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

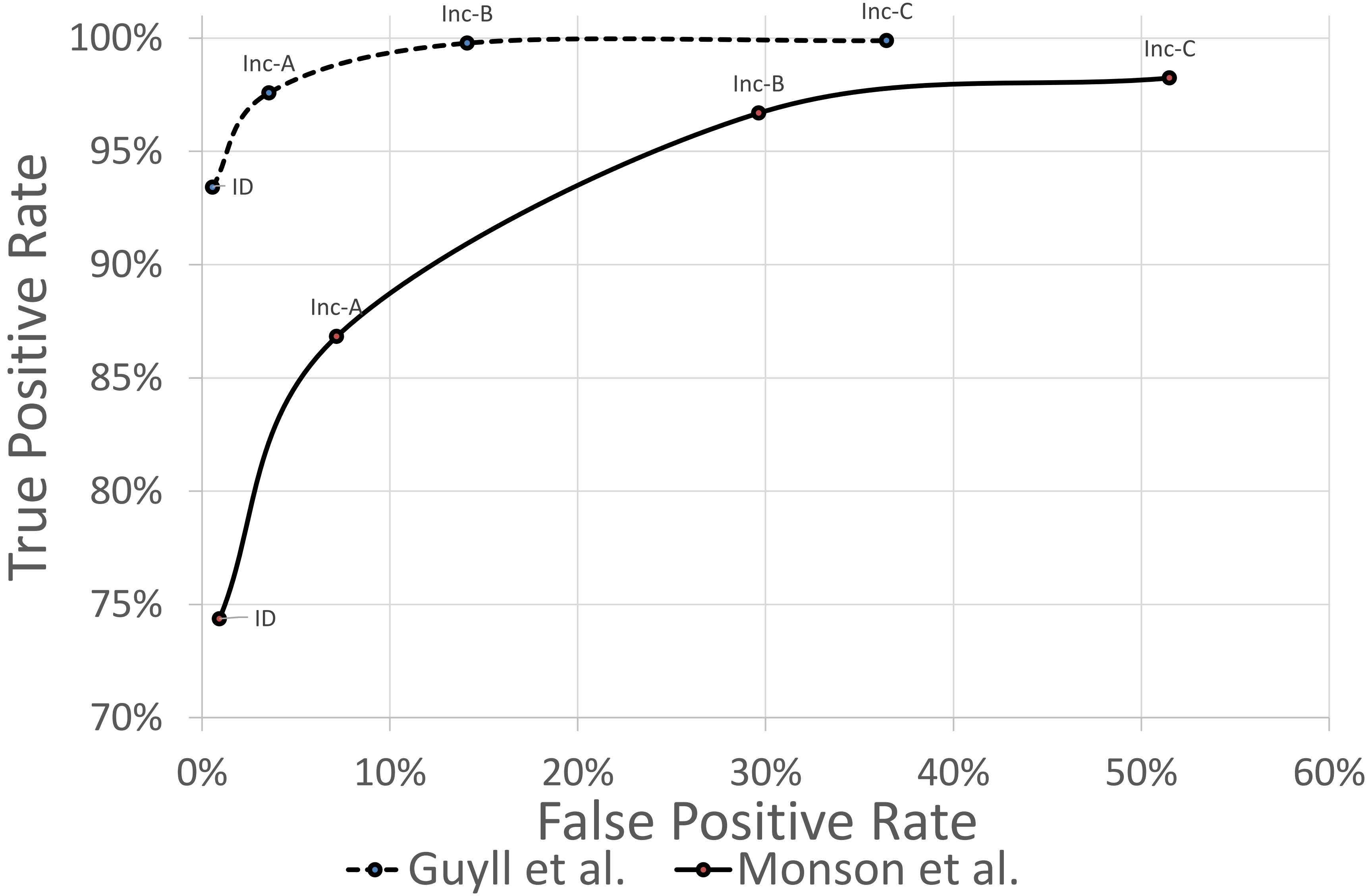
Example for Discussion

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?



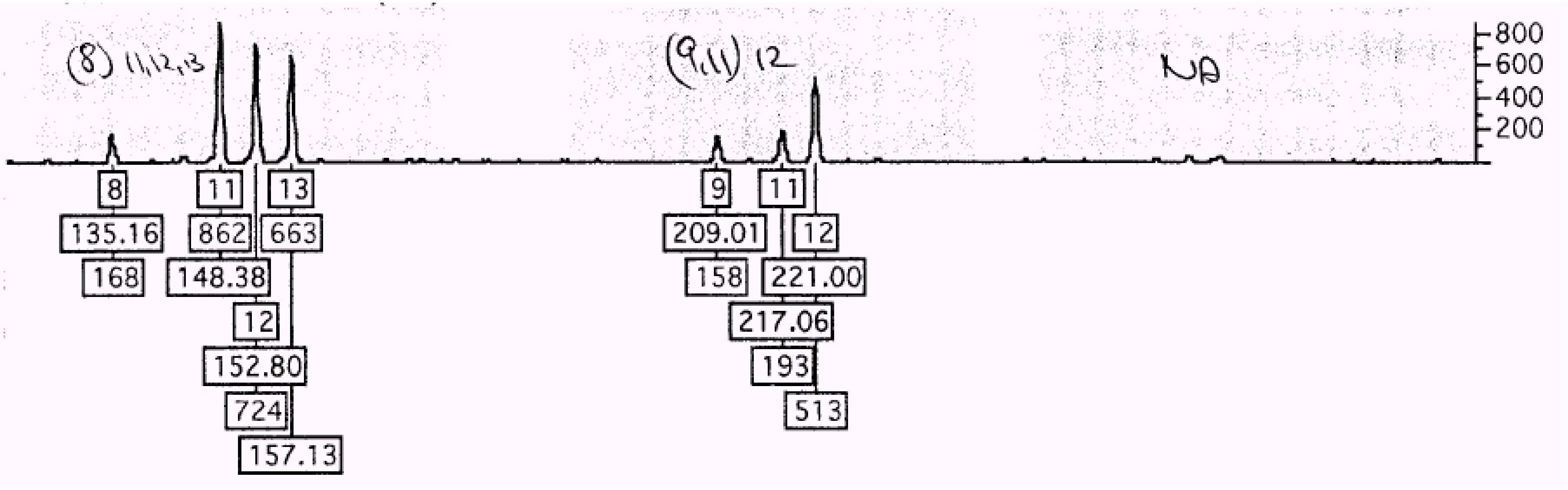
ROC Curves Cartridge Comparison based on Observed Error Rates in Monson et al. (2023) and Guyll et al. (2023) at Four Decision Thresholds

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

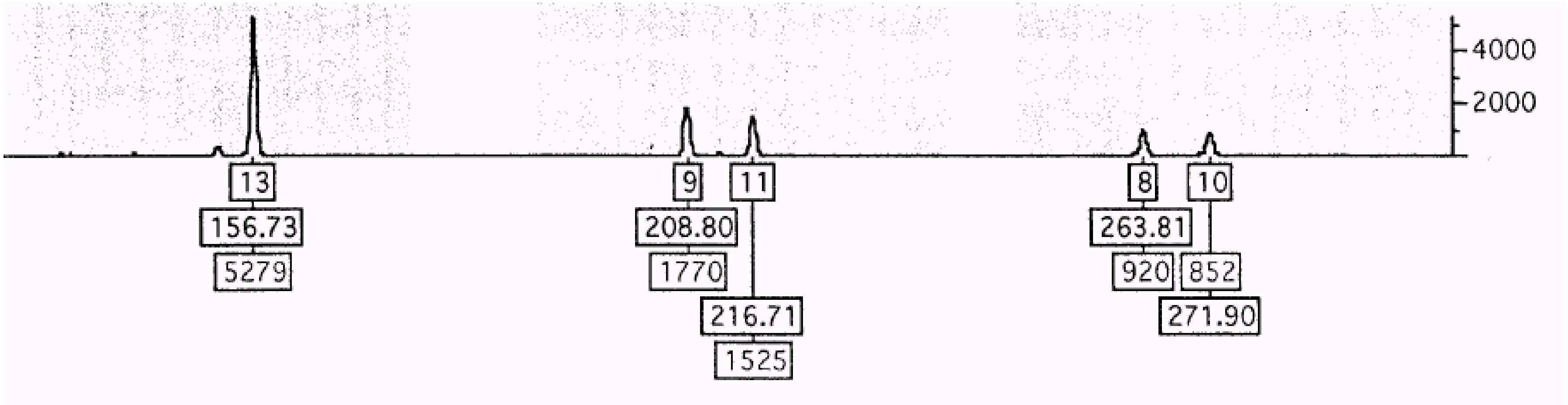
Additional caveats

- 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

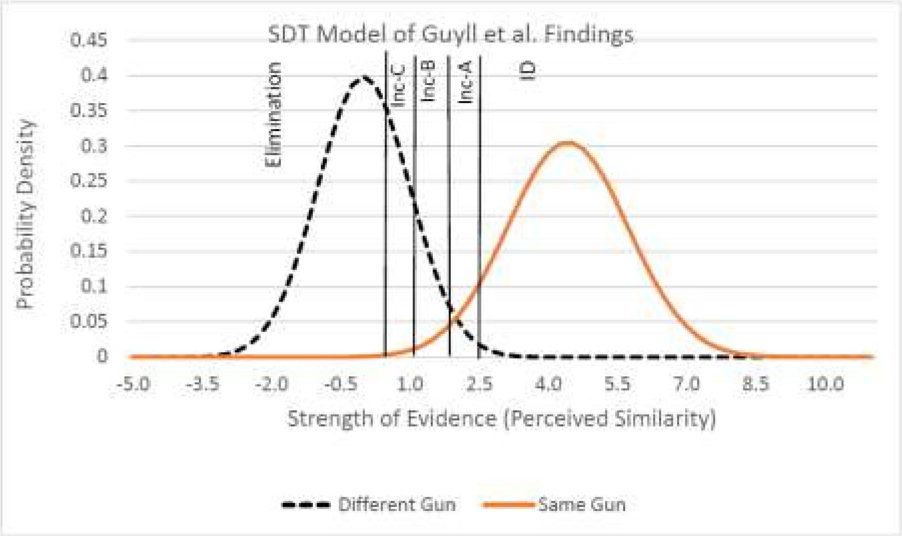


Profile of Defendant



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.

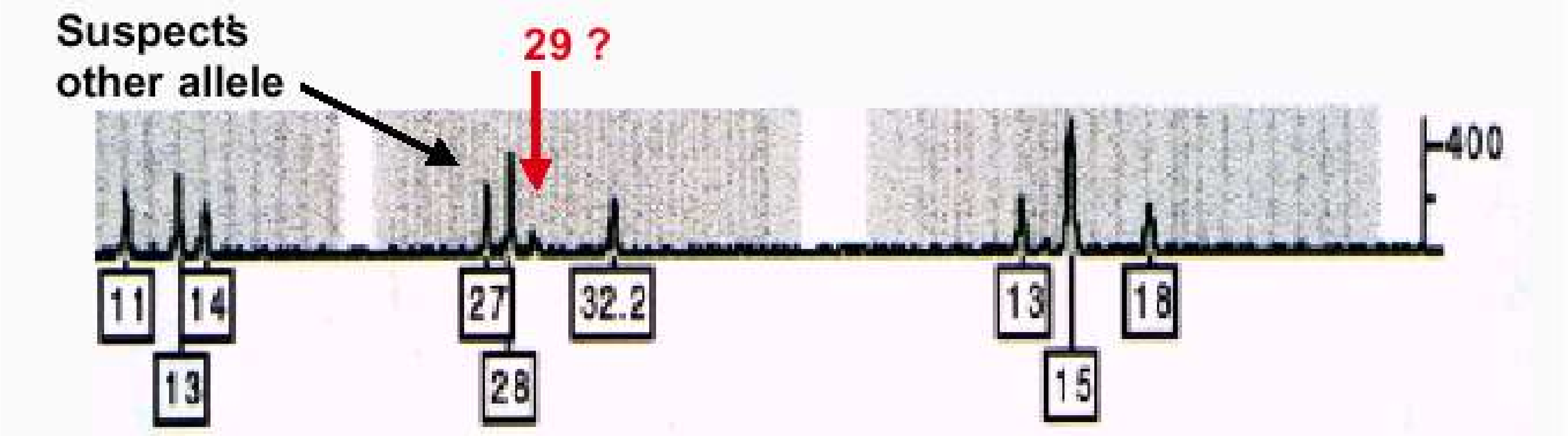
Thompson, W.C., Vuille, J., Taroni, F., & Biedermann, A. (2018). After Uniqueness: The Evolution of Forensic Science Opinion. *Judicature,* 102(1): 18-27.

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.

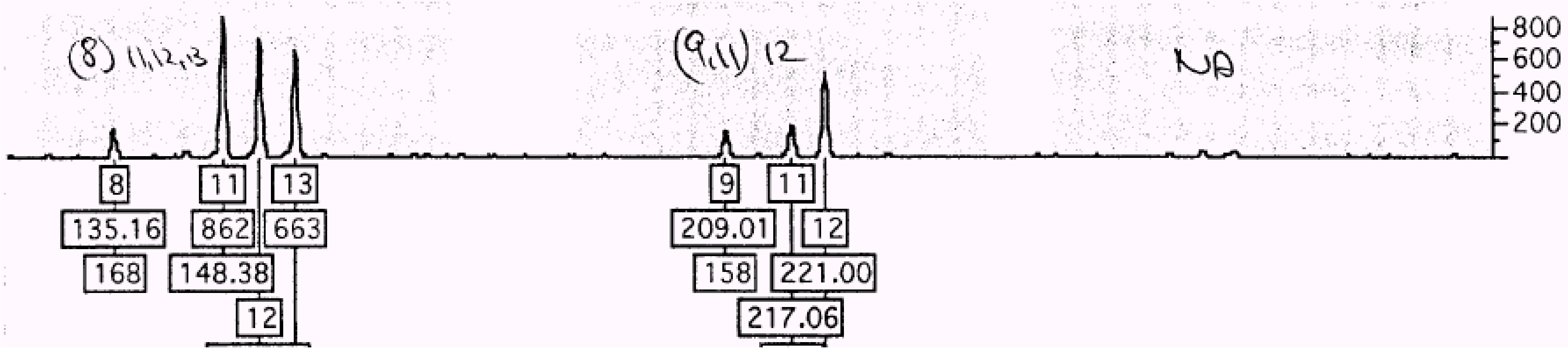
NIST Mixed Stain Study #2 (2 contributors)

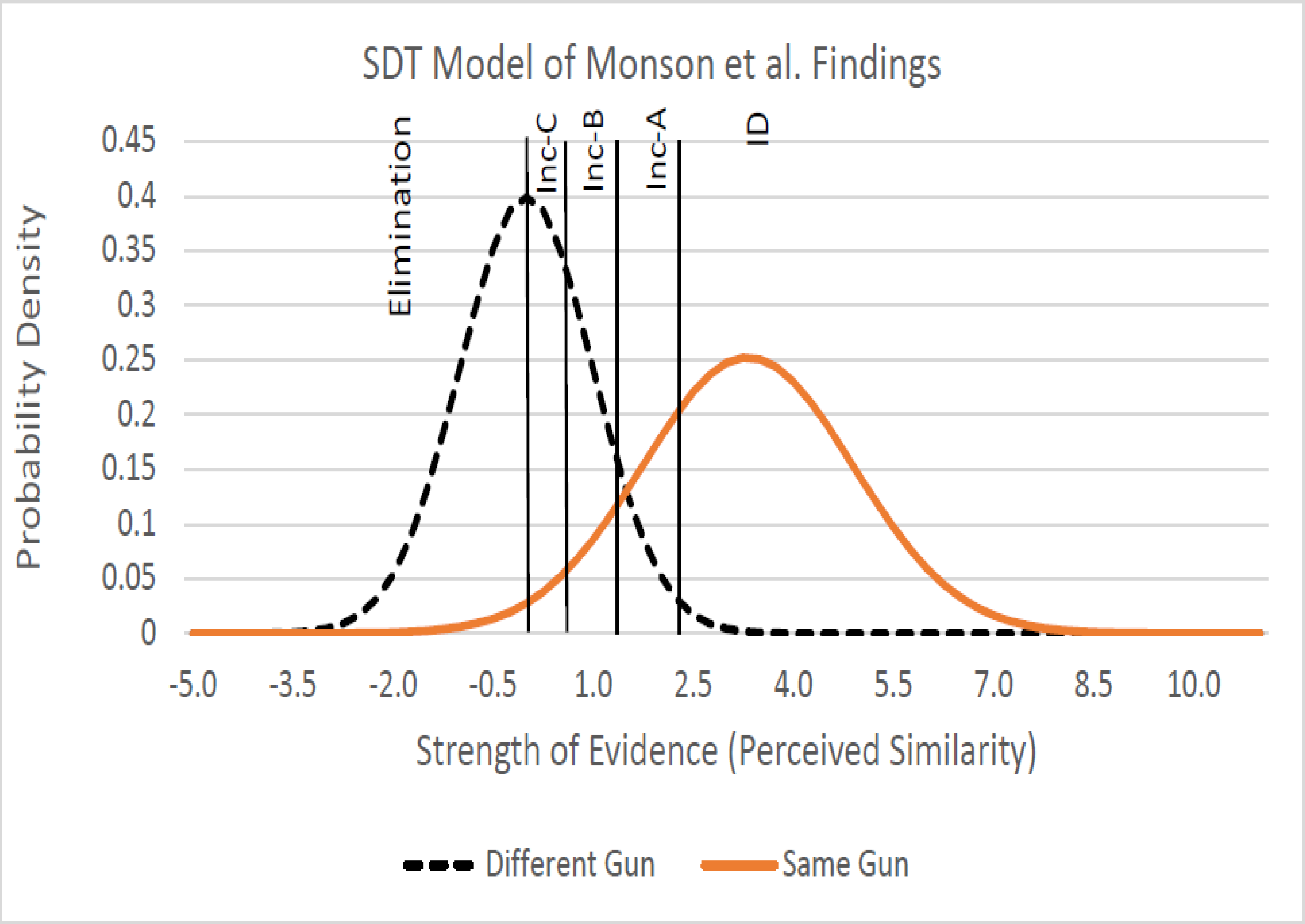
Fig. 4 Non-Amplifying Allele D21S11 locus

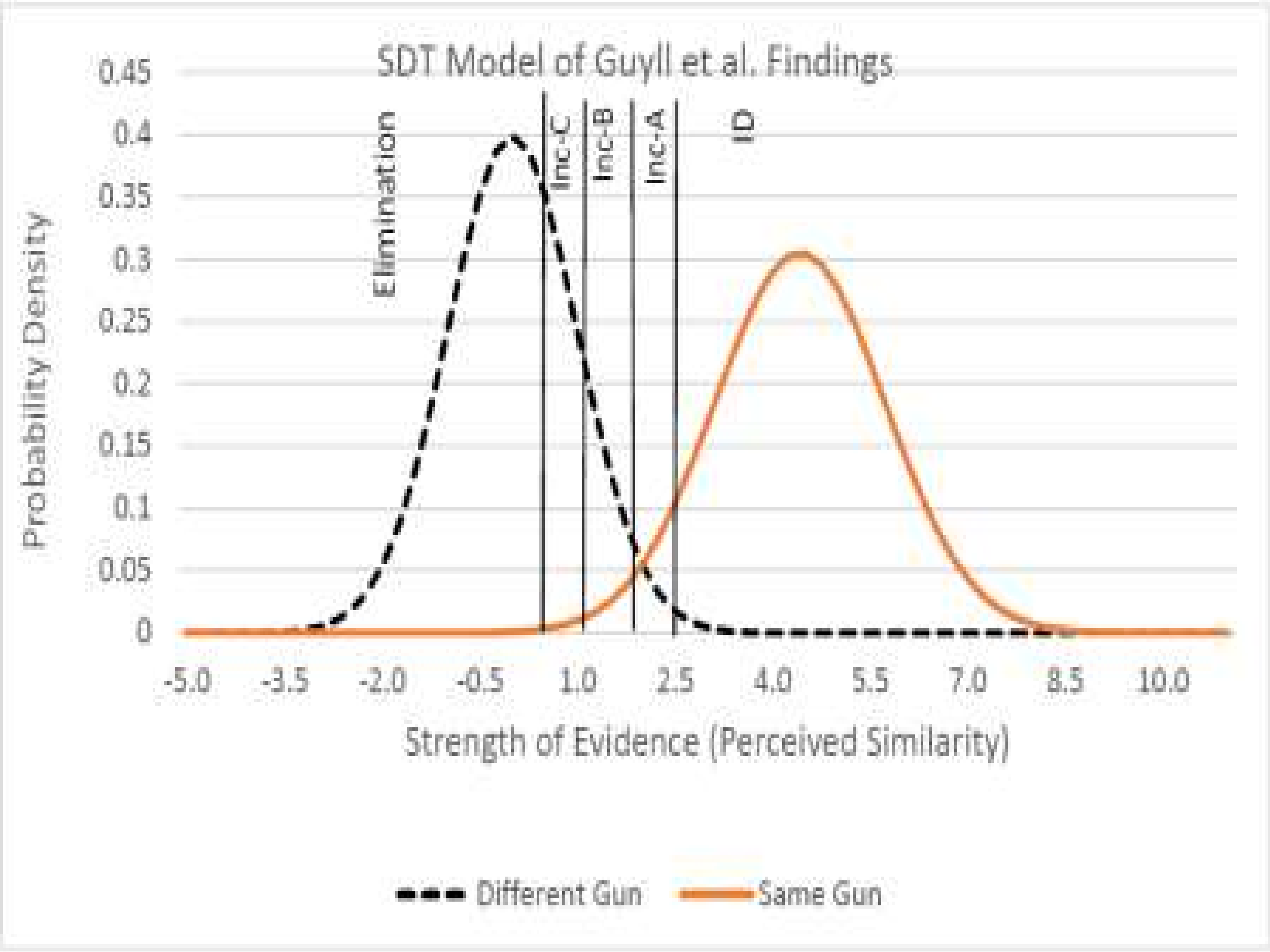

Suspects other allele

29 ?

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