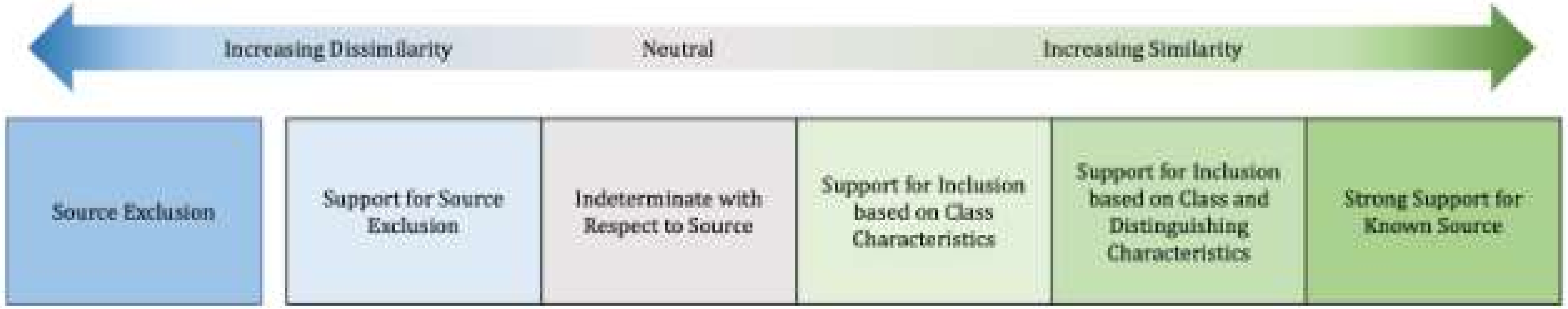
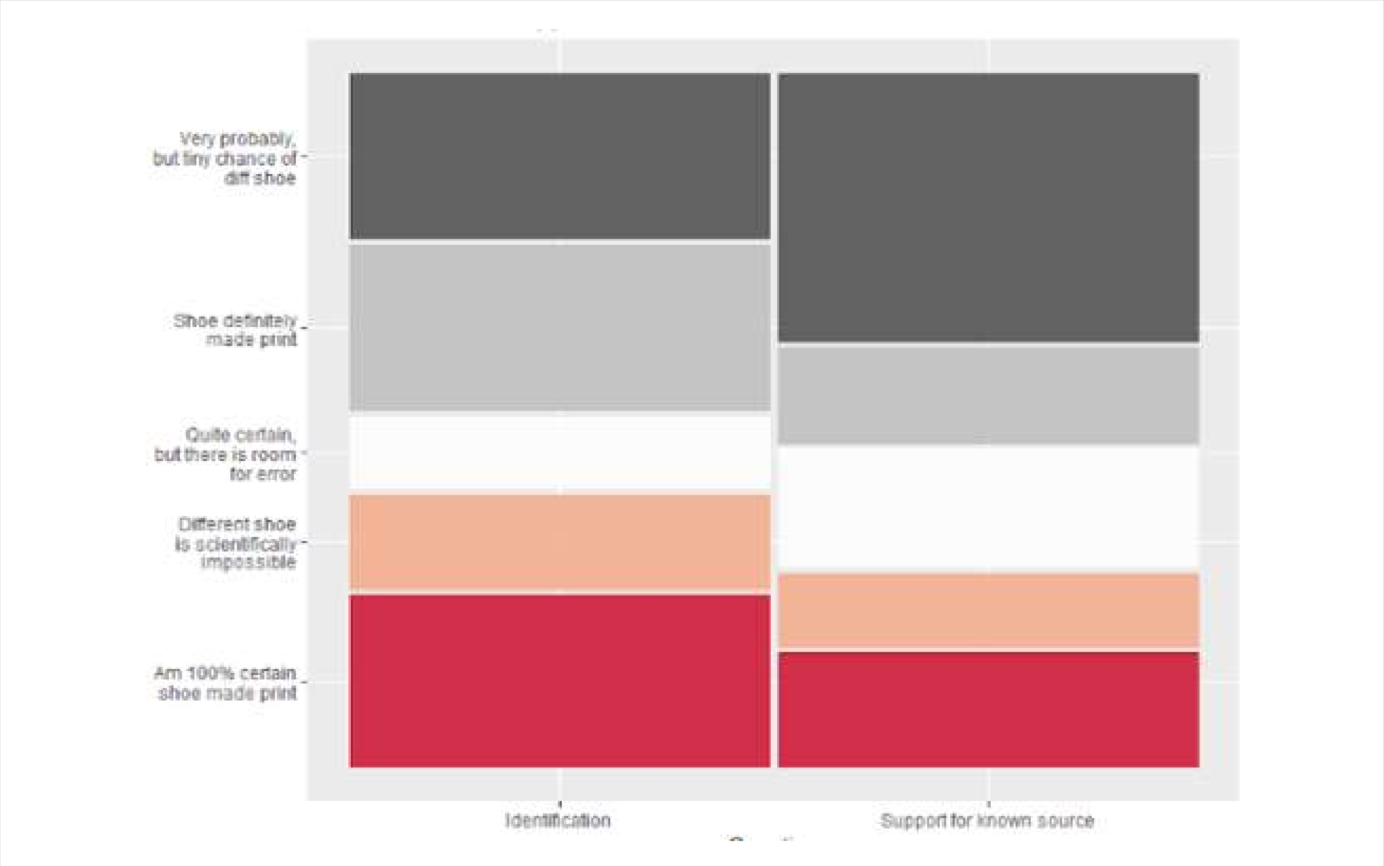


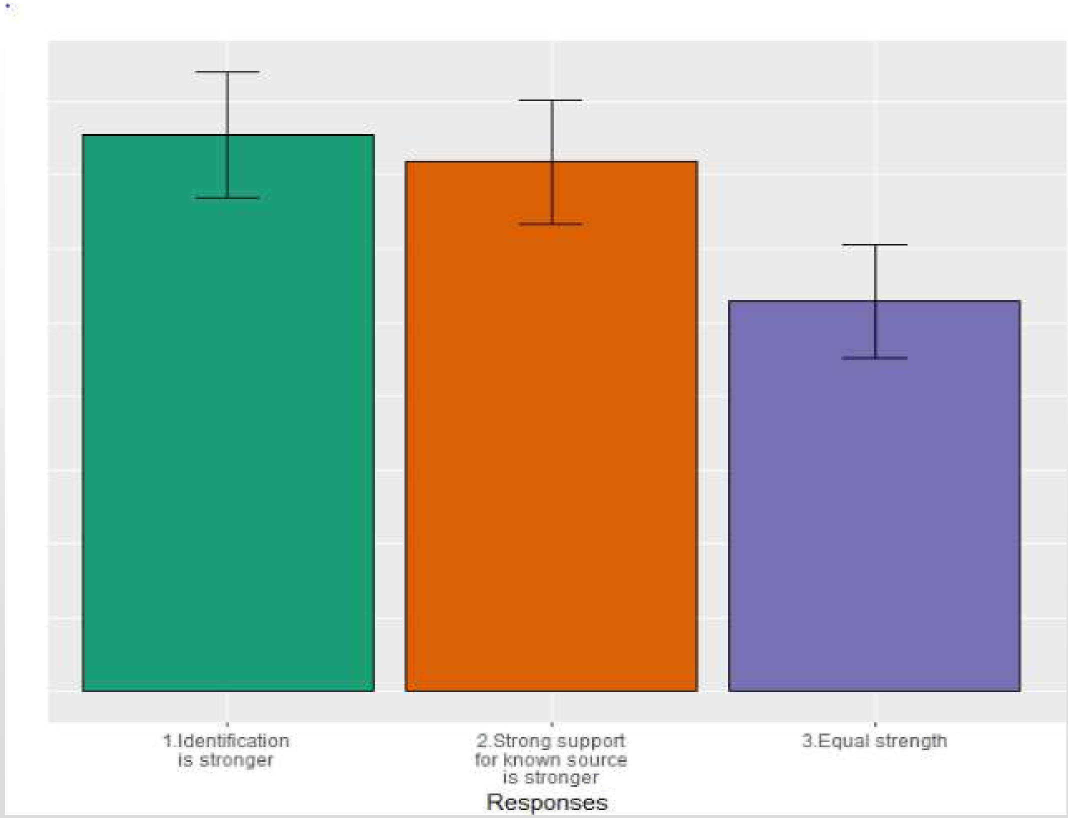
Figure 1. Articulation of source opinions for footwear and tire interpretations. The source opinion categories are displayed as concrete bins across a continuum of observed similarity/dissimilarity.

NEW SCALE



QUESTION 4 – ARE WE UNDERSTOOD?

QUESTION 5 – ARE WE CORRECTLY REPRESENTING THE STRENGTH OF SUPPORT?



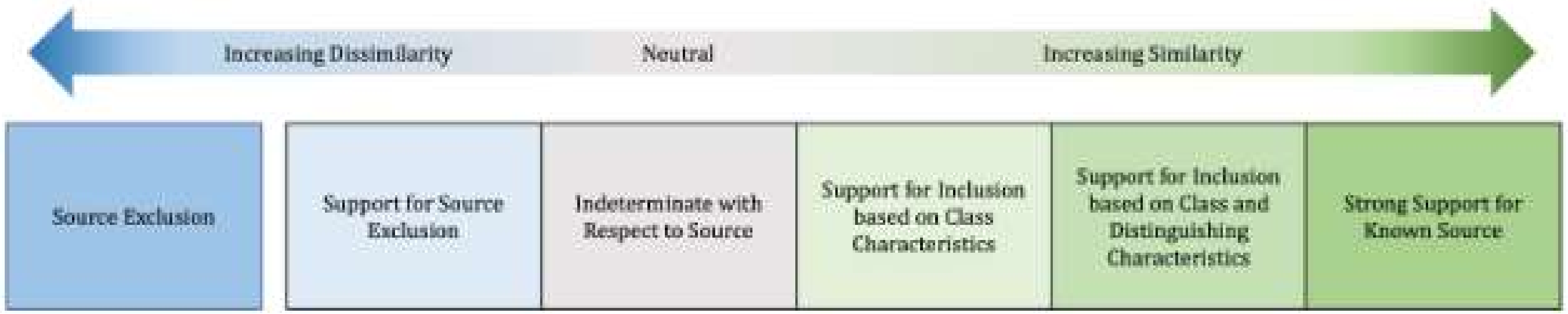


Figure 1. Articulation of source opinions for footwear and tire interpretations. The source opinion categories are displayed as concrete bins across a continuum of observed similarity/dissimilarity.

NEW SCALE

THANK YOU

- David Kanaris
- david.kanaris@alaska.gov
- Footwear & Tire Subcommittee | NIST

DM SAC Standard Guide for Image Comparison Opinions

Presented to

Communicating Forensic Findings Workshop: Current Practices and Future Directions

June 25, 2024

By Lora Sims, Director Face Center of Excellence (FaCE)

Disciplines

- Framework for discipline specific opinion scale standard.
  - Encompasses Facial and Iris Identification (FII) and the Video/Imaging Technology and Analysis (VITAL)
- Specific to comparisons of people, objects, or scenes captured in images

Disciplines

- Images of vehicles from SWGDE Technical Overview for Forensic Image Comparison v1.0
- Images of faces from Ideal Innovations, Inc. FaCE Training Database

History & Limitations of Facial Identification

- Started in law enforcement in 1840's
- Has expanded to the following applications:
  - Screening and Access Control
  - Investigative and Operational Leads
  - Intelligence Gathering for Identity Management
  - Forensic Comparison
- Limitations:
  - Image quality/resolution/size
  - Lighting conditions
  - Distance of camera to subject
  - Image formats/compression
  - Expression/Aging/Weight Changes/Health Changes/Intentional Modifications
  - Comparisons - subjective/opinion based

Historical Scales used in Facial Identification (1 of 3)

- Screening and Access Control
  - No formalized scale used
- Investigative and Operational Leads
  - Investigative Lead
  - Inconclusive
  - Not the same person

These are examples from some agencies and were not standardized across the discipline.

Historical Scales used in Facial Identification (2 of 3)

- Intelligence Gathering for Identity Management
  - Example 1:
    - Identification (match)
    - Likely
    - Inconclusive
    - Unlikely
    - Non-Identification (non-match)
  - Example 2:
    - +3 Conclusion that subjects are almost certainly the same person
    - +2 Conclusion that subjects are likely the same person
    - +1 Observation of positive indications between subjects
    - 0 Determined this comparison to be inconclusive
    - -1 Observation of negative indications between subjects
    - -2 Conclusion that subjects are likely not the same person
    - -3 Conclusion that subjects are almost certainly not the same person

These are examples from some agencies and were not standardized across the discipline.

Historical Scales used in Facial Identification (3 of 3)

- Forensic Examinations (7-point scale)
  - Example 1:
    - +3 – Extremely strong support for Identification (i.e., “Identification” or “Appears to be”)
    - +2 – Strong support for Identification, with little or no support for Exclusion
    - +1 – Some or Limited support for Identification, with less support for Exclusion
    - 0 – No Conclusion [The subject cannot be differentiated from a large segment of the population.]
    - -1 – Some or Limited support for Exclusion, with less support for Identification
    - -2 – Strong support for Exclusion, with little or no support for Identification
    - -3 – Extremely strong support for Exclusion (i.e., “Elimination”)
  - Example 2:
    - +3: The observations strongly support that it is the same person
    - +2: The observations support that it is the same person
    - +1: The observations support to some extent that it is the same person
    - 0: The observations support neither that it is the same person nor that it is different persons
    - -1: The observations support to some extent that it is not the same person
    - -2: The observations support that it is not the same person
    - -3: The observations strongly support that it is not the same person

These are examples from some agencies and were not standardized across the discipline.

History & Limitations of VITAL

- VITAL – started to become used in the 1990’s
- Limitations:
  - Image quality/resolution/size
  - Lighting conditions
  - Distance of camera to subject
  - Image formats/compression
  - Comparisons - subjective/opinion based
  - Authentication - increases in technology (automated tools, artificial intelligence)

Historical Scales used in VITAL

Continuum of Conclusions Examples For Photographic Comparative Analysis			
	Identification	Identification	Identification
			Powerful support same
	Similar	Similarities noted	Strong support same
	No conclusion, but with similarities		Moderate support same
			Limited support same
		Neither/Nor – with explanation	Inconclusive
	No conclusion		Limited support different
	No conclusion, but with dissimilarities		Moderate support different
			Strong support different
	Dissimilar		Powerful support different
	Elimination	Elimination	Elimination
	No comparison Possible	Not suitable – with explanation	No comparison Possible

History of Standard Guide for Image Comparison Opinions

- Document originally started in Facial Identification Subcommittee.
- Added VITAL Subcommittee to create a DM SAC Task Group
- Original version included the following:
  - Exclusion
  - Strong Support for Exclusion
  - Support for Exclusion
  - Inconclusive
  - Support for Common Source
  - Strong Support for Common Source

Timeline Standard Guide for Image Comparison Opinions

- Sept 2020 – document approved to send to SDO by DM SAC (previous OSAC process)
- May 2021 – document received negative ballots at SDO.
- June 2021 - OSAC invited documents that had gone through the previous process to go through the STRP process. The task group decided to do this.
- July 2022 – Completed STRP with mixed response
- Aug 2022 – Submitted for review in to the FSSB. Petition received
- Sept 2022 - Ballot at FSSB = failed
- Feb 2023 – Recirculation ballot at FSSB = failed*
- May 2023 – Recirculation ballot #2 at FSSB = passed
- June 2023 – Posted to the OSAC Registry as proposed standard

Considerations

- Multiple Disciplines
- Framework vs Interdisciplinary Standard
- Total number of options on the scale
- Confusion surrounding original decisions in the standard
  - Original version included the following:
    - Exclusion
    - Strong Support for Exclusion
    - Support for Exclusion
    - Inconclusive
    - Support for Common Source
    - Strong Support for Common Source
- Understanding conclusion-centric vs evidence-centric

Why the confusion?

- Decision maker (conclusion-centric) vs Forensic Scientist (evidence-centric)
  - Some labs function as both the forensic scientist and the decision maker
    - Reach an opinion as a forensic scientist (based on the evidence) which then informs a decision maker (who considers other things/evidence) to reach a conclusion-centric opinion

Where did we end up?

- Evidence-Centric Scale
  - **Strong Support for Different Source**
    - an opinion category for which the observed dissimilar characteristics far outweigh the observed similar characteristics or where no distinctive similarities are observed. The nature and level of the observed similarities and dissimilarities in image characteristics are much more probable given the proposition that the images depict two different sources than given the proposition that the images depict the same source.

Where did we end up?

- Evidence-Centric Scale
  - **<u>Strong Support for Different Source</u>**
  - **<u>Support for Different Source</u>**
    - an opinion category that the observed dissimilar characteristics outweigh the similar characteristics but are insufficient to reach strong support for different source. The nature and level of the observed similarities and dissimilarities in image characteristics are more probable given the proposition that the images depict two different sources than given the proposition that the images depict the same source.

Where did we end up?

- Evidence-Centric Scale
  - **Strong Support for Different Source**
  - **Support for Different Source**
  - **Inconclusive**
    - an opinion category that there is insufficient information to form an opinion of common source or different source. The nature and level of the observed similarities and dissimilarities in image characteristics are equally probable given the proposition that the images depict two different sources and given the proposition that the images depict the same source.

Where did we end up?

- Evidence-Centric Scale
  - **Strong Support for Different Source**
  - **Support for Different Source**
  - **Inconclusive**
  - **Support for Common Source**
    - an opinion category that the observed similar characteristics outweigh the observed dissimilar characteristics but are insufficient to reach strong support for common source. The nature and level of the observed similarities and dissimilarities in image characteristics are more probable given the proposition that the images depict the same sources than given the proposition that the images depict the two different sources.

Where did we end up?

- Evidence-Centric Scale
  - **Strong Support for Different Source**
  - **Support for Different Source**
  - **Inconclusive**
  - **Support for Common Source**
  - **Strong Support for Common Source**
    - an opinion category that the observed similar characteristics far outweigh the observed dissimilar characteristics. The nature and level of the observed similarities and dissimilarities in image characteristics are much more probable given the proposition that the images depict the same sources than given the proposition that the images depict the two different sources.

Where did we end up?

- Evidence-Centric Scale
  - **Strong Support for Different Source**
  - **Support for Different Source**
  - **Inconclusive**
  - **Support for Common Source**
  - **Strong Support for Common Source**

Questions

Lora Sims
lora.sims@idealinnovations.com

Communicating Forensic Findings: *Forensic Document Examination*

Miriam Angel, Los Angeles Police Department

June 25-26, 2024

What does a Forensic Document Examiner (FDE) do?

Examine documents to determine…

- Who wrote them
- Whether they are authentic
- Facts about them, such as how they were created or handled

FDE Expertise – Handwriting

Studies on FDE vs. Layperson performance on handwriting and signature comparisons

- Similar rates of correct association responses
- Laypersons make significantly more incorrect associations

Differentiability of Handwriting

- "…FDEs appreciate the sources and range of natural variation both between and within individuals.
- The causes of intra- and inter-writer variation, and the arguments for why intra-writer variation is smaller than inter-writer variation, have deep roots in motor control theory." (*Forensic Handwriting Examination and Human Factors, 2021*)

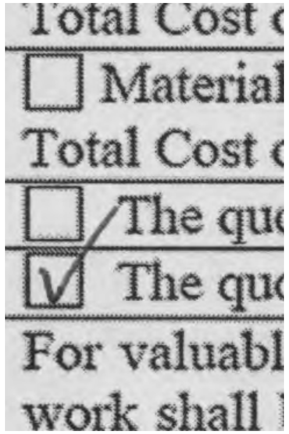
Handwriting is the Result of a Behavior/Activity

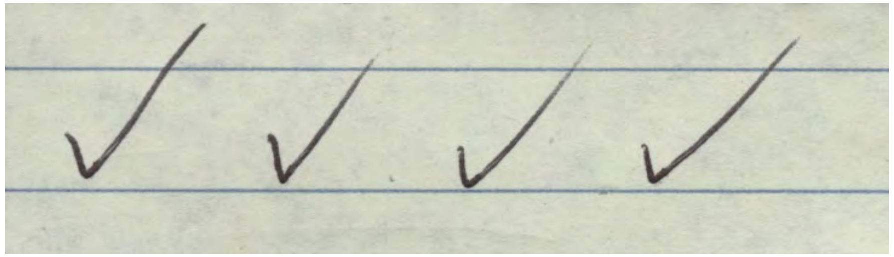
- A person's handwriting can vary beyond normal due to
  - internal or external factors (such as illness or uncomfortable writing conditions)
  - purposeful change (disguise or simulation)

Basis for a Handwriting Opinion Supporting Same-Source

- “No two people write alike”

Same writer?





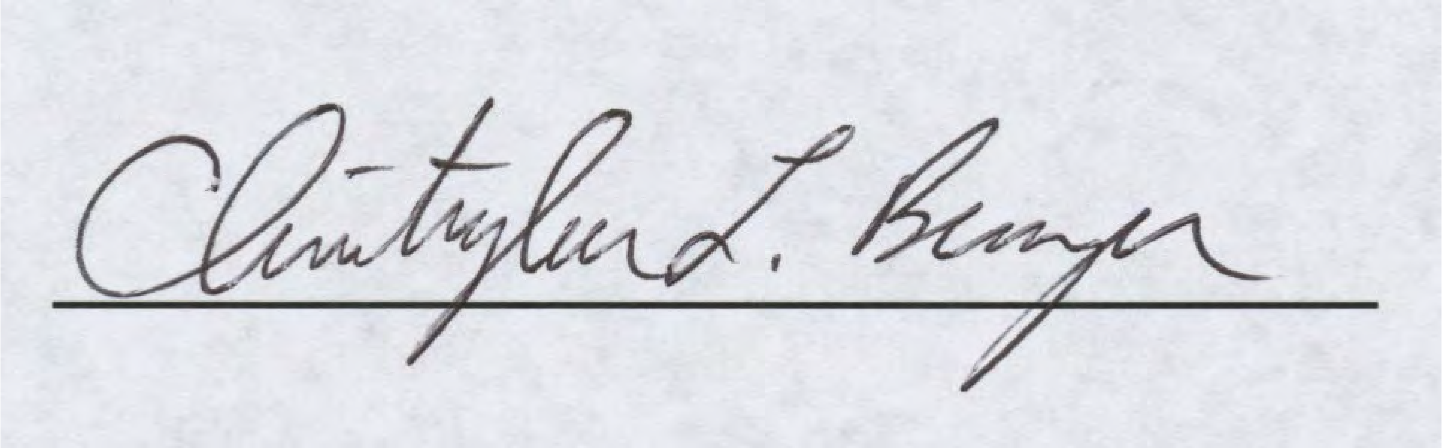
Q **K**

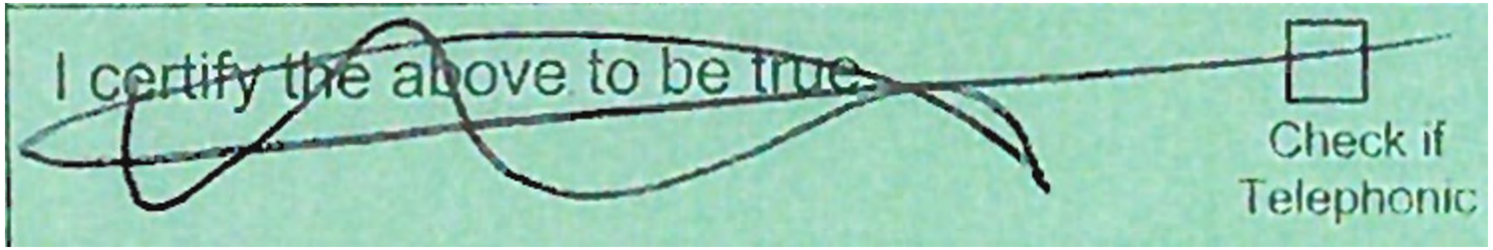
Basis for a Handwriting Opinion Supporting Same-Source

- ~~"No two people write alike"~~

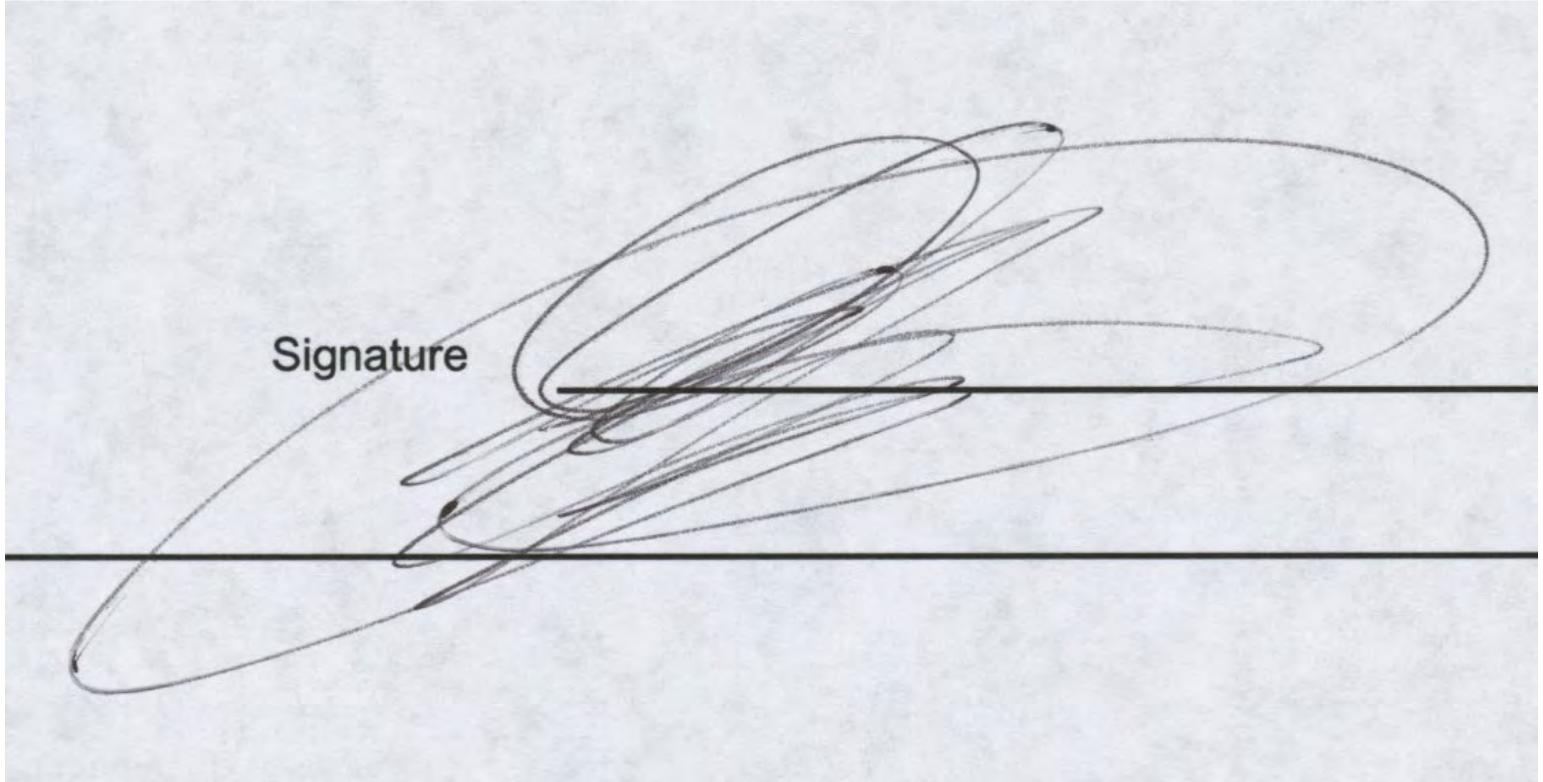
- Given a ***sufficient amount*** of ***natural*** writing, no two people are ***likely*** to produce handwriting that is exactly the same in terms of character construction, line quality, and other handwriting features.

Complexity

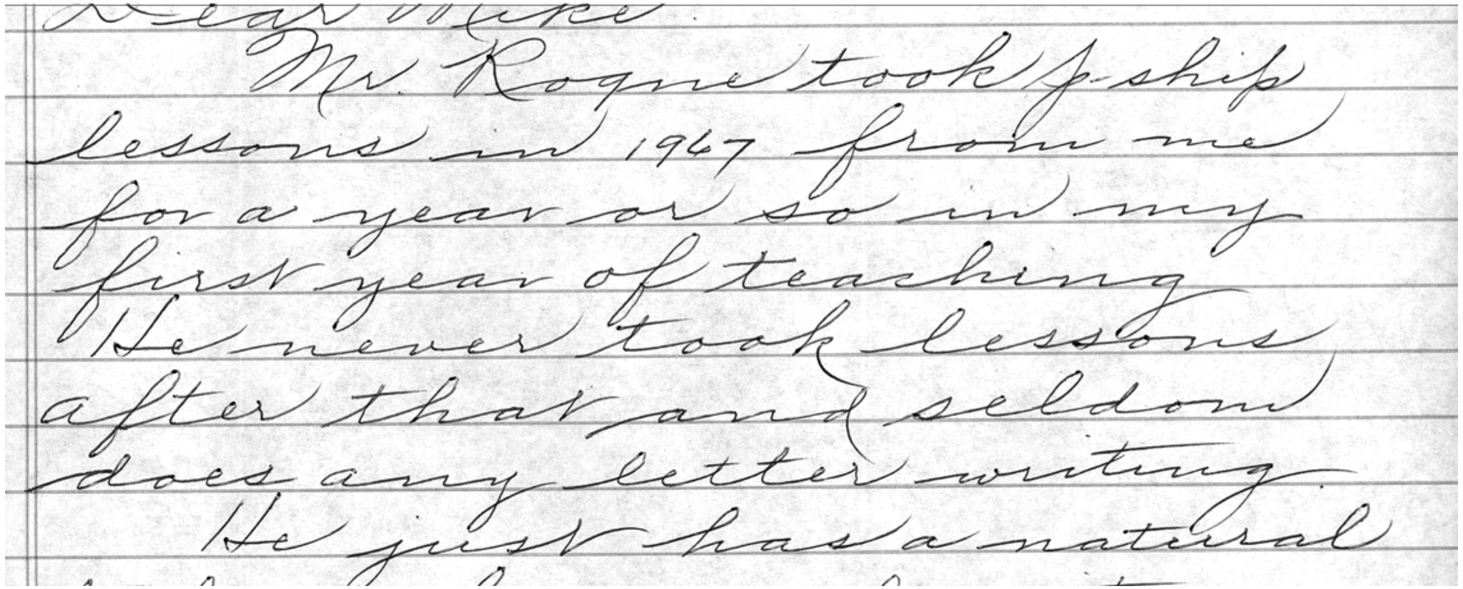




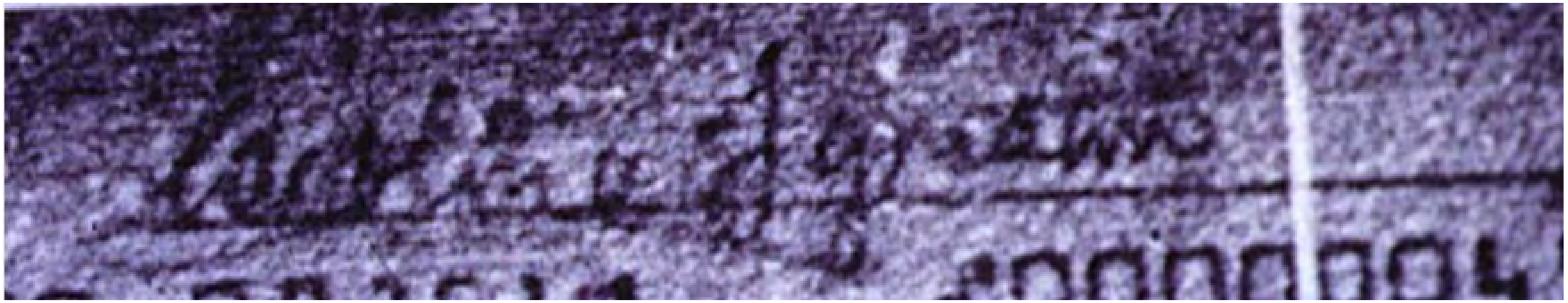
Complexity?



Limitations



Dear Mike:
Mr. Rogue took p-ship lessons in 1967 from me for a year or so in my first year of teaching. He never took lessons after that and seldom does any letter writing.
He just has a natural



Assessments are Subjective

Forensic handwriting examination, as currently practiced, is not about:

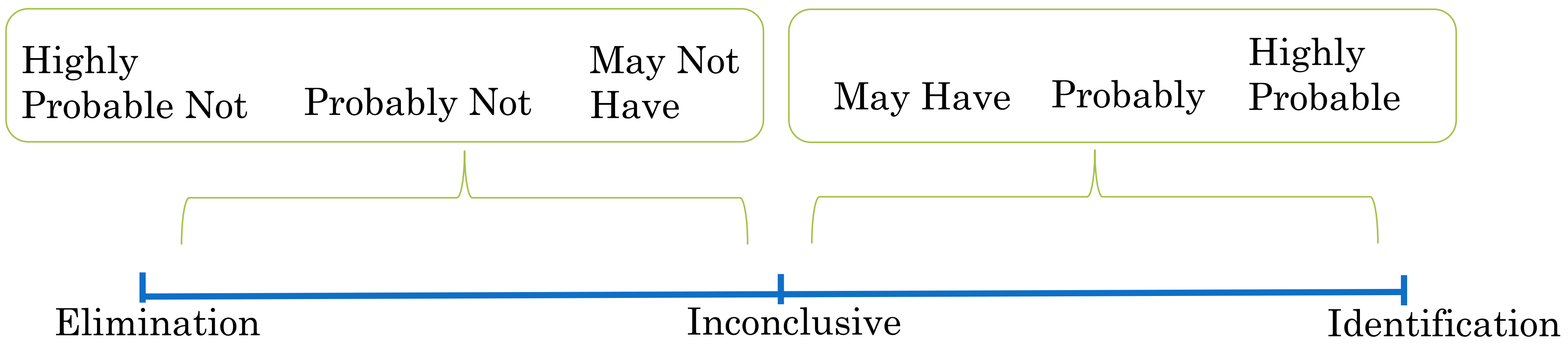
- Weighing similarities and differences (only)
- Frequency occurrence proportions of individual characteristics

Assessments are Subjective

Opinions are based on the examiner's…

- understanding of the case information and request
- knowledge of handwriting
- experience
- reasoned judgment

Current Opinion Scale



Current Conclusions Standard

4. Terminology

4.1 *Recommended Terms:*

Identification (definite conclusion of identity)—this is the highest degree of confidence expressed by document examiners in handwriting comparisons. The examiner has no reservations whatever, and although prohibited from using the word “fact,” the examiner is certain, based on evidence contained in the handwriting, that the writer of the known material actually wrote the writing in question.

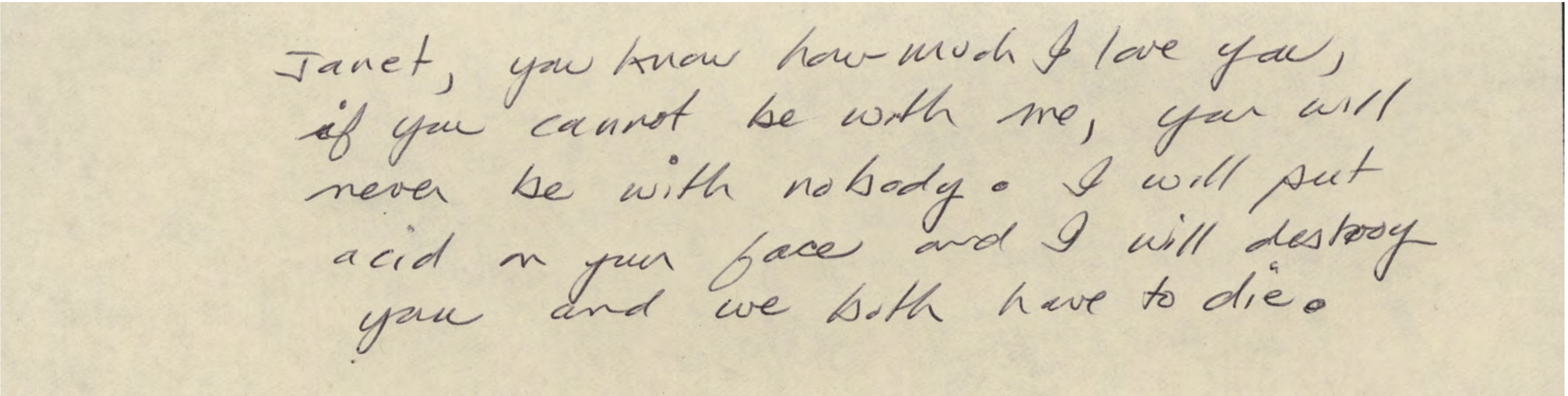
-From the *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners*

Current Conclusions Standard

strong probability (highly probable, very probable)—the evidence is very persuasive, yet some critical feature or quality is missing so that an identification is not in order; however, the examiner is virtually certain that the questioned and known writings were written by the same individual.

-From the *SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners*

Questioned (Q)


Janet, you know how much I love you,
if you cannot be with me, you will
never be with nobody. I will put
acid on your face and I will destroy
you and we both have to die.

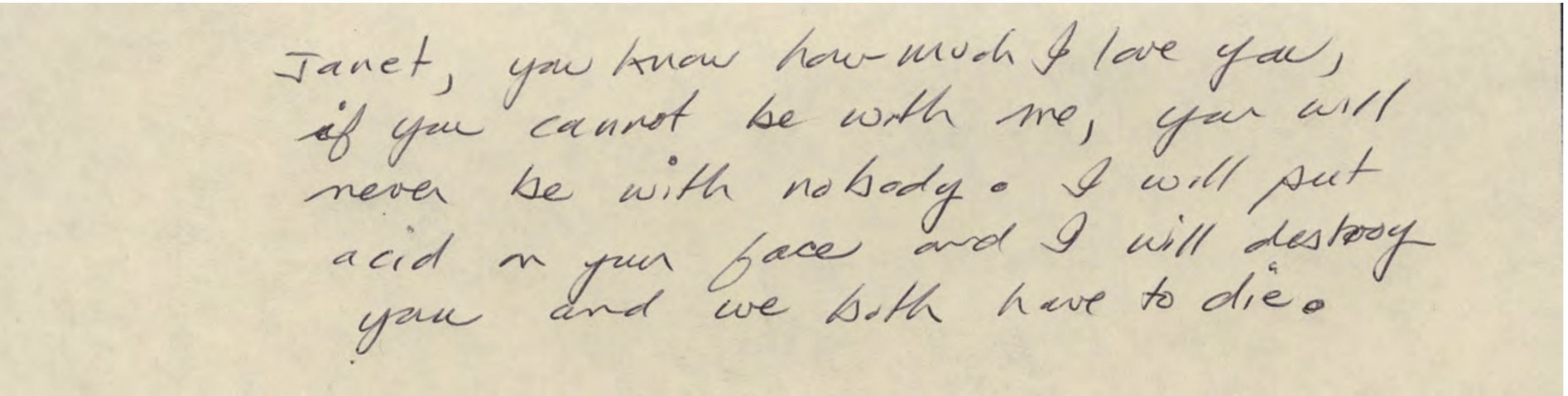
Known Writer 1

K1

Known Writer 2

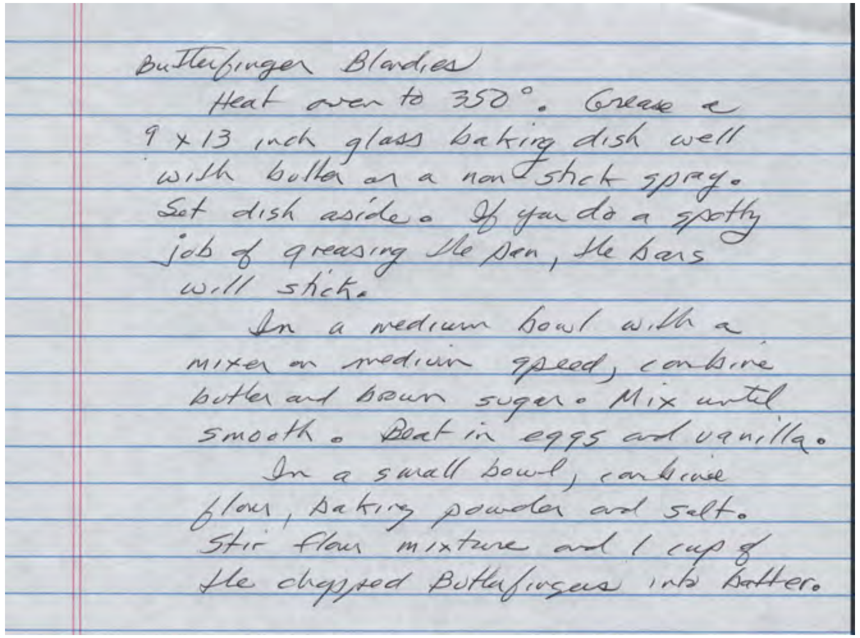
K2

Q



Janet, you know how much I love you, if you cannot be with me, you will never be with nobody. I will put acid on your face and I will destroy you and we both have to die.

K1

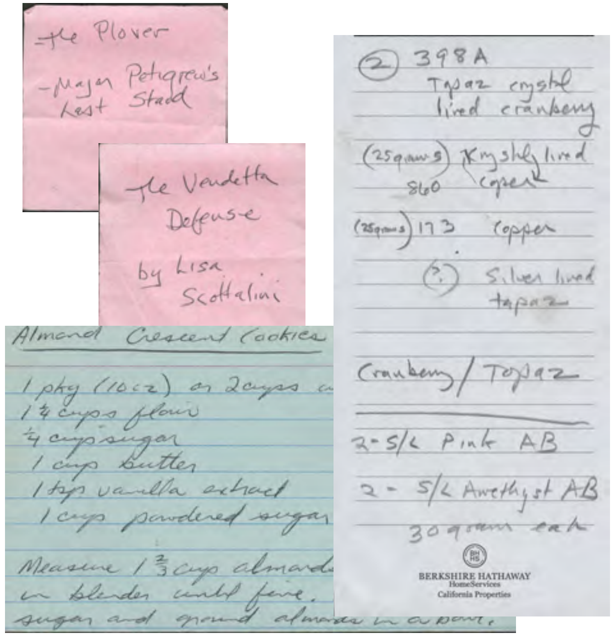


Butterfinger Blondies

Heat oven to 350°. Grease a 9 x 13 inch glass baking dish well with butter or a non stick spray. Set dish aside. If you do a spotty job of greasing the pan, the bars will stick.

In a medium bowl with a mixer on medium speed, combine butter and brown sugar. Mix until smooth. Beat in eggs and vanilla.

In a small bowl, combine flour, baking powder and salt. Stir flour mixture and 1 cup of the chopped Butterfingers into batter.



The Plover

- Major Pettigrew's Last Stand

The Vendetta Defense

by Lisa Scottoline

Almond Crescent Cookies

1 pkg (10 oz) or 2 cups
1 1/4 cups flour
1/4 cup sugar
1 cup butter
1 tsp vanilla extract
1 cup powdered sugar

Measure 1 1/3 cups almonds in blender until fine.
sugar and ground almonds in a bowl.

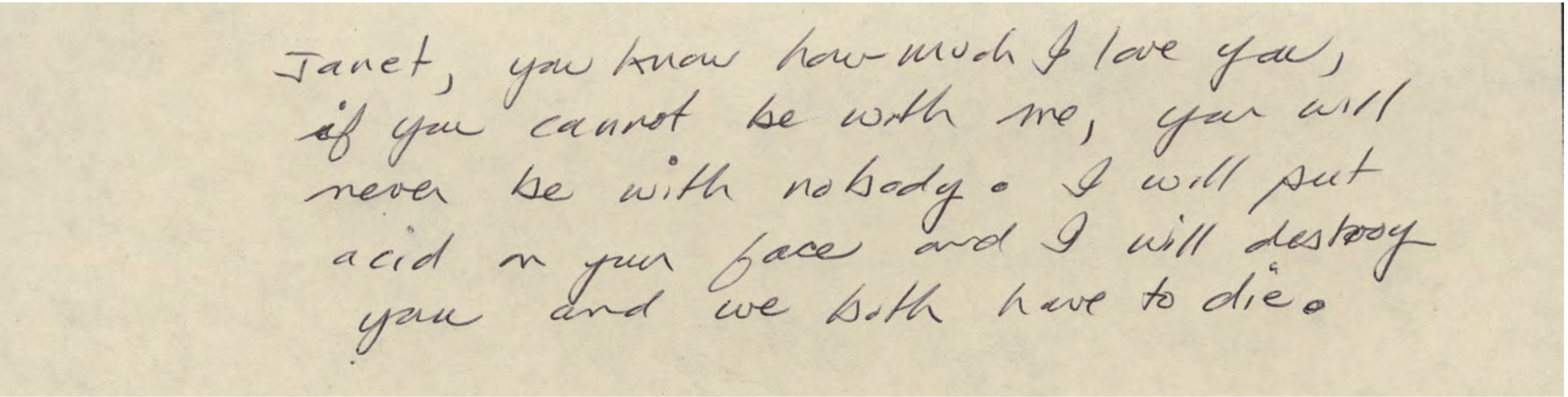
(2) 398A
Topaz crystal lined cranberry
(25 grams) Km shly lined copper
860
(25 grams) 173 copper
(?) Silver lined topaz

Cranberry / Topaz

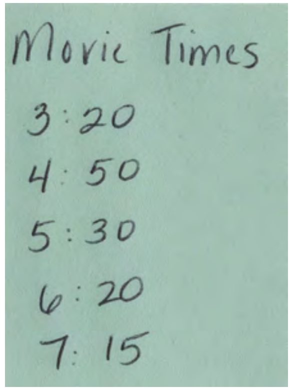
2 - S/L Pink AB
2 - S/L Amethyst AB
30 grams each

BERKSHIRE HATHAWAY
HomeServices
California Properties

Q


Janet, you know how much I love you, if you cannot be with me, you will never be with nobody. I will put acid on your face and I will destroy you and we both have to die.

K2



Opinions using current standard

- Examiner A
  - It is highly probable the K1 writer wrote Q.
  - No determination could be made whether the K2 writer wrote Q.

- Examiner B
  - It is highly probable the K1 writer wrote Q.
  - It is highly probable the K2 writer did not write Q.

Proposed Standard Opinions

- Extremely Strong Support for Different Sources
- Strong Support for Different Sources
- Moderate Support for Different Sources
- Limited Support for Different Sources
- Equal support
- Limited Support for Common Source
- Moderate Support for Common Source
- Strong Support for Common Source
- Extremely Strong Support for Common Source

“Support”

- If the probability of observing the handwriting features if proposition X is true is larger than the probability of observing these features if proposition Y is true, then the findings support proposition X over proposition Y.
- If the probability of observing these features is about the same under both propositions, the findings provide approximately equal support for each proposition.

Transition Phase

If your opinion in SWGDOC terminology is:	Your conclusion should/may be worded:
Elimination	Extremely strong support for different sources
Highly probable did not	Strong support for different sources
Probably not	Moderate (or strong) support for different sources
Indications may not have	Limited (or moderate) support for different sources
No conclusion / Inconclusive	Equal support
Indications may have	Limited (or moderate) support for same source
Probably	Moderate (or strong) support for same source
Highly Probable	Strong support for same source
ID	Extremely Strong support for same source

*Important Note: The conclusions in the second column follow from those in the first column. However, the reverse is not true; conclusions in the first column **do not** necessarily follow from those in the second column.

Evaluative Reporting Approach

Ideally, the examiner would

- Assess the probability of observing the findings if the same person wrote both samples of writing

and

- Assess the probability of observing the findings if different people wrote both samples of writing

If the findings are…	Verbal expression…
Exactly as expected	Extremely high probability
Largely as expected	Very high probability
Moderately as expected	Moderately high probability
Neutral	Balanced probability
Moderately divergent	Moderately low probability
Largely divergent	Very low probability
Completely divergent	Extremely low probability

Limitations move the probability/probabilities toward Neutral

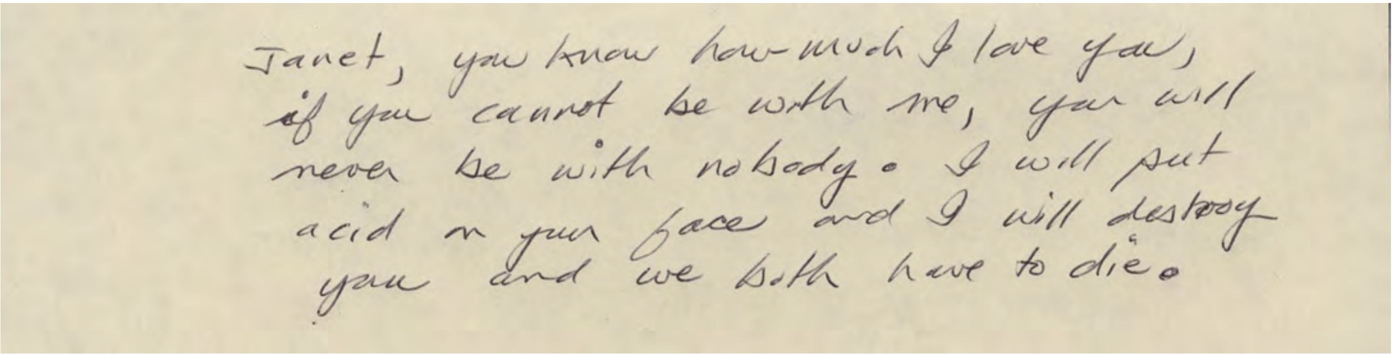
*Table adapted from Brent Ostrum's Table 3.2 in Forensic Document Examination in the 21st Century

Example Case

- H1: The writer of the known writing K wrote the questioned document Q
- H2: Someone other than the writer of the known writing K wrote the questioned document Q

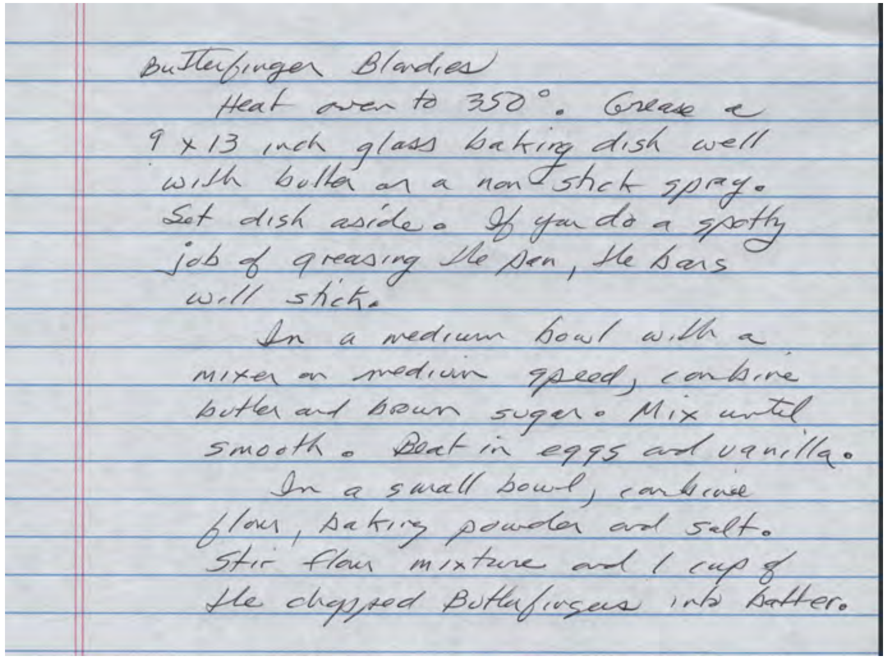
Evaluation

- Probability of the findings given H1 is **Very High**
- Probability of the findings given H2 is **Very Low**



Janet, you know how much I love you,
if you cannot be with me, you will
never be with nobody. I will put
acid on your face and I will destroy
you and we both have to die.

Q



Butterfinger Blondies
Heat oven to 350°. Grease a
9 x 13 inch glass baking dish well
with butter or a non stick spray.
Set dish aside. If you do a spotty
job of greasing the pan, the bars
will stick.
In a medium bowl with a
mixer on medium speed, combine
butter and brown sugar. Mix until
smooth. Beat in eggs and vanilla.
In a small bowl, combine
flour, baking powder and salt.
Stir flour mixture and 1 cup of
the chopped Butterfingers into batter.

K

		Probability of the findings given H1						
		Extremely Low	**Very Low**	**Low**	**Neutral**	**High**	**Very High**	**Extremely High**
Probabilities of the findings given H2	**Extremely Low**	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2	Evidence provides extremely strong support for H1 over H2 **or** Evidence provides strong support for H1 over H2	Evidence provides extremely strong support for H1 over H2
	Very Low	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2	Evidence provides extremely strong support for H1 over H2 **or** Evidence provides strong support for H1 over H2
	Low	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2
	Neutral	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2
	High	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2
	Very High	Evidence provides extremely strong support for H2 over H1 **or** Evidence provides strong support for H2 over H1	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2
	Extremely High	Evidence provides extremely strong support for H2 over H1	Evidence provides extremely strong support for H2 over H1 **or** Evidence provides strong support for H2 over H1	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2

Probabilities of the findings given H2 \ Probability of the findings given H1	Extremely Low	Very Low	Low	Neutral	High	Very High	Extremely High
Extremely Low	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2	Evidence provides extremely strong support for H1 over H2 **or** Evidence provides strong support for H1 over H2	Evidence provides extremely strong support for H1 over H2
Very Low	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2	Evidence provides extremely strong support for H1 over H2 **or** Evidence provides strong support for H1 over H2
Low	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2	Evidence provides strong support for H1 over H2 **or** Evidence provides moderate support for H1 over H2
Neutral	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2	Evidence provides moderate support for H1 over H2 **or** Evidence provides limited support for H1 over H2
High	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2
Very High	Evidence provides extremely strong support for H2 over H1 **or** Evidence provides strong support for H2 over H1	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2
Extremely High	Evidence provides extremely strong support for H2 over H1	Evidence provides extremely strong support for H2 over H1 **or** Evidence provides strong support for H2 over H1	Evidence provides strong support for H2 over H1 **or** Evidence provides moderate support for H2 over H1	Evidence provides moderate support for H2 over H1 **or** Evidence provides limited support for H2 over H1	Evidence provides limited support for H2 over H1	Evidence provides limited support for H1 over H2 **or** Evidence provides approximately equal support for H1 and H2	Evidence provides approximately equal support for H1 and H2

Opinion

The findings provide strong support for the writer of the known document (K) having written the questioned document (Q), rather than someone other than the K writer having written it.

This results because (1) the probability of the findings if K1 wrote Q is very high, due to the strong similarities between the samples in almost all handwriting features and (2) the probability of the findings if someone other than the writer of K1 wrote it is very low, due to the large amount of questioned writing, where the combination of handwriting features are unlikely to be observed in more than one person's writing.

Gaps and How to Fill Them

William C. Thompson
University of California, Irvine

NIST Workshop on Communicating Forensic Findings
June 26, 2024

Shared Goal

Forensic science findings should be reported in a way that is:

- Justifiable
- Complete
  - Uncertainties, limitations, and assumptions are disclosed
- Understandable
  - So that factfinders give the evidence the weight it deserves

But do we know enough to say what is justifiable and understandable; and can we agree on what is needed to be complete?

Gap: What statements are justifiable?

- In many disciplines we disagree about the probative value of forensic evidence for distinguishing specific propositions
  - Even in disciplines where there is an empirical basis for LRs, there may be uncertainty about some cases
  - Black-box studies of the accuracy of analysts' decisions are not as helpful for establishing probative value as studies of the frequency of features.
  - The ability of factfinders to evaluate the strength of the evidence is most important for contested cases, which are the very ones where the true probative value of the evidence is difficult to assess

What statements are justifiable?

We sometimes disagree about the proper basis for forensic scientists' opinions

- Latent print examiner influenced by "size of the population of potential suspects"
- Bloodstain pattern expert influenced by medical examiner's report
- Medical examiner influenced by suspect's statements to police in assessing cause and manner of death

- TI-Info may induce shift of decision threshold, which can dramatically undermine probative value of forensic evidence, thereby reducing accuracy of legal system
  - See Thompson, Shifting Decision Thresholds… PNAS, 2023 ("the criminalists' paradox)

Gap: What should experts say about underlying assumptions?

E.g., relevant reference population; NoC

Options

- Pick "best" assumption and report under it?
- Give an alternative report under each possible assumption?
- Average or integrate over possible assumptions?
  - Possibly testing sensitivity to priors?

See commentary on Thompson, Uncertainty in probabilistic genotyping…, JFS, 2023

Biedermann, Taroni & Thompson. Using graphical probability analysis (Bayes nets) to evaluate a conditional DNA inclusion. *Law, Probability and Risk,2011.*

Gap: Which sources of uncertainty should experts take into account, and how?

E.g., What should firearms experts say in reports and testimony about error rates in black-box studies of bullet and cartridges comparisons?

Should reports be "conservative"; if so, how?

- Rationales for "conservative" reporting
  - Presumption of innocence
    - Not convincing—See, Kaye, Hypothesis testing in law and forensic science…Harvard Law Review Forum, 2017.
  - Scientific humility—resolve uncertain by making more modest rather than bolder claim
- Example of confusion
  - Thompson, Uncertainty in probabilistic genotyping, JFS 2023
    - "conservatism" in reporting STRMix findings

Gap: Do we know enough about lay perception to evaluate whether a statement will be "understood"?

- Bali, Martire, Edmond, *Lay comprehension*…Law & HumBeh, 2021)—CASOC: consistency, inferential ability, sensitivity, orthodox updating, coherent reasoning
  - Key issue: Do those who rely on reports and testimony give forensic findings ***the weight that they deserve?***

Research Strategies

  - Perceived relative weight
  - Effects of forensic evidence on evaluations and decisions

Example for Discussion

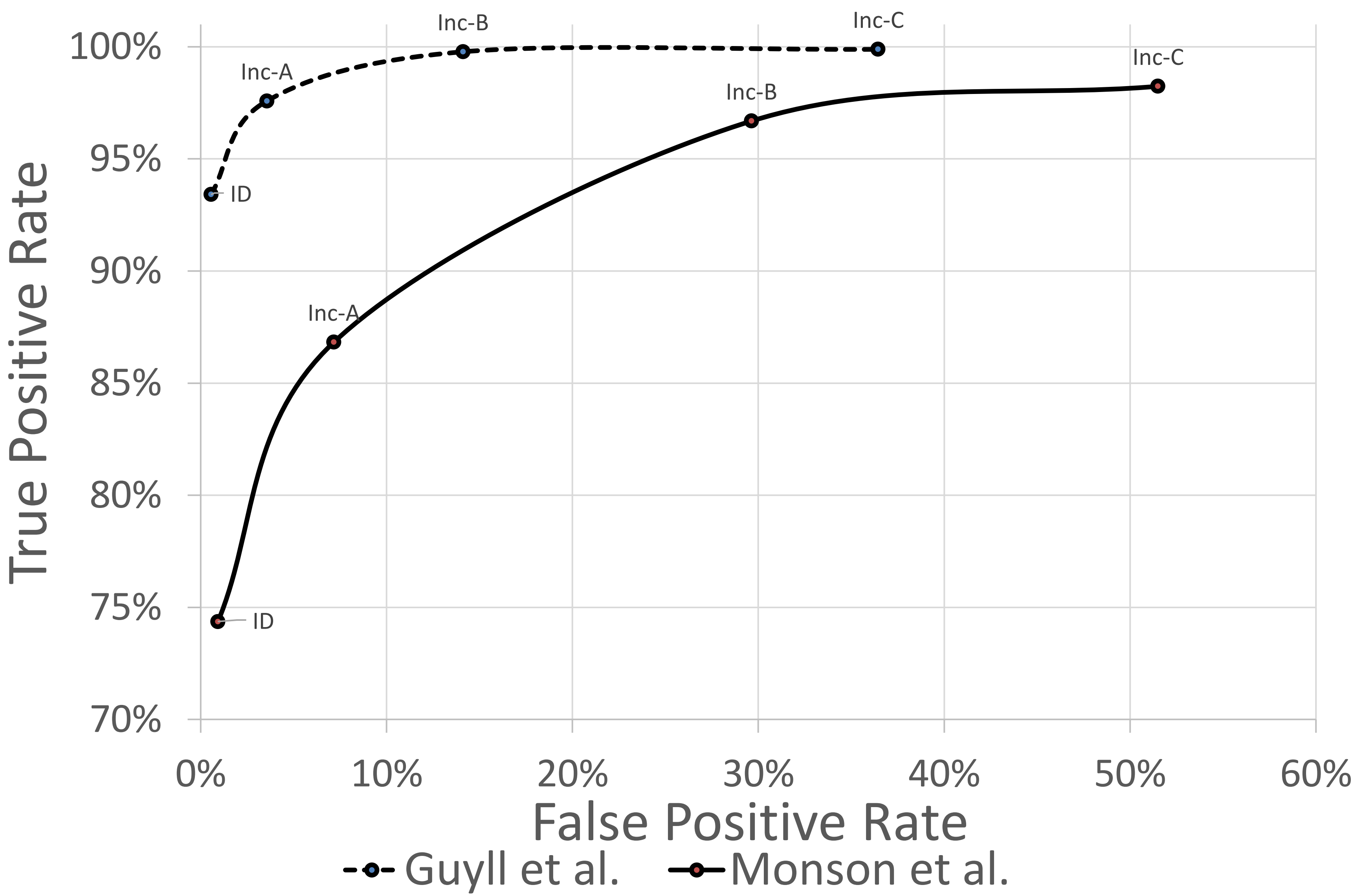
Is this statement in the DOJ ULTR on Firearms justifiable, complete and understandable?

…. the observed class characteristics and corresponding individual characteristics provide ***extremely strong support*** for the proposition that the two toolmarks originated from the same source ….(emphasis added)

Justifiable? Understandable?

- ENFSI guidelines limit claims of “extremely strong support” to LRs of 1 million and above
- Studies on perceived weight show people treat “extremely strong support” as equivalent to LRs of 100K+ or RMPs<1 in 100K
  - Thompson, et al. Perceived strength of forensic scientists’ reporting statements. Law,Prob&Risk, 2018
- Does the strength of firearms evidence justify such statement?

ROC Curves Cartridge Comparison based on Observed Error Rates in Monson et al. (2023) and Guyll et al. (2023) at Four Decision Thresholds

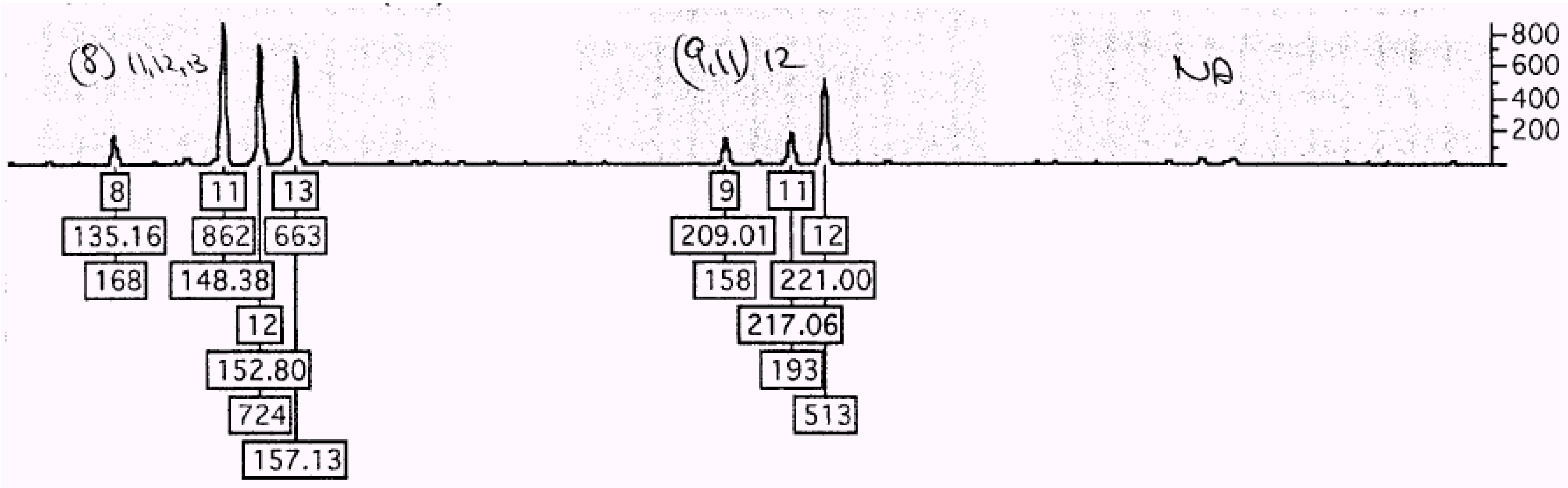


Threshold	True ID Rate	False ID Rate	LR	ENFSI Label
	Monson et al.			
ID	74.4	0.92	80.87	Moderate
Inc-A	86.8	6.64	13.07	Moderate
	Guyll et al.			
ID	93.4	0.56	166.79	Moderately Strong
Inc-A	97.6	3.36	29.03	Moderate

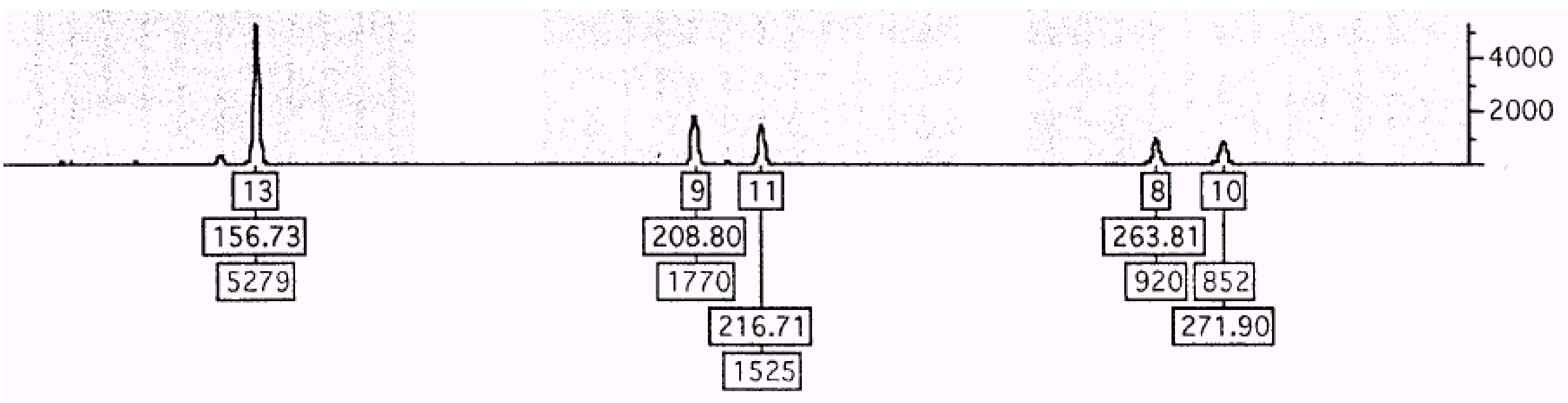
Additional caveats

- Error rates in black-box studies are aggregates that average across some important variables while ignoring others
- Underlying variables—e.g., item effects; examiner skill; decision thresholds—may have HUGE effects

DNA Profile of Hat—ProfilerPlus Yellow Loci



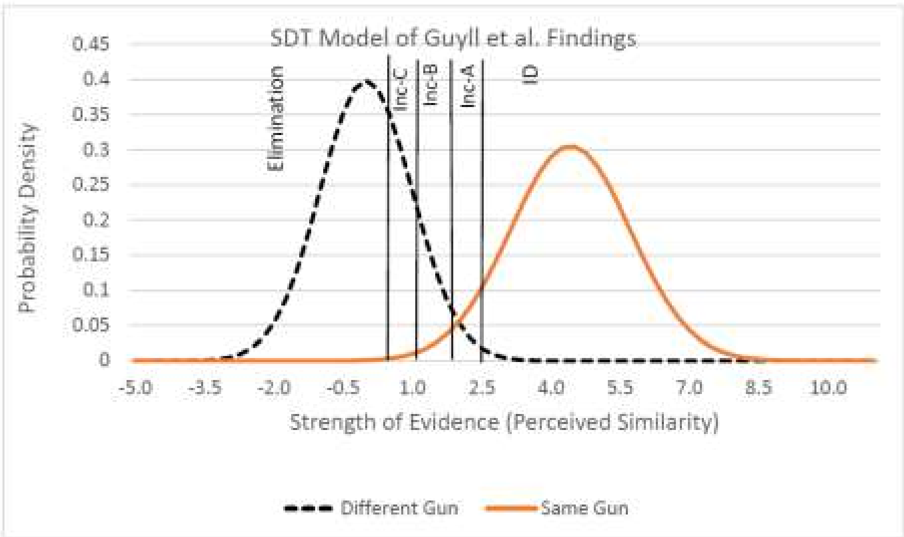
Profile of Defendant



Acknowledgements

- NIST funding through CSAFE (Center for Statistical and Applications in Forensic Evidence)(RIP)

- Contact: william.thompson@uci.edu



References

Thompson, W.C. (2023). Shifting decision thresholds can undermine the probative value and legal utility of forensic pattern-matching evidence. *Proceedings of the National Academy of Sciences, USA*, 120(41): e2301844120e.

Thompson, W.C. (2023). Uncertainty in probabilistic genotyping of low template DNA: A case study comparing STRMix™ and TrueAllele™ *Journal of Forensic Sciences*, 68(3): 1049-1063.

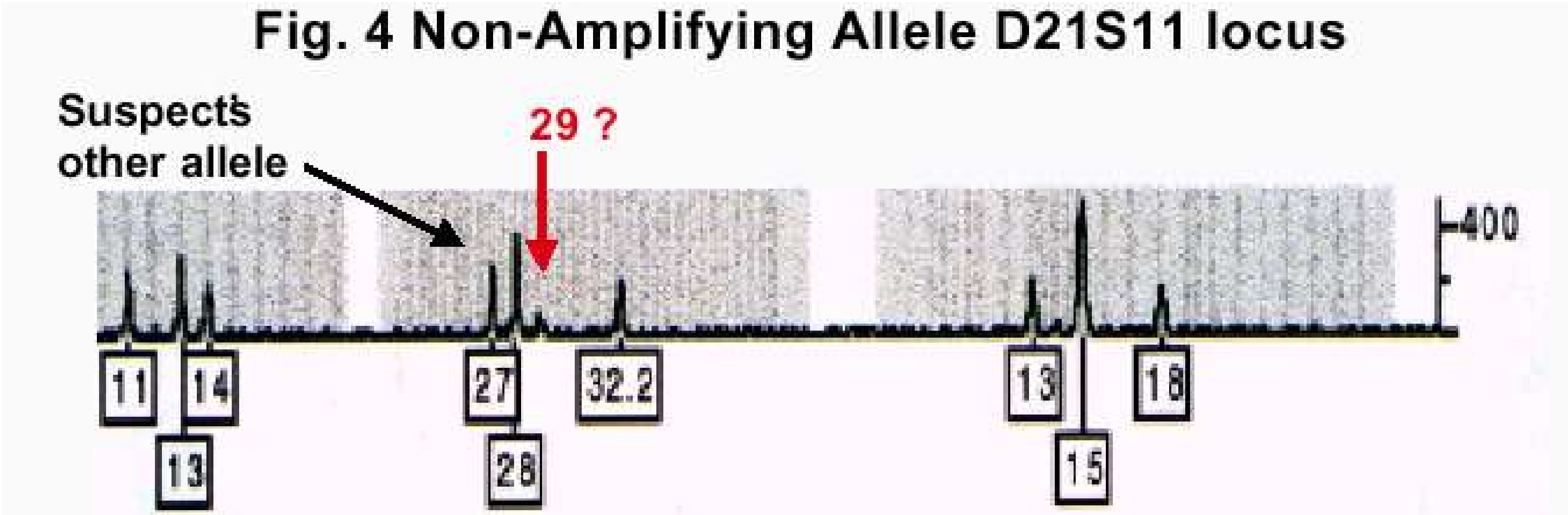
Thompson, W.C. (2018). How should forensic scientists present source conclusions? *Seton Hall Law Review*, 48(3): 774-813.

Thompson, W.C., Vuille, J., Taroni, F., & Biedermann, A. (2018). After Uniqueness: The Evolution of Forensic Science Opinion. *Judicature,* 102(1): 18-27.

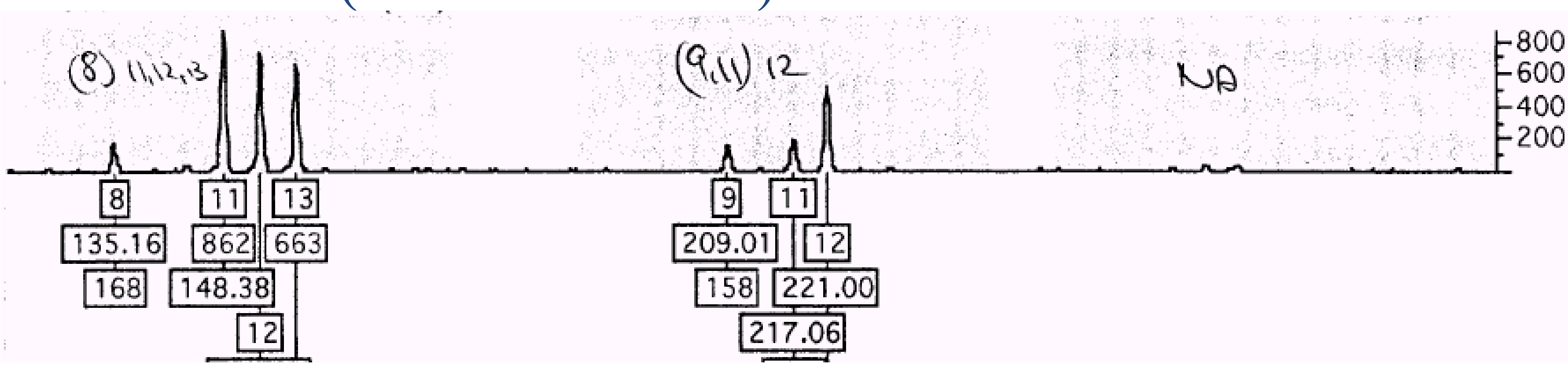
Thompson, W.C., Grady, R.H., Lai, E. & Stern, H. (2018). Perceived strength of forensic scientists' reporting statements about source conclusions. *Law, Probability & Risk,* 17(2): 133-155.

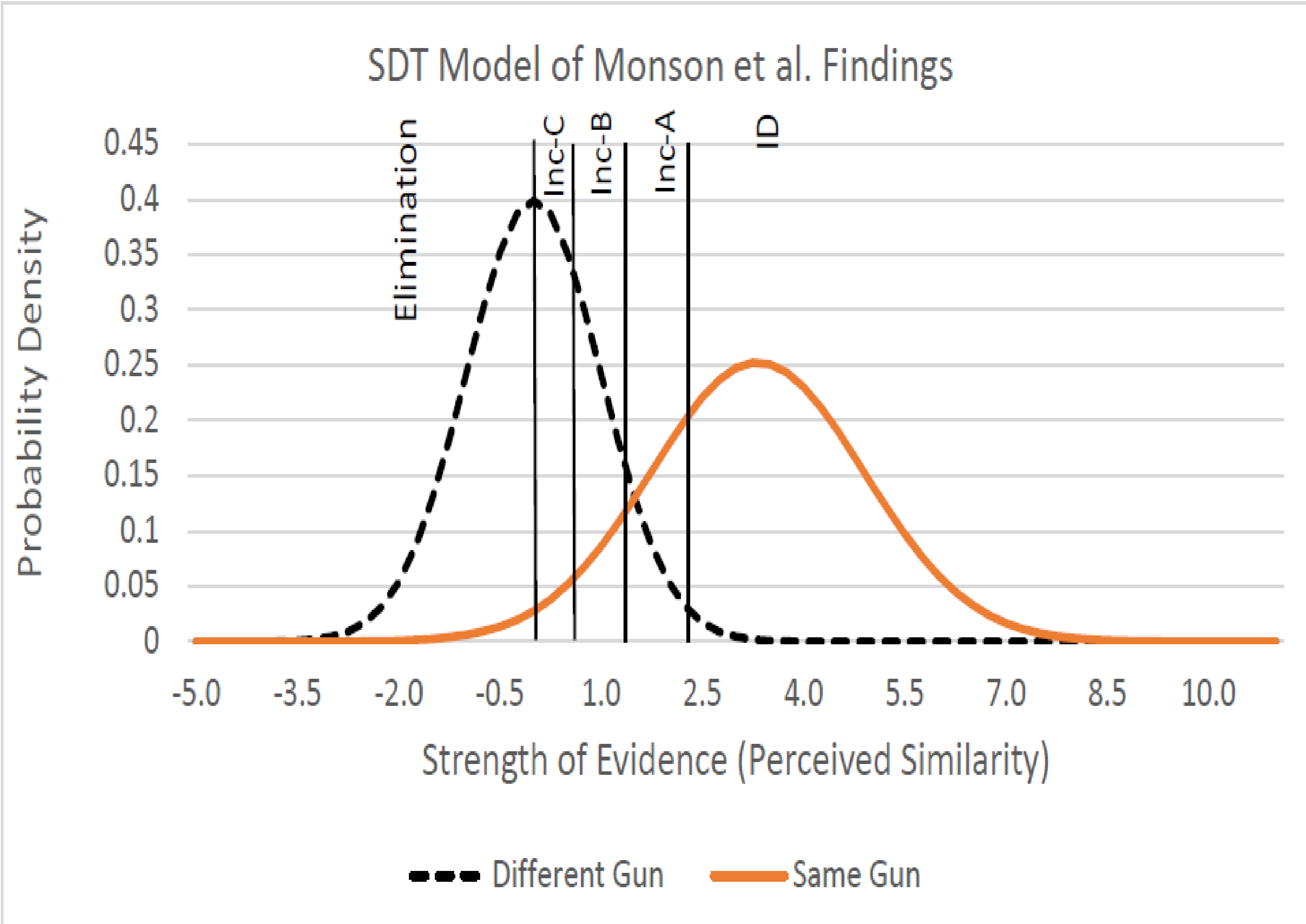
Thompson, W.C. & Newman, E.J. (2015). Lay understanding of forensic statistics: Evaluation of random match probabilities, likelihood ratios, and verbal equivalents. *Law & Human Behavior. 39(4):* 332-349.

NIST Mixed Stain Study #2 (2 contributors)

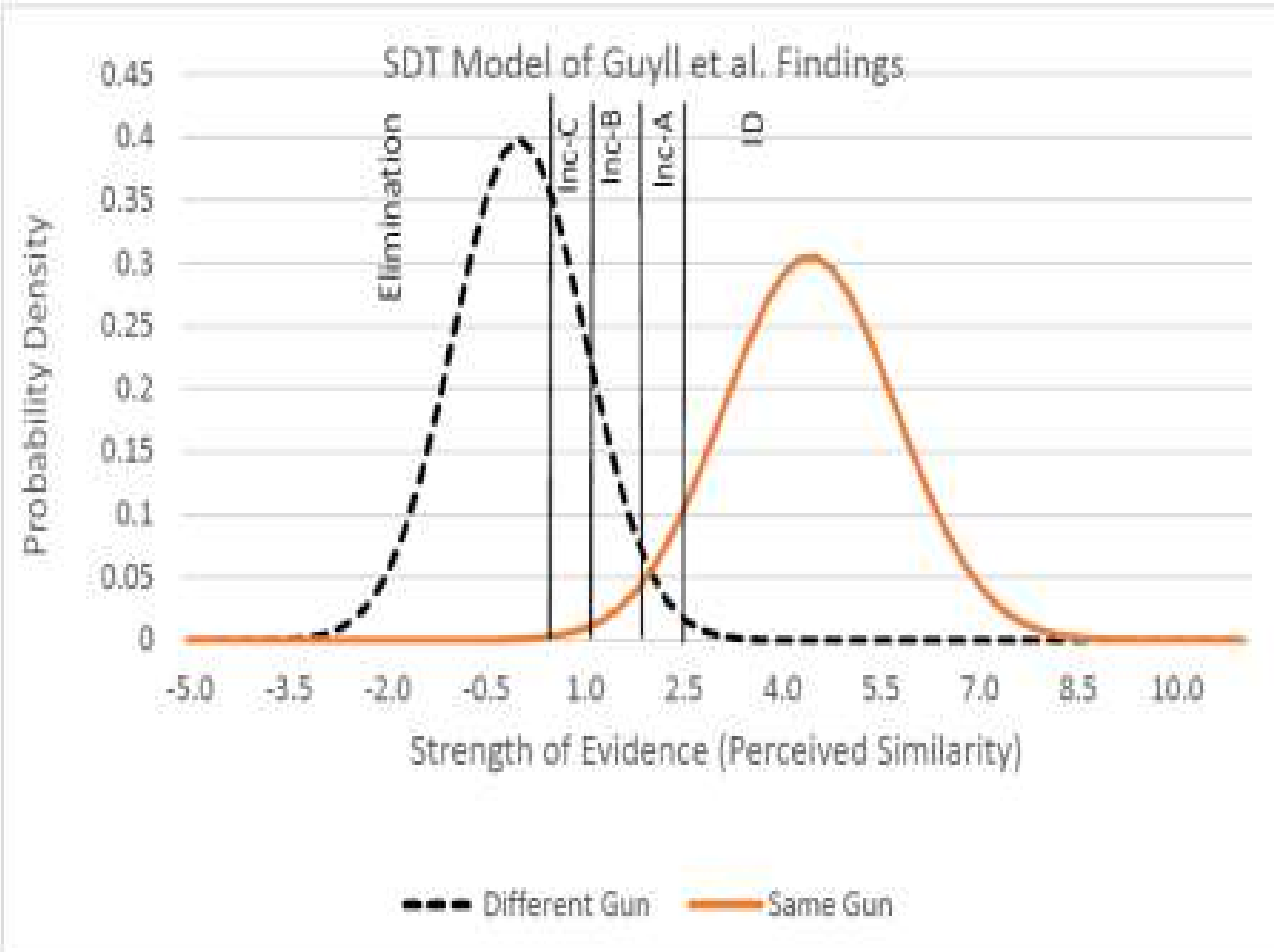


Hat alleles (2 contributors?)





SDT Model of Monson et al. Findings
Probability Density
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Elimination
Inc-C
Inc-B
Inc-A
ID
-5.0
-3.5
-2.0
-0.5
1.0
2.5
4.0
5.5
7.0
8.5
10.0
Strength of Evidence (Perceived Similarity)
Different Gun
Same Gun


SDT Model of Guyll et al. Findings
Probability Density
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Elimination
Inc-C
Inc-B
Inc-A
ID
-5.0
-3.5
-2.0
-0.5
1.0
2.5
4.0
5.5
7.0
8.5
10.0
Strength of Evidence (Perceived Similarity)
Different Gun
Same Gun

Communicating Forensic Findings Workshop: Current Practices and Future Directions

Session 5: Gaps and How to Fill Them

Identifying Gaps & Limitations via SOPs

Kate Philpott

"A primary purpose for validation studies then is to push the system until it fails in order to understand the potential limitations – to define the scope of method (and interpretation) reliability"

(Butler, Validation Webinar (2014), slide 12)

PLANNING (including study design)

EXPERIMENTS and DATA COLLECTION

DATA EVALUATION

SOP DEVELOPMENT

Clearly defined criteria and limitations tightly connected to validation data promote interpretations that are:

- Justifiable
- Complete
- Understandable *(ala Bill T)*

Recommendation 3.8: Forensic science service providers' standard operating procedures should provide criteria for assessing and documenting when a probabilistic genotyping interpretation should be rejected.

HOW TO “PUSH THE SYSTEM UNTIL IT FAILS”

First step: understanding what factors challenge the system

The factors that make DNA mixture interpretation challenging are well known:

- ✓a large and/or unknown **number of contributors**;
- ✓sub-optimal amounts of **template DNA** (i.e. stochastic effects);
- ✓skewed/evenly-distributed **mixture ratios**;
- ✓**allele sharing** between two or more contributors to a mixture (as well as between true contributors and non-contributors); and
- ✓**degradation/inhibition** (including varying degrees of degradation between contributors) of template DNA

“FACTOR SPACE”

Validation study design → limits of analysis

Four-person mixtures	Three-person mixtures	Two-person mixtures	Total target template (pg)
1:1:1:1, 4:4:1:1, 6:3:1:1, 13:3:3:1	1:1:1, 3:3:1, 6:3:1, 8:1:1	N/A	1000
1:1:1:1, 4:4:1:1, 6:3:1:1, 13:3:3:1	1:1:1, 3:3:1, 6:3:1, 8:1:1	1:1, 5:1, 10:1, 19:1	500
1:1:1:1, 4:4:1:1, 6:3:1:1, 13:3:3:1	1:1:1, 3:3:1, 6:3:1, 8:1:1	1:1, 5:1, 10:1, 19:1	250
1:1:1:1, 4:4:1:1, 6:3:1:1, 13:3:3:1	1:1:1, 3:3:1, 6:3:1, 8:1:1	1:1, 5:1, 10:1, 19:1	100

The lowest percentage/ratio for any of these contributors, no matter the number of contributors, is 1:19 (1/20) or 5%. Mixtures where the lowest level contributor comprises less than 5% of the sample are beyond the bounds of validation.

WARNING!

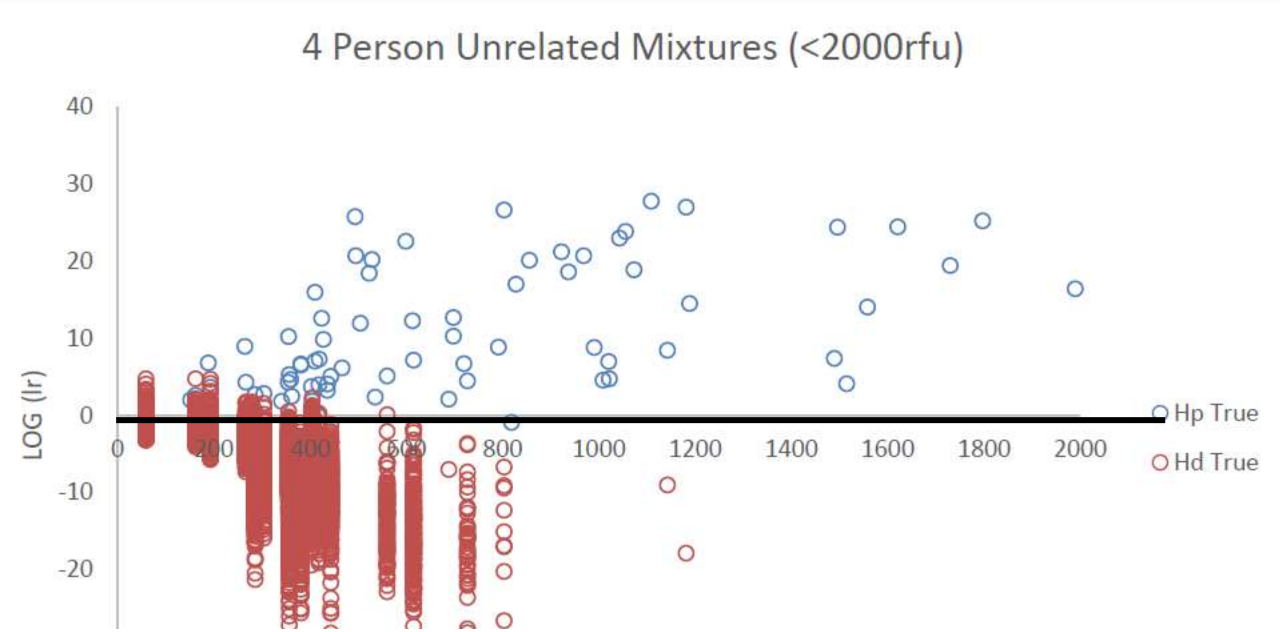
Mixtures with contributors donating as little as 5% and [X] pg of template DNA were tested during internal validation. If PGS analysis of a mixture associates a person of interest with a contributor whose estimated template DNA is lower (in terms of ratio or quantity), the sample is outside the scope of validation and should be deemed uninterpretable.

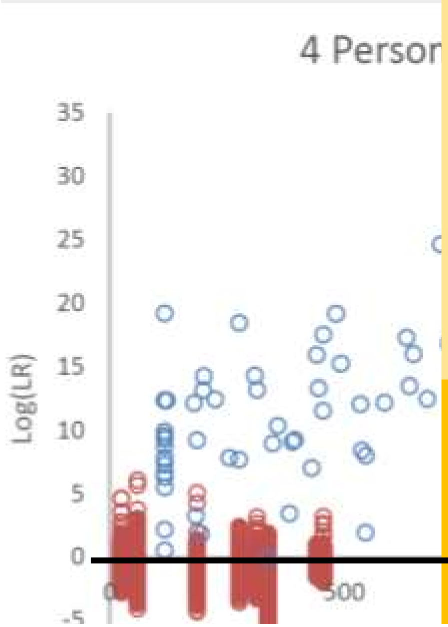
CAUTION!

Only two mixture samples tested during internal validation involved contributors donating as little as [X] pg of template DNA. If PGS analysis of a mixture associates a person of interest with a contributor donating similar amount of DNA, there are high levels of uncertainty associated with this analysis. Extreme caution should be exercised if interpretation is attempted.

Validation results → limits of reliable

E.g. *False inclusions can occur with low level contributors in DNA mixtures. In validation, false inclusions were observed for mixture components with average peak heights approaching 600 rfu. For mixtures with high levels of allele sharing (e.g. related individuals), false inclusions were observed for higher level components (~1250 rfu). Extreme caution should be exercised in interpretation when similar conditions are or may be present.*





E.g. *During validation, LRs associated with false inclusions were observed to be as high as 1,000,000 in mixtures with high levels of allele sharing and up to 100,000 for other mixture samples. Higher non-contributor LRs may occur in casework. This caveat shall be presented whenever an analyst is providing testimony that culminates in a LR.*

The plots in Figure D1 can help inform the limits of STRmix™, particularly the lower limit of DNA where an H_p true hypothesis results in a *LR* greater than 1 and the limit where false positives may arise (a *LR* greater than 1 where H_d is true).

false inclusions or exclusions were observed for single source samples. Because validation samples are specifically chosen to create mixtures with varying alleles, it is expected for casework samples to show a slightly larger range of false inclusions and exclusions.

Standard Operating Procedures

<u>Interpretability</u>

- If samples under comparison contain a partial profile, for example as a result of allele drop-out, stochastic effects, or an incomplete profile from locus dropout due to inhibition or degradation, or is a complex mixture, the DNA profile may or may not be interpretable and may be considered unsuitable for comparison.

- The following scenarios may be considered to determine if a DNA profile or portion of a DNA profile is unsuitable for comparison. (Note: this does not cover all possible scenarios):
  a. Data of limited or poor quality
  b. Mixture profiles or portions of a mixture profile where the presence of allelic drop-out is reasonable.
  c. Profiles or portions or profiles that exhibit excessive homozygosity
  d. Samples where the number of contributors cannot be determined
  e. Complex mixtures (e.g., >4 contributors, allele sharing between multiple contributors, drop-out…)

ANSI/ASB 040

4.2.4 The limitations of the interpretation methods used such as characterizing and defining the maximum number of contributors, and issues associated with low-level data, low-level contributors and potential contamination events.

4.2.5 Criteria for defining what are interpretable data versus data that cannot be interpreted.

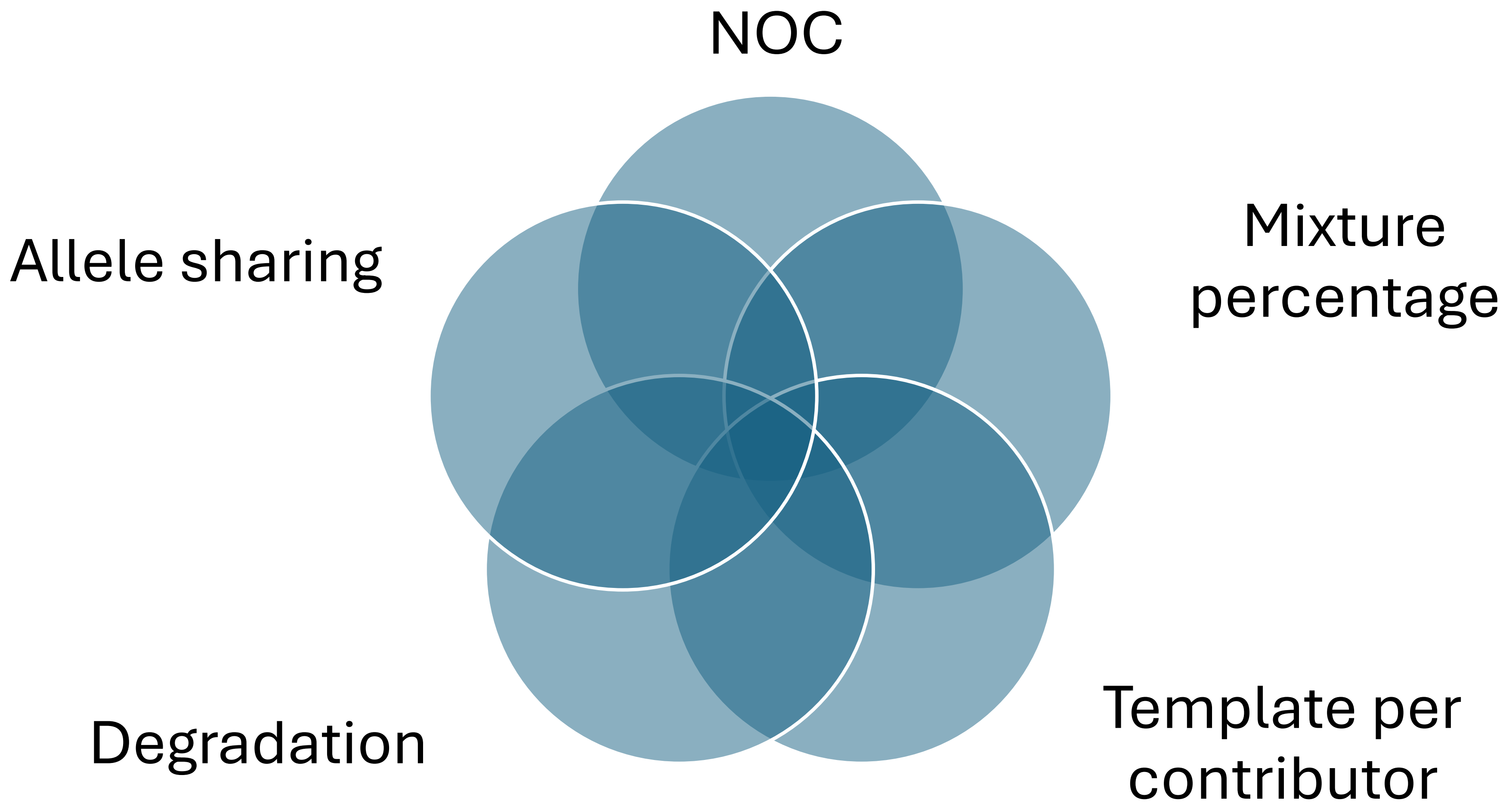
4.2.6 Criteria for defining data that are suitable for comparison versus data that are unsuitable for comparison.

Precedent for including boundaries and limitations in SOPs

2. When determining the number of contributors to a mixture, total allele number in the mixture and more discriminating loci should be considered.
3. The number of contributors to a mixed sample may be inferred based on the locus that exhibits the greatest number of allelic peaks. Counting the total number of alleles detected at the autosomal loci may provide guidance towards defining a minimum or finite number of contributors present in the mixture.
4. When using the total allele count to determine a finite number of contributors, the counting of the autosomal alleles assumes that the mixture profile has no alleles that are below the analytical threshold and therefore undetected. Also, the allele count does not take into consideration possible genotype combinations or modelling of the DNA profile (e.g. stutter or drop-in) with probabilistic genotyping.
5. If there is reason to believe that there may be undetected alleles (e.g., possibility of inhibition or degradation for one or more contributors, the possibility of stochastic effects or drop-out), counting of the autosomal alleles might only be useful in determining the minimum number of contributors.
6. Figure 7 illustrates the number of total alleles observed in two, three, and four person mixtures simulated from known samples typed with PowerPlex Fusion 5C. This data does not include allele counts for the SE33 locus. A mixture with a total of 68 autosomal alleles is more likely to consist of only two contributors than to consist of three or more contributors. A mixture with a total of 87 autosomal alleles is more likely to consist of only three contributors than consist of two or four contributors.

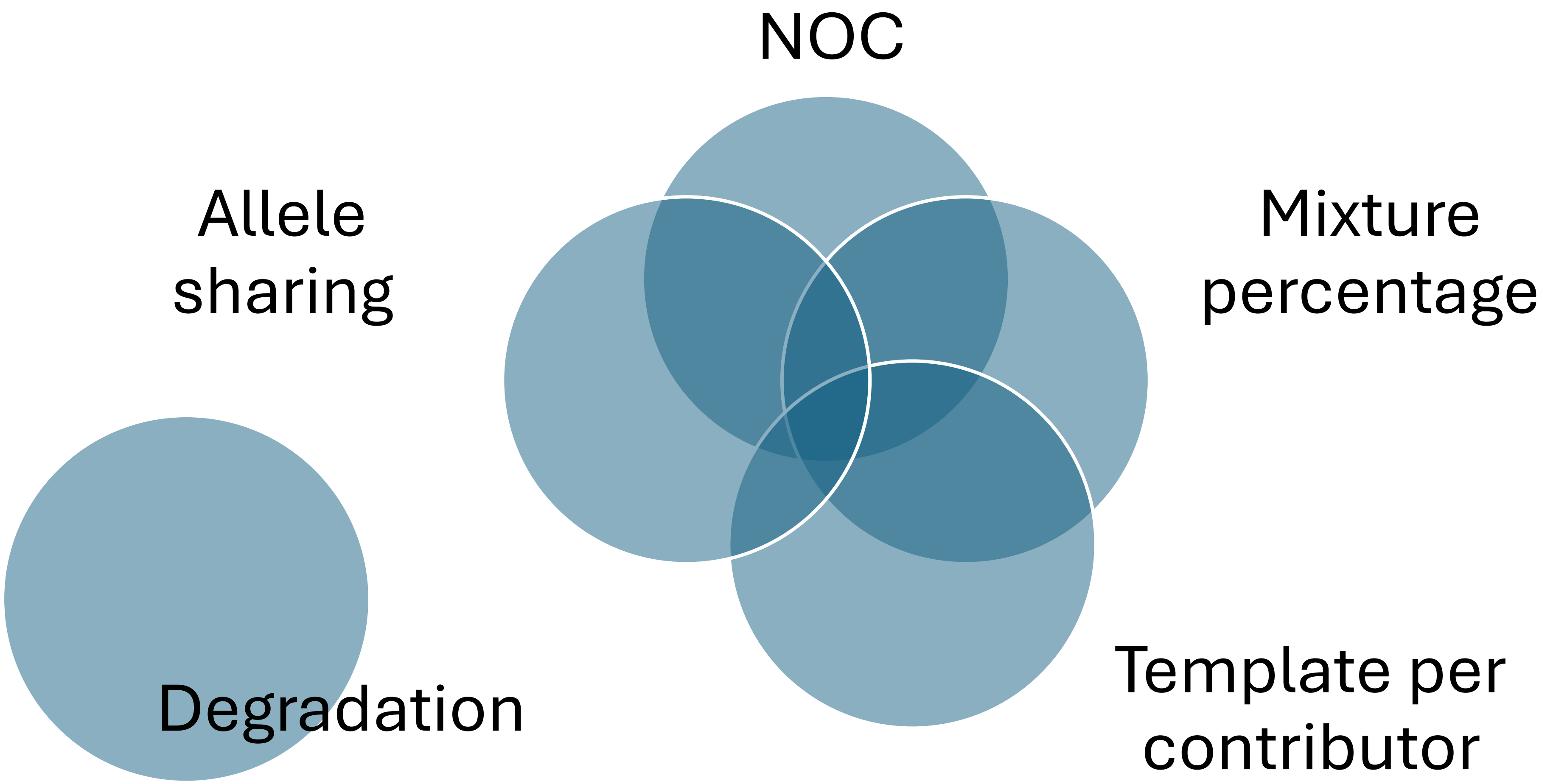
How can we do better?

- SOPs that clearly define areas of out-of-bounds (i.e. clearly describe factor space)
- SOPs that clearly communicate limitations within tested factor space



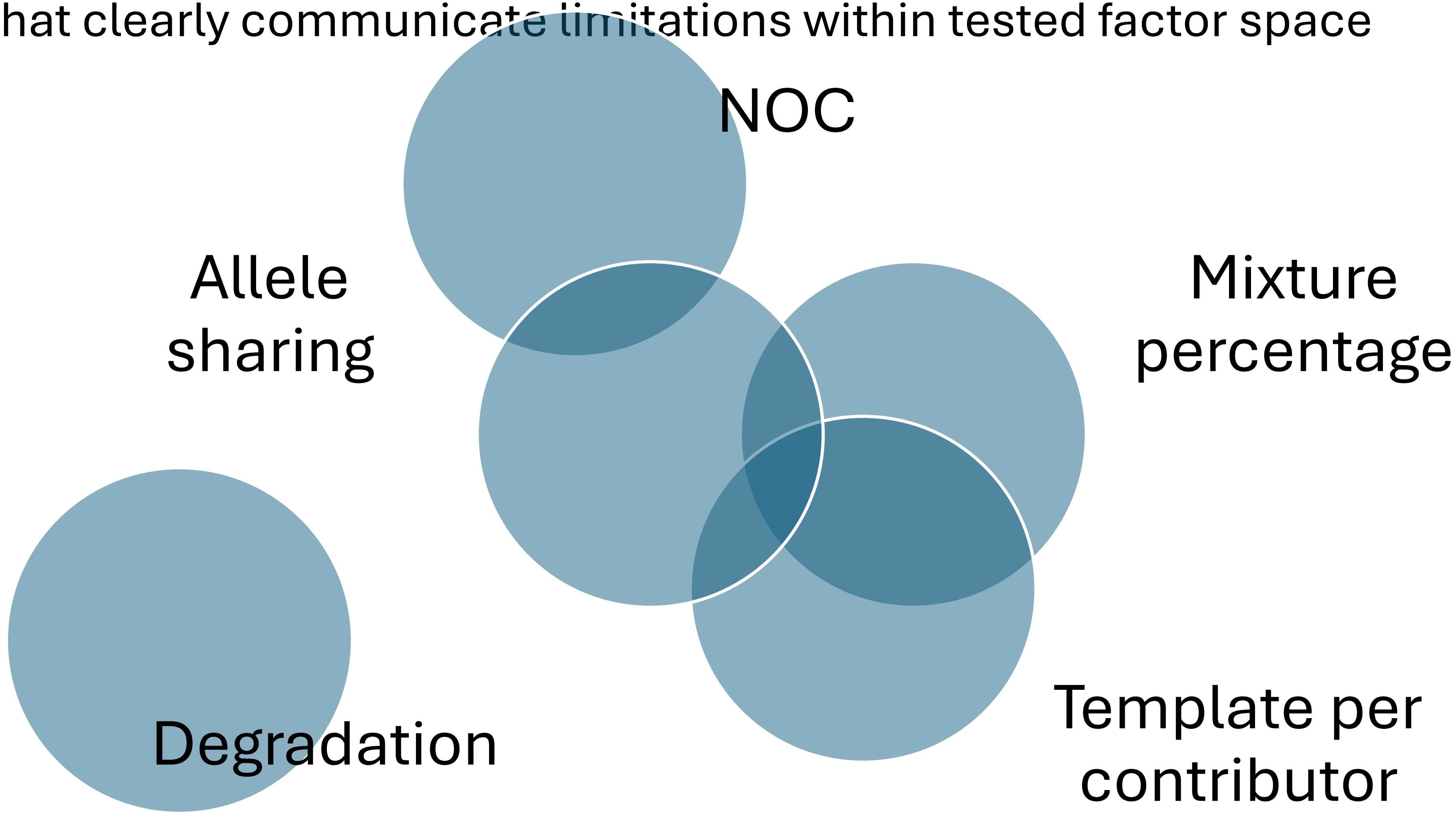
How can we do better?

- SOPs that clearly define areas of out-of-bounds (i.e. clearly describe factor space)
- SOPs that clearly communicate limitations within tested factor space



How can we do better?

- SOPs that clearly define areas of out-of-bounds (i.e. clearly describe factor space)
- SOPs that clearly communicate limitations within tested factor space



Statistical Theory for Likelihood Ratios in Forensic Analysis

Daniel Rabinowitz

Department of Statistics
Columbia University

Notation

Individual(s)

D	**D**efendant
S	**S**ource
A	Perpetrator of the alleged **A**ct

Their characteristic(s)

$$\theta_D, \;\; \theta_S \text{ distributed according to } \pi(\theta), \;\; \theta_A$$

Data

Y_1	evidence about θ_S modeled via $f^{\theta_S}(y)$
Y	all the rest of the evidence

Facts of Consequence to Culpability

$$\theta_S = \theta_D \;\longrightarrow D = S \;\longrightarrow\; D = A$$

State's Presentation of Y_1

Likelihood Ratio

- State $D = S$
  - operationally: $Y_1 \sim f^{\theta_D}(y)$
- Defense $D \neq S$
  - operationally: $Y_1 \sim \int \pi(\theta) f^{\theta}(y) d\theta$
  - Not $\theta_D \neq \theta_S$ because want to compute LRs?
  - or because point null embedded in composite alternative?
  - Turing's rule and "protection" on average.

State's Witness Reports

$$LR = \frac{f^{\theta_D}(Y_1)}{\int \pi(t) f^t(Y_1) dt} = \pi(\theta_D | Y_1) / \pi(\theta) \text{ is large,}$$

"providing strong evidence" for ... the State's hypothesis.

Strength of Evidence for State's Hypothesis

Probability depends on LR and target, prior

$$\frac{P\{\text{target}\} \times LR}{P\{\text{target}\} \times LR \,+\, (1 - P\{\text{target}\})}$$

$P\{\theta_S = \theta_D \mid Y_1\}$

$$\frac{\pi(\theta_D) \times LR}{\pi(\theta_D) \times LR \,+\, (1 - \pi(\theta_D))} \quad \text{(miniscule)}$$

or $P\{S = D \mid Y_1,\, Y\}$

$$\frac{P\{S = D \mid Y\} \times LR}{P\{S = D \mid Y\} \times LR \,+\, (1 - P\{S = D \mid Y\})}$$

"Providing strong evidence for the prosecution's hypothesis" a *de facto* (and unusual) instruction about the standard for culpability? Also, implicitly, testifying beyond personal knowledge.

State's Finder-of-Fact's Calculation

$P\{S = D \mid Y_1,\ Y\}$

$$P\{S = D \mid Y_1,\ Y\} = \frac{P\{S = D \mid Y\}LR}{P\{S = D \mid Y\}LR + (1 - P\{S = D \mid Y\})}$$

$$P\{S = D \mid Y\} = (a_D + b_D) \Big/ \sum_i (a_i + b_i)$$

followed by $P\{A = D \mid Y_1,\ Y\}$

$$\begin{aligned} P\{A = D \mid Y_1,\ Y\} &= P\{S = D \mid Y_1,\ Y\}P\{A = D \mid S = D,\ Y\} \\ &\quad + (1 - P\{S = D \mid Y_1,\ Y\})\frac{a_D}{(a_D + b_D)} \end{aligned}$$

How helpful for the purposes of FRE 702(a) is it to report the LR without explaining the subsequent calculations?

As Applied Challenges to

"providing strong evidence for" the hypothesis

$$\sum_i (a_i + b_i) \text{ is large,}$$

$$\text{or most } (a_i + b_i) \text{ are larger}$$

the Model for θ_S and Y_1

$$\text{(Posterior probability given } Y_1 \text{ that) } \theta_S \text{ is consistent with } \pi(\theta)$$

$$Y_1 \text{ is consistent with } \int f^{\theta}(y)\pi(\theta)d\theta$$

Validation

- Leverage statistics measuare "edge" not "hole."

Issues

- Power (sample size, State's quantile and the Confrontation Clause)
- Burden Shifting and Multiple testing
- Statistical versus practical significance

Reporting Uncertainty

Whose burden?

- Standard in *the discipline*: burden on analyst to report uncertainty ... *including* uncertainty in assessments of uncertainty.

Uncertainty in LR

- Propagate standard errors.
- Sensitivity analyses for assumptions. Subjective priors are judicial admissions of uncertainty?
  - random, known distribution
  - random, uncertain distribution
  - not random, subjective beliefs about distribution
  - randomness as a deliberate fiction because
    - only* Bayes' rules are admissible (as statistics term of art),
    - need Bayes factor to have an LR.

  "approximations" and "restrictions" are assumptions, too.
- Validation of π and f^θ to complement "black box" studies.

Using a validation study model to minimize wrongful convictions

J.D. Schmid

What can the court system do to adequately ensure that DNA evidence is not being used to convict innocent people?

Forensic Science International: Genetics 17 (2015) 155–162

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

ELSEVIER

Secondary and subsequent DNA transfer during criminal investigation

CrossMark

Ane Elida Fonneløp [a,c,*], Thore Egeland [a,b], Peter Gill [a,c]

[a] Norwegian Institute of Public Health, Oslo, Norway
[b] IKBM, Norwegian University of Life Sciences, Ås, Norway
[c] University of Oslo, Oslo, Norway

- "With the introduction of new multiplexes and instrumentation…there has recently been a rapid change in the technology that has greatly increased sensitivity of detection so that a DNA profile can routinely be obtained from only a few cells. Research to assess the risks of passive transfer has not kept pace with this development; hence the 'hidden' risk of innocent DNA transfer at the crime-scene is currently not properly understood."

- 2019 AAFS Annual Meeting, Workshop 10, DNA Mixture Interpretation Principles: Observations From a National Institute of Standards and Technology (NIST) Scientific Foundation Review (Sheila Willis presentation)

- Peter Gill, Misleading DNA Evidence (2014)

- NIST DNA Mixture Interpretation: A Scientific Foundation Review (Draft Report 2021) – Chapter 5

NISTIR 8351-DRAFT

DNA Mixture Interpretation:
A NIST Scientific Foundation Review

Key takeaway 5.4 "DNA statistical results such as a subsource likelihood ratio do not provide information about how or when DNA was transferred, or whether it is relevant to a case. **Therefore using the likelihood ratio as a stand alone number without context can be misleading.**"

DNA Evidence & Innocent Suspects

- Farah Jama
- Lukis Anderson
- Adam Scott
- Kevin Brown
- Amanda Knox/Raffeale Sollecito

"For every error discovered, there are an unknown number that are undiscovered." Gill, Misleading DNA Evidence, at p. 21.

OPTIONS FOR THE COURTS

1. Do nothing
2. Allow expert opinion testimony on activity-level propositions
3. Rely on procedural safeguards short of exclusion, i.e., cautionary jury instructions, Defense expert testimony
4. Exclude DNA evidence in situations where the case circumstances demonstrate an unreasonable risk that the DNA is not associated with the crime.

What can the court system do to adequately ensure that DNA evidence is not being used to convict innocent people?

- **Criminal Justice System**

DNA Evidence

GOALS

Distinguish true positives from true negatives

Laboratory: Distinguish between true contributor and true non-contributor

Criminal Justice System: Distinguish between guilty and innocent

METHOD

<u>Laboratory</u>: Standard Operating Procedures

<u>Criminal Justice System</u>: Constitutional Rules, Rules of Procedure, Rules of Evidence, Statutes, Case Law

Values

Criminal Justice System:

"[A] fundamental value determination of our society [is] that it is far worse to convict an innocent man than to let a guilty man go free." In re Winship, 397 U.S. 358, 372 (1970)(Harlan, J. concurring)

"The maxim of the law is … that it is better that ninety-nine … offenders should escape, than that one innocent man should be condemned." Schlup v. Delo, 513 U.S. 298, 325 (1995)

Does the method work?

Laboratory: Validation Studies

Criminal Justice System: ???

Red Flags

Wrongful Conviction Research:

Brandon Garrett, Judging Innocence (2008)

"The Elephant in the Room" workshop series published in the Albany Law review (2016)

John Morgan, *Forensic Testimony Archeology: Analysis of Exoneration Cases and its implications for forensic science testimony and communications* (2023)

Red Flags

Gross, et al., *Rate of false convictions of criminal defendants who are sentenced to death,* National Academy of Sciences (2014)

- estimated that at least 4.1% of inmates sentenced to death would be exonerated if the execution was indefinitely delayed.
- Expressed that death cases are not representative of all of the cases in the Criminal Justice System and that the 4.1% was likely an underestimate of the actual rate of wrongful convictions

Red Flags

Jury composition:

"The evidence regarding the impact of the jury pool is straightforward and striking…[I]n cases with no blacks in the jury pool, black defendants are convicted at an 81% rate and white defendants at a 66% rate. When the jury pool includes at least one black potential juror, conviction rates are almost identical: 71% for black defendants and 73% for white defendants."

Anwar, et al., *The impact of jury race in criminal Trial*, 127 Quarterly Journal of Economics 1017, (2012)

Red Flags

Limiting Instructions:

Deliberating groups were obedient to judge's limiting instruction concerning prior convictions but convicted persons at a higher rate when they knew about a person's prior record

Vidmar & Hans, American Juries: the Verdict, p. 162 (2007)

Red Flags

"The reliability test adopted in Rule 702 appears, at least in written appellate opinions, to be rarely used in practice to test reliability and, when used, it tends to exclude defense witnesses."

Brandon L. Garrett & M. Chris Fabricant, *The Myth of the Reliability Test*, 86 Fordham L. Rev. 1559, 1581 (2018).

Red Flags

"To be blunt: expert testimony in civil cases is habitually and stringently assessed under the Daubert factors. The same cannot be said of expert testimony in criminal cases. Rather, criminal cases favor admissibility over a rigorous assessment of reliability (the so called 'weight v. admissibility' argument)."

Jessica G. Cino, *An Uncivil Action: Criminalizing Daubert in Procedure and Practice to Avoid Wrongful Convictions*, 119 W. Va. L. Rev. 651, 656 (2016)

Red Flags

"An analysis of post-Daubert decisions demonstrates that whereas civil defendants prevail in their Daubert challenges, most of the time criminal defendants almost always lose their challenges to government proffers. But when the prosecutor challenges a criminal defendant's expert evidence, the evidence is almost always kept out at trial."

Peter J. Neufeld, *The (Near) Irrelevance of Daubert to Criminal Justice and Some Suggested Reforms*, 95 Am. J. Pub. Health S 107, S109 (2005)

OPTIONS FOR THE COURTS

1. Do nothing
2. Allow expert opinion testimony on activity-level propositions
3. Rely on procedural safeguards short of exclusion, i.e., cautionary jury instructions, Defense expert testimony
4. Exclude DNA evidence in situations where the case circumstances demonstrate an unreasonable risk that the DNA is not associated with the crime.

VALIDATION

1. Ground truth samples
2. Samples bear the characteristics of what we expect to see in case work
3. Samples should account for all of the variables that can impact the accuracy of a conclusion in casework
4. The study should determine what circumstances cause the method to fail
5. Higher degrees of validation are required when the consequence of a failure is high

BENEFITS OF MOCK JURY STUDIES

- We can use fact scenarios where that approximate ground truth.
- We can control mock jurors' exposure to different conditions
- We can assess whether specific procedures help mock jurors get the correct result

Future research

1. Look at circumstances where the system failed or where the recovery of Irrelevant DNA could very easily have caused the system to fail under slightly different circumstances. (Annie Le)

2. Create fact scenarios based on demonstrated instances of indirect transfer in the lab or TPPR studies

Contact information

- Email: schmid.jd@gmail.com
  john.schmid@pubdef.state.mn.us

- Phone: 218-349-1372 (cell)
  218-302-8823 (office)

Guidance documents for evaluative reporting in forensic science: European developments

Communicating Forensic Findings Workshop: Current Practices and Future Directions

Alex Biedermann
University of Lausanne (Switzerland)
Faculty of Law, Criminal Justice and Public Administration

25th and 26th June, 2024
National Institute of Standards and Technology (NIST)
National Cybersecurity Center of Excellence (NCCoE)

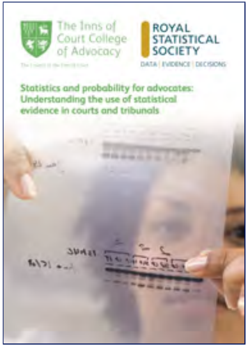
UNIL | Université de Lausanne

What will I talk about?

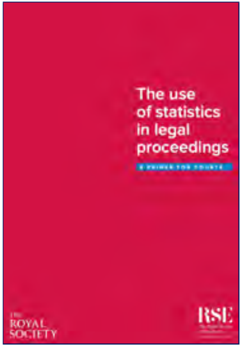
Practitioner guidance for evaluative thinking (2010–2020)

Royal Statistical Society (2010)

European Network of Forensic Science Institutes (ENFSI, 2015)



Inns of Court College of Advocacy & Royal Statistical Society (2017)



The Royal Society & Royal Society of Edinburgh (2020)

The ENFSI Guideline for Evaluative Reporting in Forensic Science:

What is the institutional background and context of development of the ENFSI Guideline?

ENFSI Board: “Statement regarding interpretation”

Ref code: BRD-GEN-004 issued 6th July 2010 , Annual Meeting Prague

Statement

ENFSI wishes to promote consistent and reliable scientific information through the whole forensic process from the scene of crime to court. It recognizes that it is of vital **importance to interpret** potential forensic evidence in the **context of a case**, using **all available information***; **reporting results of analysis alone may be misleading**.

The ENFSI Board plans to undertake actions to **agree a standard for the interpretation** of forensic evidence and to provide the necessary **support for implementing** this standard within the membership.

*Today we would say **task-relevant information**

Direct Grants – “Monopoly Programmes”

ENFSI’s Monopoly Position and Action Grant

In 2009 the **European Commission** has decided to grant ENFSI the **monopoly position** concerning forensic science in Europe. As a result of this decision the EC allocated a specific **action grant** for ENFSI to spend on project work.

ENFSI’s “M1 Project” Sheila Willis (Principal Investigator)

Developing and implementing a **standard** for the **evaluation of forensic evidence** whilst providing some **educational support** in statistics for practitioners across Europe.

Dr. Sheila Willis*

*Image source: https://x.com/sideliner2

In what sense is the ENFSI Guideline more than just a guideline?

More than just a guideline: Roadmap towards change

[Roadmap]

It is recognized that the implementation of the *Guideline for evaluative reporting* is a challenge in itself and below is proposed the key elements of a roadmap that should help laboratories in this task.

Step 1
Managing the change

- Identifying **key personnel responsible** for the implementation
- Deciding on a **strategy** to approach each forensic discipline covered by the laboratory (focus groups, leaders in each iscipline, etc.)
- Adopting a **project plan** with defined objectives and timeline

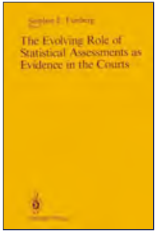
Step 2
Training

- Providing **training** a guideline (i.e. fram ircumstance, prop ratio, workshops pe
- Identifying what is **reports** (compared investigative report
- Training should inc **testing**.
- Providing **informat** stakeholders (e.g. p judiciary, mandatin to the changes asso guideline in particu information at the the reporting pract

ENFSI Guideline (2015, p. 124)

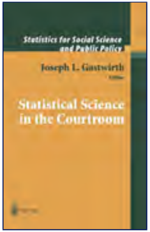
What is the theoretical background of the ENFSI Guideline?

Statistics and the Law & Forensic Inference and Statistics

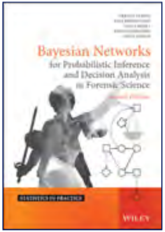


Fienberg (Ed.) (1989)

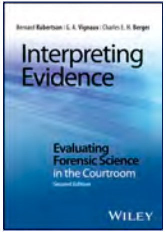
Aitken/Stoney (Eds., 1991)



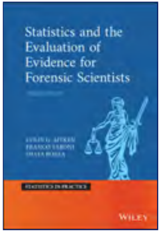
Gastwirth (Ed.) (2000)



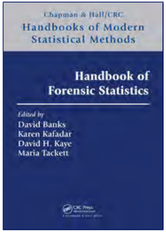
Taroni et al. (2014)



Robertson et al. (2016)



Aitken et al. (2020)



Banks et al. (2021)

*Image sources: CRC Press, Ellis Horwood, John Wiley & Sons, Springer

How original is the ENFSI Guideline?

How original is the ENFSI Guideline?

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Standards for the formulation of evaluative forensic science expert opinion

AFSP Standard (2009)

in: Science & Justice, 49, 2009, 161–164.

ENFSI Guideline

Additional features: Guidance notes, Case examples, more EU-compatible.

What is the central conceptual challenge that the ENFSI Guideline is trying to address?

The challenge of practical application

- "(...) **no mathematical result is self-applying**, and **additional argument is necessary to <u>bridge the gap</u>** from a general mathematical truth to a substantive application – in law as in any other domain." [1]

Prof. David H. Kaye*

[1] Kaye D. H. 1999, **Clarifying the burden of proof: what Bayesian decision rules do and do not do**, International Journal of Evidence & Proof, 3, 1–28.
*Image source: https://pennstatelaw.psu.edu/faculty/kaye

$\Rightarrow$ ...it's all about asking the relevant questions

Is the ENFSI Guideline “Bayesian”?

The ENFSI Guideline is about inference

Inference: the reasonable reasoning under uncertainty

Asking the relevant question

"(...) the **single most important advance** has nothing to do with technology (...). It tells us the most important lesson for the **logic of evaluative forensic science**: **consider the probability of the evidence, given the proposition**."

Dr. Ian W. Evett CBE*

I.W. Evett, Evaluation and professionalism, Science & Justice 49 (2009) 159–160, at p. 159
* Image source: https://www.principalforensicservices.com

Why are the recommendations in the ENFSI Guideline fundamental?

What is the structure of the ENFSI Guideline?

What is the structure of the ENFSI Guideline?

Document structure

1. Scope
2. Evaluative reporting
3. Standard framework
4. Guidance notes
5. Glossary

13 pages

References
Case examples
Audit template

4.1 Reporting requirements
4.2 Propositions
4.3 Data used to assess the strength of the findings
4.4 Meaning of the LR in an evaluative report

What is the scope of ENFSI Guideline (Chapter 1)?

What is the scope of the ENFSI Guideline?

Domain of application: What is evaluative reporting?

- The Guideline applies **only to evaluative reports for use in courts**.
- **Not covered**: Investigative, intelligence and technical reporting.

<u>What</u> is evaluative reporting and <u>when</u> should an evaluative report be produced (Chapter 2)?

What is evaluative reporting and when should an evaluative report be produced?

“**Evaluative reports for use in court** should be produced when two conditions are met:

1. The forensic practitioner has been asked by a **mandating authority** or **party** to examine and/or **compare material** (typically recovered trace material with reference material from known potential sources)
2. The forensic practitioner seeks to **evaluate findings with respect to** particular competing **propositions** set by the specific case **circumstances** or as indicated by the mandating authority.” (p. 6)

⇒ Section 2 of the ENFSI Guideline

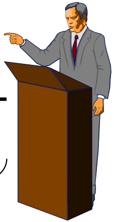
The core of an evaluative report: assessment of the value of the findings

Assessing the **value of the findings, <u>not</u> opining on the competing propositions**

- “Evaluation (...) is based on the assignment of a **likelihood ratio**.
- The likelihood ratio measures the strength of support the findings provide to discriminate between propositions of interest.” (at p. 6)

Key element of the ENFSI guideline

Probability ratio *(likelihood ratio)* = $\dfrac{\text{Probability of result } \underline{\textbf{if}} \text{ proposition 1 is true and the information.}}{\text{Probability of result } \underline{\textbf{if}} \text{ proposition 2 is true and the information.}}$



<u>focus on:</u> « the **probability of the findings <u>given</u> the propositions and the information** – and not the reverse »

What is meant by the “Standard Framework” (Chapter 3) in the ENFSI Guideline?

The “Standard Framework”

Questioning the evaluation process: Three notions to understand the standard framework

1. **(Key) issues**
2. **Propositions**
3. **Hierarchy of Propositions**

What are key issues? (Glossary, p. 21)

“The **key issue(s)** represent those aspects of a case on which a Court, under the law of the case, seeks to reach a judgement. The key issue(s) provide the general framework within which **requests** to forensic practitioners and **propositions** (for evaluative reporting) are **formally defined**.”

⇒ Evaluation is a **process**, and you may want to **question** that process critically.

What do the “Guidance Notes” of the ENFSI Guideline say (Chapter 4)?

What do the “Guidance Notes” of the ENFSI Guideline say?

4 Guidance Notes:

1. Reporting requirements: Balance, robustness, transparency and logic
2. Propositions
   - How to choose a level in the hierarchy?
   - What if no proposition(s) is (are) available?
   - When and how to change propositions?
3. Data and expert knowledge used to assess the strength of the findings and assignment of likelihood ratios
4. Meaning of the likelihood ratio in an evaluative report
   - Example of a verbal scale

What does the ENFSI Guideline say about data and expert knowledge used to assess the strength of the findings and assignment of likelihood ratios?

Data and expert knowledge for likelihood ratio assignment

ENFSI Guideline: transparency in probability assignment

- “(…) **probability assignment** is **not** arbitrary or speculative, but is based on a **body of knowledge that should be available for auditing and disclosure**.” (p. 16)
- **Data “hierarchy”**: “Such data can take, for example, the structured form of **scientific publications**, **databases** or **internal reports** or, in addition to or in the absence of the above, be part of the expert knowledge built upon **experiments** conducted under controlled conditions (including case-specific experiments), **training and experience**.” (p. 19)

What are the benefits of the ENFSI Guideline and what are the prospects for change?

Benefits and perspectives of change

Conclusions

- Principles that are well established and reasonably practicable
- Resistance

  limited resources (reduction of workflow)

  adherence to traditional modes of thinking
- Keeping the status quo also has a cost
- Reduction and control of the points of attack of the forensic expert's work: change is in the scientists' own interest

Thank you for your attention.

UNIL | Université de Lausanne

When, how, and for whom?

Polisen

Anders Nordgaard, PhD, LL.D h.c., forensic specialist

Swedish Police Authority – National Forensic Centre (NFC)

NIST Workshop, Gaithersburg, Md June 25-26, 2024

Outline

- Categorizing casework – when is evaluative reporting needed?
  - Sources of uncertainty
  - Clarifying the forensic questions
- Simplifying the expression of the value of evidence – scales of conclusions
  - Robust assignment
- Training forensic experts, investigators, prosecutors, judges and defence attorneys
  - Nordic law and normative framework
  - Who is the commissioner of forensic investigations?

Categorizing casework – when is evaluative reporting needed?

Sources of uncertainty (in forensic investigations)

Category 1

- accuracy (measurement uncertainty)
- handling of material (contamination, mixing-up etc.)
- human factor in general

Such sources should normally not motivate evaluative reporting

- Measurement uncertainty is typically provided in technical reports
- Contamination, mixing-ups, human factor error are all unacceptable errors and <u>should not</u> be quantified and accompany a forensic conclusion – Should be handled by the quality assurance system.

Category 2

- only a (random) sample of the seized material is analysed

Typical in screening analysis (e.g. identifying analysis of suspected drug material).

Conclusions should be accompanied with a statement of uncertainty reflecting the sampling error (“With 99% probability 50% of the consignment consists of Ecstasy pills”)

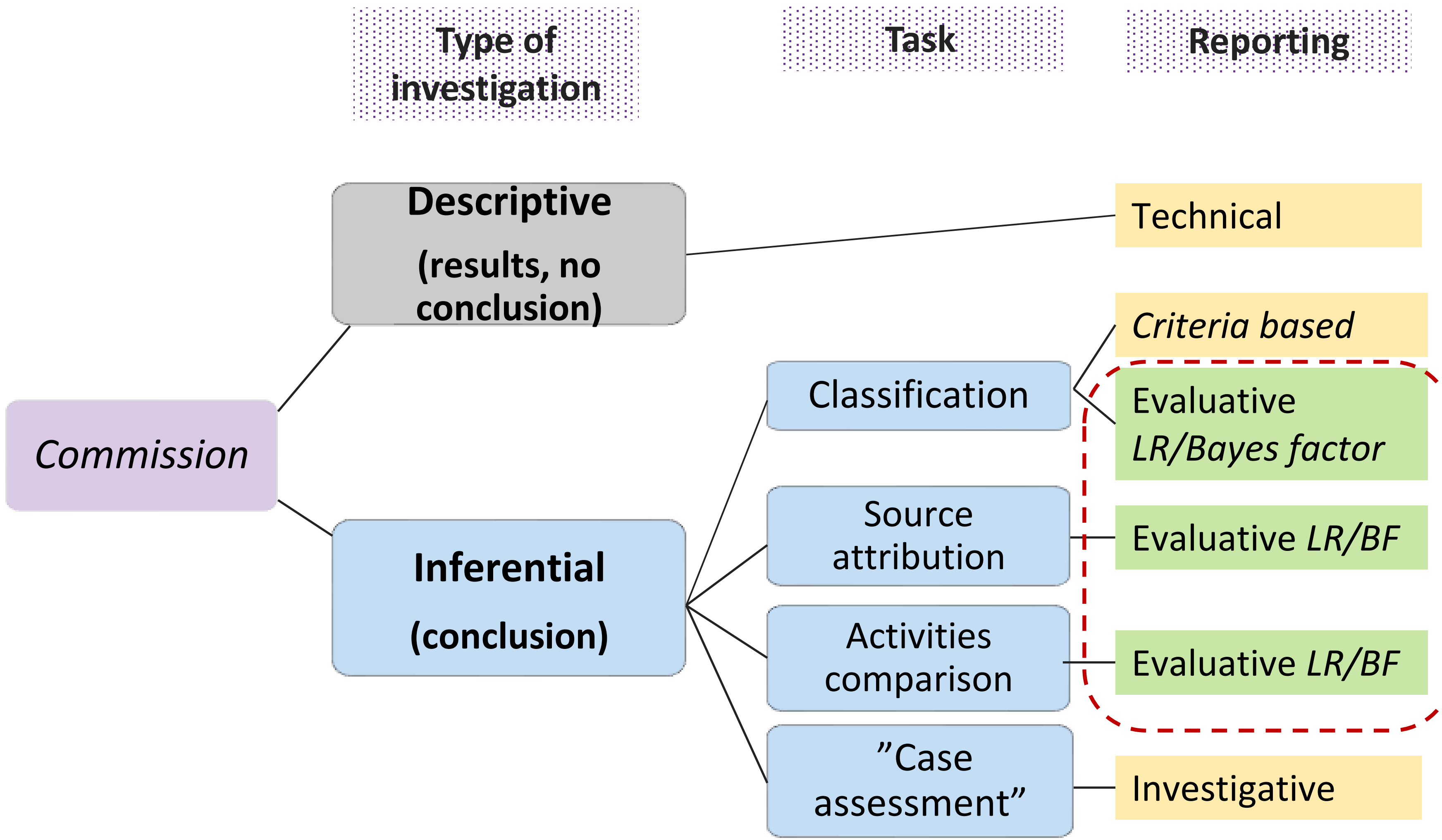
Category 3

- rarity/commonness in general of the characteristics observed/analysed
- mechanisms of transfer, persistence and background levels of material

This is the dominating source of uncertainty when the forensic question is about source attribution, competing activities or classification with no established criteria.

Requires evaluative reporting!

Clarifying the forensic questions (at NFC)



Examples

Descriptive tasks:

- What substances and with which concentrations can be found in the fire debris?
- How long is the person on the CCTV uptake?

Classification tasks with criteria-based reporting:

- Is the document a genuine Swedish passport?
- Is the electronic equipment a jammer?

Classification tasks with evaluative reporting:

- Do the fire debris contain (traces of) ignitable liquids?

Source attribution:

- Were the two scratch marks made with the same tool?
- Was part of the graffiti paint made with the spray can found with the suspect?

Activity comparison/attribution:

- Was the suspect’s pullover recently in contact with the car seat?
- Did the suspect kick the victim in the face or was he just standing aside?

Simplifying the expression of the value of evidence – scales of conclusions

The likelihood ratio (LR) (or Bayes factor (BF)) is a component of Bayes' theorem.

There are two competing hypotheses (propositions) addresses in a case:

The main hypothesis, $\boldsymbol{H_m}$ (usually forwarded by the prosecution)

The alternative hypothesis, $\boldsymbol{H_a}$

The forensic findings, $\boldsymbol{E}$ should be evaluated against these two hypotheses.

Bayes' theorem on odds form:

$$\frac{P(\boldsymbol{H_m}|\boldsymbol{E})}{P(\boldsymbol{H_a}|\boldsymbol{E})} = (\mathbf{LR/BF}) \times \frac{P(\boldsymbol{H_m})}{P(\boldsymbol{H_a})}$$

$$\frac{P(H_m|E)}{P(H_a|E)} = (\mathrm{LR/BF}) \times \frac{P(H_m)}{P(H_a)}$$

Polisen

In many accounts of forensic interpretation, the likelihood ratio is given as

$$\mathrm{LR} = \frac{P(E|H_m)}{P(E|H_a)}$$

"The probability of obtaining the forensic findings if H_m is true divided by the probability of obtaining the forensic findings if H_a is true."

However, a likelihood is not by necessity a probability. If the forensic findings are quantified on a continuous scale, *probability density functions* must be used as probative measure. Hence, a more general definition is

$$\mathrm{LR} = \frac{\mathcal{L}(H_m; E)}{\mathcal{L}(H_a; E)}$$

"The likelihood of H_m in light of the forensic findings divided by the likelihood of H_a in light of the forensic findings."

Moreover, likelihoods are defined for a particular value of a parameter or a simple hypothesis. If one or both hypotheses involved are composite, it is no longer valid to use the term likelihood ratio for the value of evidence.

The general expression for the *Bayes factor* as value of evidence is

$$BF = \frac{\sum_i \mathcal{L}(\boldsymbol{H_{m,i}}; \boldsymbol{E}) \times P(\boldsymbol{H_{m,i}} | \boldsymbol{H_m})}{\sum_k \mathcal{L}(\boldsymbol{H_{a,k}}; \boldsymbol{E}) \times P(\boldsymbol{H_{a,k}} | \boldsymbol{H_a})}$$

where $\boldsymbol{H_m} = \bigcup_i \boldsymbol{H_{m,i}}$ and $\boldsymbol{H_{m,i}} \cap \boldsymbol{H_{m,i'}} = \emptyset \ \forall\, i \neq i'$

$\boldsymbol{H_a} = \bigcup_k \boldsymbol{H_{a,k}}$ and $\boldsymbol{H_{a,k}} \cap \boldsymbol{H_{a,k'}} = \emptyset \ \forall\, i \neq i'$

Can be a challenge to assign!

Polisen

But whether we address LRs or BFs, constructing a scale of conclusions is not about giving an absolute interpretation of neither of them.

Polisen

Focus on the potential posterior probabilities (or odds) they would lead to under different settings of the prior odds.

One possibility is to take as an average setting the maximum entropy.

Likelihood Ratio scale

Maximum prior entropy would be even odds, i.e. $P(\boldsymbol{H_m}) = P(\boldsymbol{H_a}) = 0.5$

Assuming exhaustive hypotheses, the posterior probability will be

$$P(\boldsymbol{H_m}|\boldsymbol{E}) = \frac{LR}{LR + 1}$$

Decide upon how many $\boldsymbol{H_m}$-supporting levels you wish to have in the scale.

For each level (or a subset of levels), decide upon a sufficiently high posterior probability – reflecting end-users' interpretation of levels of probability with respect to their appreciation of evidentiary strength.

With even prior odds you can for each level deduce the LR as

$$P(\boldsymbol{H_m}|\boldsymbol{E}) = \frac{LR}{LR+1} \qquad LR = \frac{P(\boldsymbol{H_m}|\boldsymbol{E})}{1 - P(\boldsymbol{H_m}|\boldsymbol{E})}$$

At NFC, we chose 4 levels (in 2004): +1, +2, +3, +4

For +2 we decided that this is a level that with even prior odds should give a posterior probability of at least 99%.

99% is a probability level for which there is a wide-spread acceptance among legal professionals that something is corroborated (however, not proven).

+2 should thus – if prior odds are even or higher - be sufficient for detention.

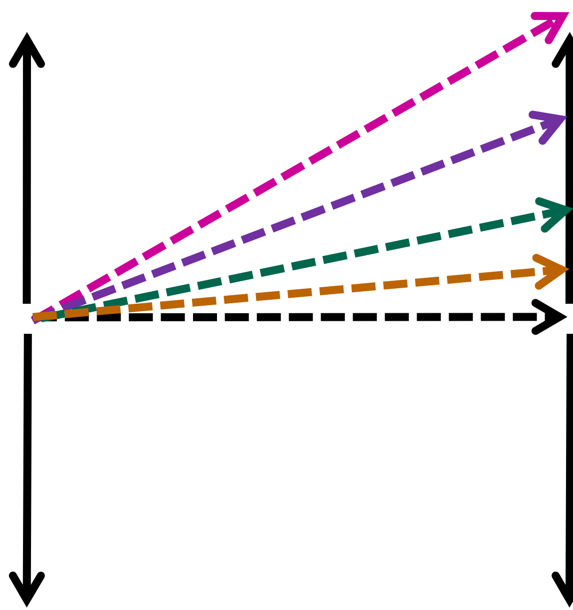
For the highest level, we set the LR to be at least one million – this magnitude was at the time a lower limit for a full match in DNA (siblings excluded).

With even prior odds a LR of one million gives a posterior probability of 0.999999.

Sufficiently high for considering $\boldsymbol{H_m}$ proven beyond reasonable doubt.

Level +1 and level +3 would correspond to LRs given a regular increase of the intervals between the levels.

The rest is math and rounding-off.



Level	Posterior probability $P(H_h \mid E)$	Lower limit for LR
+4 :	**> 0.999999**	**$10^6 \leq$ LR**
+3:	**> 0.9998**	**6000 ≤ LR**
+2:	**> 0.99**	**100 ≤ LR**
+1:	**> 0.86**	**6 ≤ LR**
0:	**between 0.14 and 0.86**	**(1/6 < LR < 6)**

Scale of conclusions used at NFC:

Scale level	Magnitude of the likelihood ratio (V)	”Explanation”
		The findings are deemed…
+4	at least one million	…at least one million times more probable…
+3	between 6000 and one million	…at least 6000 times more probable…
+2	between 100 and 6000	…at least 100 times more probable…
+1	between 6 and 100	…at least 6 times more probable…
0	between 1/6 and 6	… approximately equally probable…
		…if the main hypothesis is true compared to if the alternative proposition is true
–1	between 1/100 and 1/6	…at least 6 times more probable…
–2	between 1/6000 and 1/100	…at least 100 times more probable…
–3	between 1/(one million) and 1/6000	…at least 6000 times more probable…
–4	at most 1/(one million)	…at least one million times more probable…
		…if the alternative hypothesis is true compared to if the main proposition is true

Robust assignment

It is most often the magnitude of the LR that is of interest, not its precise value.

In most forensic disciplines it is a difficult and time-consuming task to come up with a precise value of the LR – betting preferences work theoretically but are hard to imply to a community that never bets.

With a scale of conclusion, the forensic expert is instructed to report the level they are convinced is reached – but with no obligation to be more precise.

Produces a conservative report in the sense that the defence can always refer to the lower limit of an interval.

End-users *learn* successively what is a high value of evidence and what is a low value of evidence.

Training forensic experts, investigators, prosecutors, judges and defence attorneys

Polisen

Some points about Nordic Law

Nordic law is quite similar between the Nordic countries (Sweden, Denmark, Norway, Finland and Iceland)

One common thing is the free sifting of evidence – almost no evidence rules.

Sweden has a bit more adversarial system than Norway and Denmark.

NFC (and other Nordic labs) reports to the preliminary investigation with the Police (i.e. most often to the prosecutor) – not to the court.

But we must take into consideration what would be understood by the court.

Training…

Polisen

The in-lab trainee program at NFC

- (at least) 2 years of training followed by formal examination before being approved to be case responsible, sign reports and give statement of witness in court
- General part (to a large extent digital)
  - Basic module (to all personnel working to any extent with forensic evidence) – comprises one section on introductory evaluative reporting and forensic assessment
  - Add-on modules depending on function
    - Evaluative reporting for classification and source attribution task using assessed probabilities
    - Evaluative reporting for comparison/attribution of activities
    - Evaluative reporting with continuous probability distributions
    - Investigative reporting comprising evaluative steps
    - Evaluative reporting for combining evidence

Polisen

- Function-specific part
  - Casework training (incl. evaluative reporting) under supervision

Crime scene investigators

- 1.3 years trainee program at NFC preceded by one year training at a police squad and mixed-up with such training during the program
- Investigative reporting with evaluative steps (model developed at NFC)
- 4 case studies from volume crime to severe crime
- Individual examination on an individual report on one of the case-studies (examination focussing on the investigative/evaluative part)

Police investigators

- Specific activities for target groups (comprising som evaluative reporting)
- *NEW!* Evaluative reporting part of curriculum of trainee program for investigators of severe crime

Prosecutors

- 3.5 days' basic training at NFC within the prosecutor's compulsory trainee program
  - Evaluative reporting taught via a complex case study throughout the days
- 3 days' continuing training for experienced prosecutors
  - 2-3 hours of evaluative reporting focussed on evaluation against hypotheses at activity level and combination of evidence

Judges

- 3 days' basic training at NFC within the compulsory trainee program
  - 2 hours general evaluative reporting and application in different forensic disciplines
- 3 days' continuing training for experienced judges
  - 3-4 hours of evaluative reporting focussed on the role of the judge

Defence attorneys

- Upon request, 2 days' training with 2-3 hours general evaluative reporting and application in different forensic disciplines
- Specific activities (half-day courses) at regional assemblies

Netherlands Forensic Institute
Ministry of Justice and Security

Communicating LR-conclusions in forensic reports

NIST 25-26 July 2024

Marjan Sjerps, Rolf Ypma

Numerical LR

- LRs are reported in all DNA reports
- In smaller volume in other areas such as glass, automatic speaker identification, fingerprints, telecom, authorship

Special communication challenge: database search results

- Searches generate many “good” leads but also a few “false” leads
- It’s special because other evidence may be missing completely
- NFI adds text box warning in reports

Verbal LR

“The findings are far more probable* when the fragment comes from the window than when the fragment comes from some other glass object”

Verbal equivalent	LR
Approximately equally probable	1-2
Slightly more probable	2-10
More probable	10-100
Appreciably more probable	100-10,000
Far more probable	10,000-1 million
Extremely probable	> 1 million

Footnote (1)

This term is part of a standard verbal scale (the left column in the table). This scale is used when the scientist has no or insufficient numerical data to explicitly substantiate a numerical conclusion.

The selection of the specific verbal term is based on expert knowledge, experience in research and casework, etc.

To promote the transparency for the reader and the uniformity among the different experts the NFI has defined the verbal terms numerically. These definitions are expressed in orders of magnitude and are listed in the right column in the table below.

For example, the term ‘slightly more probable’ means that the probability of observing the results of the investigation is considered 2 to 10 times larger when one hypothesis is true than when the other hypothesis is true.

Footnote (2)

The conclusion expresses the evidential strength of the results regarding the hypotheses. The conclusion does not represent the probability that a particular hypothesis is true. That probability depends on other evidence and information outside the domain of forensic expertise and falls outside the scope of this report.

More information about this way of concluding is available in the professional annex “De reeks waarschijnlijkheidstermen van het NFI”. This annex is, among others, available through the NFI website www.forensischinstituut.nl .

Education

- Reports include link to professional annex
- Courses / e-learnings for
  - police
  - prosecutors
  - judges
  - defence lawyers
- Infographics are being developed

Nederlands Forensisch Instituut
Ministerie van Veiligheid en Justitie

Vakbijlage

De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs

Inhoudsopgave

1. De vakbijlage algemeen

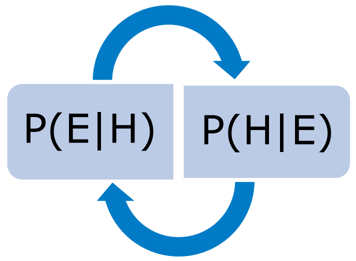
Het Nederlands Forensisch Instituut (NFI) kent een groot aantal typen onderzoeken. Normaal gesproken gaat elk onderzoeksrapport van het NFI vergezeld van een vakbijlage. Deze dient als toelichting op het onderzoek en heeft een zuiver informatief karakter. Achterin de vakbijlage zijn een verklarende woordenlijst en een overzicht van bron- en literatuurverwijzingen opgenomen.

2. Inleiding

In veel gevallen kan de forensisch onderzoeker de vraag die de opdrachtgever hem stelt niet met een volmondig ja of nee beantwoorden. Er is dan een bepaalde mate van onzekerheid over de conclusie. Bij voorkeur wordt deze onzekerheid getalsmatig uitgedrukt, bijvoorbeeld in de vorm van een kans of een interval. Maar in sommige onderzoeken kan de onderzoeker zijn conclusie slechts formuleren in verbale termen van waarschijnlijkheid. Hierbij gebruiken NFI-onderzoekers, waar van toepassing, een standaardreeks van termen om hun conclusiete formuleren. Deze standaardreeks is gebaseerd op inzichten die volgen uit het zogeheten 'Bayesiaanse model' voor de interpretatie van bewijs. In deze vakbijlage wordt dit model besproken aan de

What goes wrong? /prosecutor's fallacy

1. We see this very often
2. With any form of conclusion:
   - Random match probabilities
   - Numerical LRs
   - Verbal LRs
   - 'Ticks on a line'[1]
3. Many variants, such as 'base rate fallacy'
4. Possibly serious consequences with small prior



1: de Keijser, Elffers, Kok & Sjerps, 2009, "Bijkans begrepen", [Bijkans begrepen? | Publicatie | Forensischinstituut.nl]

What goes wrong? /'comparison fallacy'

Irrelevant comparison for interpreting an LR:

- what is normal in a field,
- position in table,
- other cases

Verbal equivalent	LR
Approximately equally probable	1-2
Slightly more probable	2-10
More probable	10-100
Appreciably more probable	100-10,000
Far more probable	10,000-1 million
Extremely probable	> 1 million

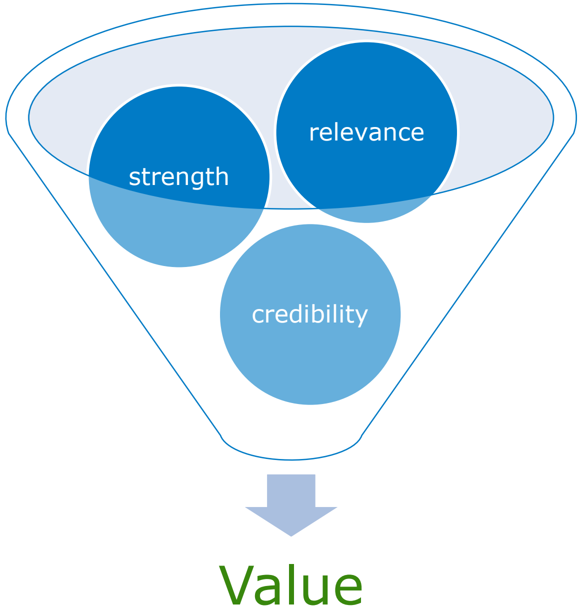
E.g. , for LR=200

- "LR is usually smaller in this area, so very strong evidence"
- "LR would be tossed away in DNA, so very weak evidence"
- "4th position in table of 6, so weak…/…strong evidence"
- "I have seen other reports (in other cases) with higher / lower LRs, so weak…/…strong evidence"

What goes wrong? /'Relevance fallacy'

Ignoring the relevance and focus only on LR

- Small LR at activity level can be more valuable than large LR at source level



Schum 2009 A science of evidence: contributions from law and probability doi:10.1093/lpr/mgp002

What goes wrong? /'Rule based reasoning'

Rule based reasoning, e.g. :

- "DNA on moveable object = trash can"
- "LR smaller than 1 billion = trash can"
- "LR smaller than 1 million = trash can"
- "verbal LR in lower category of verbal scale = trash can"
- "Database search partial match = trash can"
- "Expert knew context = trash can"

What goes wrong? /combining LR

- Large LR with small prior (we put a warning in the report)
- Small LR with large prior
- Ignoring prior
- ‘Salami strategy’: discarding every piece of ‘uncertain’ evidence until there is nothing left

Netherlands Forensic Institute
Ministry of Justice and Security

What works well?

- Explaining LR framework in different ways (numbers, pictures, words, formulas)
- “Everyday” comparisons (windforce, matching facial features)
- Relating posterior probability to prior (table, graphs, various situations)
- Avoiding probabilistic terms:
  - “results fit better with A than B”
  - “Conclusion is like adding pebble stone in prosecutor’s scale”

- Being transparent about errors and uncertainties (expected from scientist)

Main message from the three presentations

1. Logic of LRs is leading
2. Invest in a training and continuing education program for experts
3. Numbers are unequivocal, words are not:
   - 1 mile = far (snail)
   - 1 mile = close (athlete)
4. Communicating LRs is difficult, but not impossible:
   - Train and test experts
   - Provide communication tools: leaflets, visuals, suitable wordings, e-learnings…
   - More research on effective communication is very welcome
5. Statisticians IN THE LAB

WHY IT MATTERS

MIND THE GAP

THE STAGE: TEXAS 2015/2016

- Issues with application of CPI/CPE statistic to DNA mixture evidence in DC.
- Same issues existed in Texas, to a greater or lesser degree depending on the laboratory.
- State embarked on a massive review of DNA mixture cases where CPI/CPE used.
- During this process, prob gen was around the corner.
- And we sped that train along....with training but without an ENFSI roadmap.

HOWARD WAYNE LEWIS: CAPITAL MURDER

- In 2018, Lewis was convicted of murdering his 18-month old son. The baby was hanged over the bathroom door. His grandmother (and caretaker) was beaten to death.
- Lewis was estranged from mom and lived in Dallas; the murder occurred in Huntsville (2.5 hour drive).
- The jury sentenced Lewis to death.
- When the DNA testing was done in this case, Texas DPS had just finished validating and launching STRmix.

HOWARD WAYNE LEWIS: CAPITAL MURDER

- In 2018, Lewis was convicted of murdering his 18-month-old son. He was sentenced to death.
- When the DNA testing was done in this case, Texas DPS had just finished validating and launching STRmix.
- They had—at the time—a likelihood ratio range they deemed to be “inconclusive.” For minifiler, the DNA analyst **thought** the range was .01 to 1000.


Global
HD

HOWARD WAYNE LEWIS: CAPITAL MURDER

At the time the following was true:

- They were **used to** an inconclusive range, and having it was perceived as a “conservative” approach;
- They didn’t have confidence in their ability to explain that the lower LR’s may include adventitious matches from their validation;
- Dad (ESR) said it was okay.

Testimony re: Mini-Filer Analysis of White Rope

Q. One more thing. Can you tell me what the likelihood ratio for Howard Lewis was?

MR. BLAZEK: Your Honor, I made that objection early with regard to the other witness.

THE COURT: Right.

MR. BLAZEK: I would make that same objection to her and ask it be considered at this time.

THE COURT: Yes. The objection is overruled. You may answer.

THE WITNESS: May I refer to my notes?

THE COURT: If you can, yes.

A. 995.

Q. (By Ms. Stroud) And when you get to a thousand you're included?

A. It would be over a thousand, so a thousand and one

Mini-Filer Analysis of White Rope

would be an inclusionary rage.

Q. And he was?

A. 995.

MS. STROUD: I pass the witness, Your Honor.

MR. BLAZEK: May I have a running objection?

THE COURT: Yes, you may.

And you know, I guess maybe I was in a different courtroom or maybe it's because I'm just a lawyer and I didn't take science classes and I'm not a DNA analyst. But do you remember this? And you remember those girls with all that education, do you remember what they said? This is a number scale. And the way we decide where you go on this scale is by a number. And do you remember when I asked Clare, so you're telling me that when you use words like excluded and inconclusive, you're not speaking the same language we are? And she said no, ma'am. Those are our words. They don't mean what you think they mean.

Mr. Blazek wants to stand up here and tell you that that DNA exonerated Howard Lewis. And guess what? I can give you 995 reasons why he's not exonerated. 995. It takes 1,000 to be included. 995.

CALL FROM FRANK BLAZEK (DEFENSE)

- "I think the DNA analyst misstated the upper end of the inconclusive range." It should have been 10,000 for minifiler.

THINGS ARE GETTING BETTER

8 YEARS IN….

- Not all labs have transitioned to STRmix.

- DPS ended up having to review all its “inconclusive” cases because a handful of them actually were exclusionary LR’s (where NOC was overstated). AND THEY DID IT.

- Not all labs have transitioned to prob gen but most are in the process.

- Communication is better but **still hard**.

ON THE HORIZON: WE MUST ADDRESS

- Evaluative Reporting in All Other Disciplines: How? When?

- Activity Level Reporting
  - Right now, we answer "is it possible" questions all day;
  - We can transition to talking about the probability of the evidence given two mutually exclusive activity propositions. HOWEVER:
- There is no UNIL here—not even close.
- Risk of ad hoc pseudo evaluations on the stand is HUGE
- No idea how to properly analyze the data/build a Bayes net;
- Traceability!
- Transparency!