EXPLAINABILITY IN THE COURTS: ALGORITHM-GENERATED *PROOF*

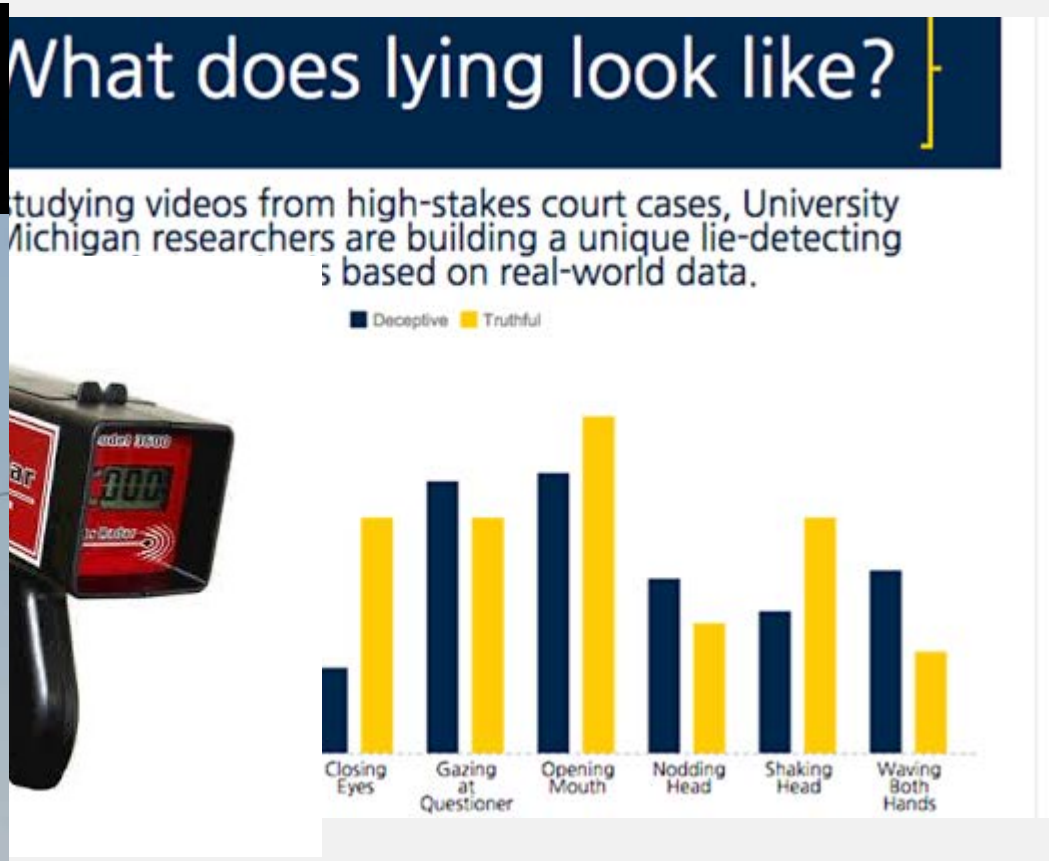
Andrea Roth

Assistant Professor, UC Berkeley School of Law

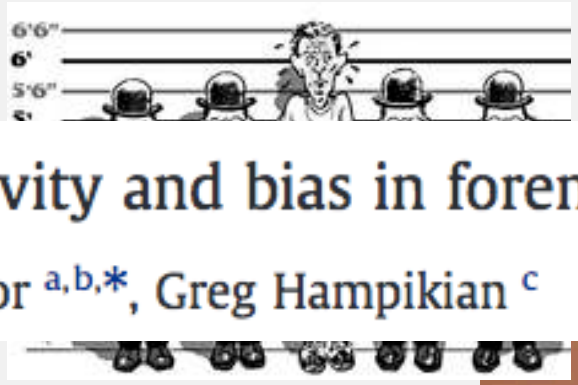
NYU Algorithms and Explanations Conference

April 28, 2017

WHAT IS ALGORITHM-GENERATED PROOF?



WHAT IS IT REPLACING?



Subjectivity and bias in forensic DNA mixture interpretation ☆

Itiel E. Dror [a,b,*], Greg Hampikian [c]

SO WHAT'S THE PROBLEM?

- Will jurors draw an appropriate inference from machine-generated proof?
- Are there non-accuracy-related individual interests of the litigants (e.g. dignity) implicated by machine-generated proof?
- Will the public view verdicts based on such proof as legitimate?

SO WHAT'S THE PROBLEM?

- Will jurors draw an appropriate inference from machine-generated proof?
- Are there non-accuracy-related individual interests of the litigants (e.g. dignity) implicated by machine-generated proof?
- Will the public view verdicts based on such proof as legitimate?

PEOPLE V. HILLARY

TYPICAL PGS REPORT

CYBERGENETICS REPORT:

TrueAllele assumed that the evidence sample data (Item 1SF) contained two or three contributors, and objectively inferred evidence genotypes solely from these data. The victim's genotype was assumed in some calculations, and degraded DNA was considered. Following genotype inference, the computer then compared a genotype from this evidence item to provided reference (Items 1EF, 10, 11 and 12) genotypes, relative to reference populations, to compute LR DNA match statistics. Based on these results:

A match between the vaginal slide sperm fraction (Item 1SF) and **[Boyfriend]** (Item 11) is:

2.82 million times more probable than a coincidental match to an unrelated Black person,
11 million times more probable than a coincidental match to an unrelated Caucasian person, and
73.3 million times more probable than a coincidental match to an unrelated Hispanic person.

A match between the vaginal slide sperm fraction (Item 1SF) and **[Defendant]** (Item 12) is:

1.62 quintillion times more probable than a coincidental match to an unrelated Black person,
43.6 quintillion times more probable than a coincidental match to an unrelated Caucasian person, and
330 quintillion times more probable than a coincidental match to an unrelated Hispanic person.

April 1, 2015

Jerry D. Varnell, Contract Specialist
Procurement Section, Department of Justice
Federal Bureau of Investigation
935 Pennsylvania Avenue, N.W.
Washington, District of Columbia 20535

Cybergenetics TrueAllele makes full use of the DNA typing data, as described in many papers and presentations[2]. However, New Zealand's product does not fully consider all the DNA data and possible solutions.

TrueAllele resolves DNA mixtures without any knowledge of a comparison reference genotype, thus is entirely objective. STRmix uses the comparison reference in its calculations, and is therefore a subjective approach.

TrueAllele has been demonstrated to give consistent match statistics in dozens of validation studies and on hundreds of complex DNA mixtures. STRmix can give different answers based on how an analyst sets their input parameters.

IS EXPLAINABILITY NECESSARY TO AVOID INFERENTIAL ERROR BY JURORS FROM ALGORITHM-GENERATED PROOF?

EXPLAINABILITY AND HUMAN EXPERTS

Legal question: How do you cross-examine a computer?

August 29, 2016 12:25 AM




HOW EXISTING LAW REGULATES ALGORITHM-GENERATED PROOF

- Machine sources of information are generally (and correctly) not considered "declarants" under the hearsay rule
- Not considered "witnesses" for Confrontation Clause purposes
- Only regulated by:
  - Daubert/Frye, if it's the basis of a human expert's opinion
  - Authentication rules (eg FRE 901, showing that a "process or system . . . produces an accurate result")

LIMITS OF VALIDATION AND ERROR RATES

- Idealized conditions versus marginal cases where machines offer the most added value
- Error rate alone doesn't make clear the relative plausibility of various false positive scenarios
- No easily determined baseline for determining how erroneous certain assertions, e.g. match statistics, are
- Even if method is sound, does current version of software operationalize it correctly?

LIMITS OF TRANSPARENCY ALONE

- Inscrutability; TrueAllele has 170,000 lines of code
- Resource asymmetry among litigants
- Intellectual property concerns with disclosing source code
- BUT… source code can be valuable

Jul 20, 2015

The Forensic Commission and the DNA Subcommittee

New York State

Commissioner Michael Green (Forensic Science Commission Chairman)

Dr. Dwight Adams (DNA Subcommittee Chairman)

This statement is incorrect. There was a minor miscode in an early version of STRmix. This took a particular set of circumstances to "fire" and so occurred very rarely. The effect was a minor change in the LR which could go either way and was typically smaller than a factor of 10 in a very large number. There was no false positive in the case (confirmed with Queensland), nor in any case of which we are aware in Australia, New Zealand or the US.

I have written to Mr Scheck three times now asking for clarification of his statement or a retraction. I have not received a reply. Because of this I write to you directly to set the record straight.

Yours faithfully

John Buckleton, DSc, Ph.D. FRSNZ.

SO WHAT'S THE PROBLEM?

- Will jurors draw an appropriate inference from a machine-generated proof?
- Are there non-accuracy-related individual interests of the litigants (e.g. dignity) implicated by machine-generated proof?
- Will the public view verdicts based on such proof as legitimate?

INSCRUTABILITY, ACCOUNTABILITY, AND DIGNITY

- Sixth Amendment right to *know* one's accuser
- Personhood argument against certain mechanical interpretations of human emotions
- "Push button justice"

SO WHAT'S THE PROBLEM?

- Will jurors draw an appropriate inference from a machine-generated proof?
- Are there non-accuracy-related individual interests of the litigants (e.g. dignity) implicated by machine-generated proof?
- Will the public view verdicts based on such proof as legitimate?

LEGITIMACY

- Whether explainability in algorithm-generated proof promotes legitimacy is likely context-specific
  - Field Sobriety Tests?
  - George Fisher, *The Jury's Rise as Lie Detector* (1997)
  - *Trial by Mathematics* and greater deference to verdicts based on an "actual belief" in guilt?

WAYS TO ENHANCE EXPLAINABILITY OF ALGORITHM-GENERATED PROOF

- Allow pretrial access for modifying parameters and "tinkering"
- Require pretrial disclosure of other relevant machine statements (e.g. other runs under different parameters)
- Require detailed account of all analytical assumptions on "hot button" issues underlying machine's conclusion
- Require, if needed, live testimony of designers or inputters
- Jury education (e.g. on likelihood ratios)

QUESTIONS?

BerkeleyLaw
UNIVERSITY OF CALIFORNIA

Boalt Hall