Overcoming Bias in DNA Mixture Interpretation

American Academy of Forensic Sciences
February, 2016
Las Vegas, NV

Mark W Perlin, PhD, MD, PhD
Cybergenetics, Pittsburgh, PA

Cybergenetics

Cybergenetics © 2003-2016

DNA

Does
Not
Advocate

Gold standard of forensic evidence

However, ... there may be problems ... with how the DNA was ... interpreted, such as when there are mixed samples

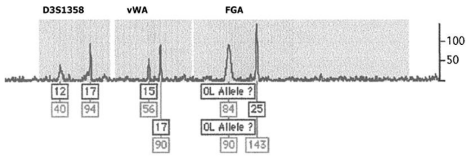
Law, Probability and Risk (2009) **8**, 257–276 doi:10.1093/lpr/mgp013
Advance Access publication on July 28, 2009

Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation†

WILLIAM C. THOMPSON*
Department of Criminology, Law and Society, University of California, Irvine, CA 92697, USA

[Received on 17 November 2008; revised on 19 April 2009; accepted on 24 April 2009]

Forensic DNA analysts tend to underestimate the frequency of matching profiles (and overestimate likelihood ratios) by shifting the purported criteria for a 'match' or 'inclusion' after the profile of a suspect becomes known—a process analogous to the well-known Texas sharpshooter fallacy. Using examples from casework, informal and naturalistic experiments, and analysts' own testimony, this article demonstrates how *post hoc* target shifting occurs and how it can distort the frequency and likelihood ratio statistics used to characterize DNA matches, making matches appear more probative than they actually are. It concludes by calling for broader adoption of more rigorous analytical procedures, such as sequential unmasking, that can reduce the sharpshooter fallacy by fixing the target before the shots are taken.



Defendant	D3S1358	vWA	FGA
Tom	17,17	15,17	25,25
Dick	12,17	15,17	20,25
Harry	14,17	15,17	20,25
Sally	12,17	15,15	20.22

FIG. 2. Electropherogram of a saliva sample and four suspect profiles.

Science and Justice 51 (2011) 204–208

Contents lists available at SciVerse ScienceDirect

Science and Justice

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

Subjectivity and bias in forensic DNA mixture interpretation ☆

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

[a] Institute of Cognitive Neuroscience, University College London (UCL), London, UK
[b] Cognitive Consultants International (CCI), London, UK
[c] Departments of Biology and Criminal Justice, Boise State University, USA

ARTICLE INFO

Article history:
Received 30 July 2010
Received in revised form 30 July 2011
Accepted 4 August 2011

Keywords:
Human cognition
Bias
Forensic decision making
Contextual influences
DNA interpretation

ABSTRACT

The objectivity of forensic science decision making has received increased attention and scrutiny. However, there are only a few published studies experimentally addressing the potential for contextual bias. Because of the esteem of DNA evidence, it is important to study and assess the impact of subjectivity and bias on DNA mixture interpretation. The study reported here presents empirical data suggesting that DNA mixture interpretation is subjective. When 17 North American expert DNA examiners were asked for their interpretation of data from an adjudicated criminal case in that jurisdiction, they produced inconsistent interpretations. Furthermore, the majority of 'context free' experts disagreed with the laboratory's pre-trial conclusions, suggesting that the extraneous context of the criminal case may have influenced the interpretation of the DNA evidence, thereby showing a biasing effect of contextual information in DNA mixture interpretation.

© 2011 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.

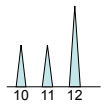
Case context impact

	With context	Without context
Include	2	1
Exclude		12
Inconclusive		4

DNA mixture

Genotype 1 | Genotype 2 | Data

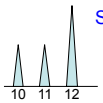
10, 12 ⊕ 11, 12 ⇒



(oversimplified cartoon diagram)

Interpret #1: separate

Data | Genotype 1 | Genotype 2



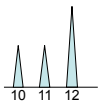
Separate ⇒

Genotype 1	Genotype 2
10, 10 @ 10%	10, 10 @ 10%
10, 11 @ 20%	10, 11 @ 10%
10, 12 @ 40%	10, 12 @ 10%
11, 11 @ 10%	11, 11 @ 10%
11, 12 @ 10%	11, 12 @ 40%
12, 12 @ 10%	12, 12 @ 20%

Unmix the mixture

Interpret #2: compare

Data | Genotype 2



10, 10 @ 10%
10, 11 @ 10%
10, 12 @ 10%
11, 11 @ 10%
11, 12 @ 40%
12, 12 @ 20%

Compare ⇕ with 11,12

Match statistic = $\frac{\text{Prob\{match\}}}{\text{Prob\{coincidence\}}} = \frac{40\%}{4\%} = 10$

Cognitive bias

Illogical thinking affects decisions

- Anchoring – rely on first information
- Apophenia – perceive meaningful patterns
- Attribution bias – find causal explanations
- Confirmation bias – interpretation confirms belief
- Framing – social construction of reality
- Halo effect – sentiments affect evaluation
- Oversimplification – simplicity trumps accuracy
- Self-serving bias – distort to maintain self-esteem

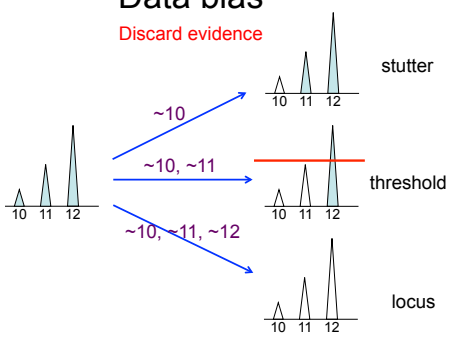
Contextual bias

Background information affects decisions

- Academic bias – beliefs shape research
- Educational bias – whitewash damaging evidence
- Experimenter bias – expectations affect outcomes
- Inductive bias – tilt toward training examples
- Media bias – selecting mass media stories
- Motivational bias – reaching desired outcome
- Reporting bias – under-report undesirable results
- Social desirability bias – want to be seen positively

Data bias

Discard evidence



Genotype bias

Actual		Desired
10, 10 @ 5%		10, 10 @ 0%
10, 11 @ 5%		10, 11 @ 0%
10, 12 @ 75%	⇨	10, 12 @ 100%
11, 11 @ 5%		11, 11 @ 0%
11, 12 @ 5%		11, 12 @ 0%
12, 12 @ 5%		12, 12 @ 0%

RMP – random match probability
- analyst chooses only one genotype
- inflates DNA match statistic

Perlin, M.W. "Inclusion probability for DNA mixtures is a subjective one-sided match statistic unrelated to identification information." *Journal of Pathology Informatics*, 6(1):59, 2015.

Match bias

CPI – combined probability of inclusion
- analyst begins by including the suspect
- unrealistic, unproven model
- random number generator
- lacks probative value

LR – likelihood ratio
- analyst ignores much of the data
- calculation requires suspect genotype
- introduces "phantom" peaks (drop out)
- considers few genotype possibilities

Process bias

(1) Choose, alter, discard, edit, and manipulate the DNA **data signals**

(2) Compare defendant's **genotype** to edited **data** & decide if he is in the DNA evidence

(3) If he is "included", then **calculate** a DNA mixture statistic

Hidden cognitive and contextual bias largely determine the outcome

Presented as unbiased science

Software bias

Why labs choose mixture software

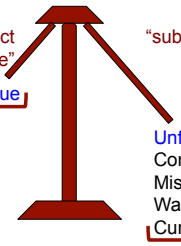
• Puts analyst in charge	Confirmation bias
• Results confirm belief	Confirmation bias
• Simplifies the problem	Oversimplification
• Gets desired answer	Motivational bias
• The FBI uses it	Social desirability bias
• Familiar process	Social desirability bias

Relevance (FRE 403)

Admissibility of biased DNA evidence

Rule 401
"evidence makes a fact more or less probable"

Probative value inflated



Rule 403
"substantially outweighed by a danger of:"

Unfair prejudice
Confusing the issues
Misleading the jury
Wasting time
Cumulative evidence
"DNA"

Cross examination

Hundreds of effective questions can elicit bias

"Did you know the defendant's genotype during your analysis of the evidence?"

"Doesn't knowing your customer's desired answer bias your decisions?"

"Have any scientific studies demonstrated otherwise?"

Sequential unmasking

Human DNA review proposal (reduce bias):

1. **First** analyze the crime scene data, without knowing context or references
2. **Then** compare with reference samples

But there is potential **bias** in choosing data, conducting analysis, and making comparisons.

Human analysts can always introduce bias.
Why is a human even involved in this process?
Why not use an unbiased computer instead?

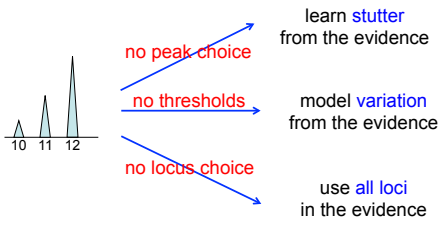
Unbiased interpretation

Use an objective **computer** to:

1. **Examine** *all* DNA data, without having suspect's genotype
2. **Separate** genotypes of each DNA mixture contributor, considering *all* possible solutions
3. **Compare** genotypes only *afterwards* to calculate match statistics

Eliminate all human involvement
to overcome cognitive & contextual bias
in DNA mixture interpretation

No data bias – use all evidence



No genotype bias – objective

Actual		Desired
10, 10 @ 5% 10, 11 @ 5% 10, 12 @ 75% 11, 11 @ 5% 11, 12 @ 5% 12, 12 @ 5%	⟺	Use the actual genotype probability

Do not change probability

No match bias – accurate

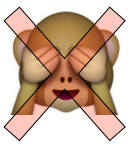
CPI – combined probability of inclusion
- random number generator
- bad forensic science
- review all past cases

LR – likelihood ratio
- don't ignore any data
- don't use suspect genotype
- don't concoct "phantom" peaks
- use all genotype possibilities

No process bias – remove analyst

(1) Do not change **data signals**

(2) Do not use **defendant genotype**



(3) **Calculate** accurate DNA match statistic

Eliminate cognitive and contextual bias from the process

Present unbiased science

No software bias – true stats

Accurate, objective, thorough, validated

- Puts analyst in charge
- Results confirm belief
- Simplifies the problem
- Gets desired answer
- The FBI uses it
- Familiar process

Examine all the data
without human choice
↓
Separate genotypes
consider all solutions
↓
Compare genotypes
stats decide outcome

TrueAllele® information

http://www.cybgen.com/information

- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

http://www.youtube.com/user/TrueAllele
TrueAllele YouTube channel

Cybergenetics

YouTube

perlin@cybgen.com