Forensic Genetics

Module 15 – Session 6

Schedule – Day 2

Session 5	Population Structure and Relatedness	8:00-8:50
Review	Session 5	9:05-9:55
Session 6	Reporting and Presenting LRs	10:10-11:00
Review	Session 6	11:30-12:20
Session 7	Y-STR Profiles	12:35-1:25
Session 8	Incorporating Relatives	1:40-2:30

Session 6 - Reporting and Presenting LRs

- Formulating Propositions
  - Hierarchy of propositions
  - Guidelines
  - Issues

- Communicating LRs
  - Verbal scales
  - Bias

- Fallacies
  - Prosecutor’s fallacy
  - Defendant’s fallacy
  - Uniqueness fallacy
  - Association fallacy

Likelihood Ratio

The LR assigns a numerical value in favor or against one proposition over another:

$$\text{LR} = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)},$$

where H_p typically aligns with the prosecution case, H_d is a reasonable alternative consistent with the defense case, and I is the relevant background information.

Setting Propositions

- The value for the LR will depend on the propositions chosen: different sets of propositions will lead to different LRs.

- Choosing the appropriate pair of propositions can therefore be just as important as the DNA analysis itself.

Hierarchy of Propositions

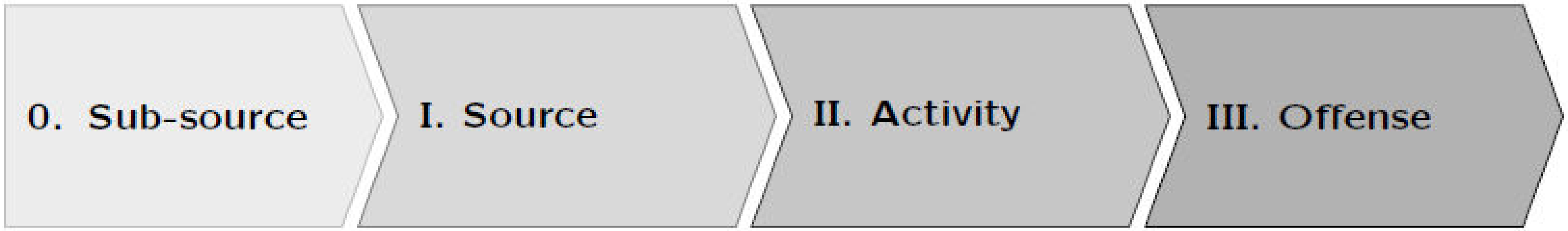
Evett & Cook (1998) established the following hierarchy of propositions:

Level	Scale	Example
III	Offense	H_p: The suspect raped the complainant. H_d: Some other person raped the complainant.
II	Activity	H_p: The suspect had intercourse with the complainant. H_d: Some other person had intercourse with the complainant.
I	Source	H_p: The semen came from the suspect. H_d: The semen came from an unknown person.
0	Sub-source	H_p: The DNA in the sample came from the suspect. H_d: The DNA in the sample came from an unknown person.

Hierarchy of Propositions

- The *offense* level deals with the ultimate issue of guilt/ innocence, which are outside the domain of the forensic scientist.
- The *activity* level associates a DNA profile or evidence source with the crime itself, and there may be occasions where a scientist can address this level.
- The *source* level associates a DNA profile or evidence item with a particular body fluid or individual source.
- The *sub-source* level refers to the strength of the evidence itself. This is usually the level a DNA reporting analyst will spend most of their time.

Hierarchy of Propositions



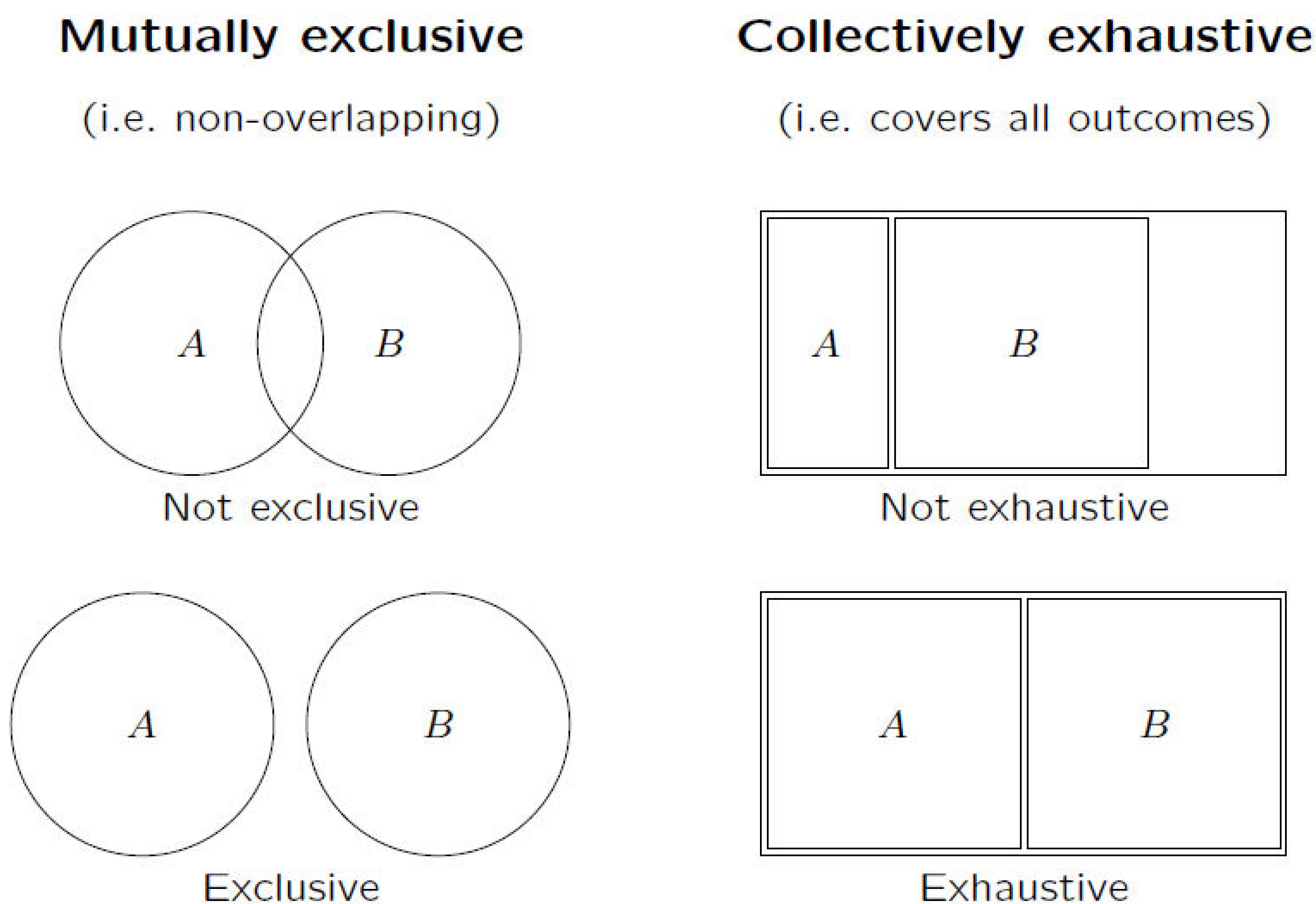
- A forensic scientist can provide information in relation to propositions which are intermediate to the ultimate issue.
- The higher the level of propositions, the more information is needed on the framework of circumstances.
- Since different levels rely on different assumptions to consider, strength of the evidence estimates will change significantly at each level.

Setting Propositions

Some useful principles for setting hypotheses:

- Propositions should address the issue of interest;
- Propositions should be based on relevant case information;
- Propositions should not include irrelevant details;
- Propositions should be (close to) MECE.

MECE Definition



Background Information

- **Relevant background information** can help set appropriate propositions. E.g. the origin of clothing or intimate vs. non-intimate swab can help determine if it is reasonable to assume a known contributor.

- **Irrelevant background information** is not needed and may contribute to bias decision making (e.g. criminal history, confession, presence or lack of other evidence).

Formulating Propositions

- The prosecution hypothesis (H_p) is usually known, or more or less straightforward to set.

- However, the defense are usually under no requirement to offer a proposition, and often they do not.

- If a defense stance is not available, a sensible proposition can be chosen.

Formulating Propositions

What if multiple alternative hypotheses are relevant?

- Report the 'most relevant' LR (and provide the rest in the appendix);
- Provide all considered propositions and corresponding LRs;
- Report only the lowest LR to provide a lower bound for the LR.

Formulating Propositions

What about the number of contributors?

This is an important component of mixture interpretation. Most approaches assume that the NoC is known.

- Underestimating the NoC is usually conservative (minor contributors may be incorrectly excluded).

- Overestimating the NoC may not be conservative (non-contributors may not be excluded).

- For major contributors the NoC has little effect on the LR.

Number of Contributors

What about the number of contributors?

This is an important component of mixture interpretation. Most approaches assume that the NoC is known.

- The MAC method does not always work (e.g. when we have four alleles, but the POI is homozygous).
- Multiple LR_n values may be calculated for varying number of contributors n and the most conservative one is usually presented.
- Machine learning approaches have been proposed to assess the NoC[1].

[1] PACE: Probabilistic Assessment for Contributor Estimation - A machine learning-based assessment of the number of contributors in DNA mixtures (Marciano & Adelman, 2017).

Number of Contributors

Another option is to calculate a weighted average[1]:

$$\text{LR} = \sum_{n=1}^{N} \text{LR}_n \, \text{Pr}(\text{NoC} = n),$$

where prior independence is assumed:

$$\text{Pr}(\text{NoC} = n | H_p) = \text{Pr}(\text{NoC} = n | H_d)$$

The ISFG also recognizes that there may be situations where different number of contributors in H_p and H_d are needed. Non-equivalence of the prior seems a rare event and may be difficult to interpret.

[1] Contributors are a nuisance (parameter) for DNA mixture evidence evaluation (Slooten & Caliebe, 2018).

Formulating Propositions

What about relatives?
The LR can accommodate for this, which we will see in the next section.

What if the DNA got there by some other means?
This indicates a different level of propositions. The discussion will likely move to transfer and contamination.

Propositions are formed based on information available at that time. If this information changes, or the defense want any other propositions considered, it may be necessary to update or add LR calculations.

Reporting LRs

As can be seen from the definition of the likelihood ratio

$$\text{LR} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)},$$

- an LR > 1 supports the prosecution hypothesis, meaning that the evidence is more likely if H_p is true than if H_d is true;
- an LR < 1 supports the defense hypothesis;
- an LR $= 1$ is consistent with the observations being equally likely under the considered hypotheses.

Reporting LRs

The likelihood ratio is usually reported using phrases such as:

> *"The evidence is . . . more likely if the suspect is the donor of the sample than if someone else is the donor of the sample"*.

It is important to note that the LR is not an absolute measure of the weight of evidence, but is dependent on the underlying hypotheses.

How to express the LR in terms of a verbal 'equivalent'?

Verbal Scales

A verbal scale for evidence interpretation, applied to the prosecution proposition:

Likelihood Ratio	Verbal Equivalent
$1 < \text{LR} \leq 10$	Limited support (for H_p)
$10 < \text{LR} \leq 100$	Moderate support (for H_p)
$100 < \text{LR} \leq 1\,000$	Moderately strong support (for H_p)
$1\,000 < \text{LR} \leq 10\,000$	Strong support (for H_p)
$10\,000 < \text{LR} \leq 1\,000\,000$	Very strong support (for H_p)
$1\,000\,000 < \text{LR}$	Extremely strong support (for H_p)

The equivalent for H_d is given by taking the reciprocal.

Verbal Scales

The association of words with numbers is subjective and arbitrary.

LR	1	$1-10$	$10-10^2$	10^2-10^3	10^3-10^4	10^4-10^6	$>10^6$
Evett & Weir (1998)	–	l	l	m	s	vs	vs
Evett (2000)	–	l	m	ms	s	vs	vs
Martire (2015)	–	w or l	m	ms	s	vs	es
Taroni (2016)	n	l	m	s	vs	es	es

Using verbal scales of neutral (n), weak (w), limited (l), moderate (m), moderately strong (ms), strong (s), very strong (vs) and extremely strong (es).

Presenting Evidence

There are a lot of difficult issues that arise in interpreting DNA samples and presenting complex scientific evidence to non-expert judges and juries.

A sufficiently deep understanding of the principles can help an expert witness to make well-informed judgments and find good solutions to the problem of satisfying goals such as clarity, precision and simplicity.

> *"How forensic evidence is presented is at least as important as what is presented".*

> *". . . it is not only what forensic experts say but how they say it that must be considered".*

Heuristics and Biases

Valid probabilistic reasoning is not easy, so people often use various tricks, rules of thumb, habits, etc., to reason in daily life. These are called *heuristics*.

Heuristics may suffice for most practical situations, but can lead to systematic errors in probabilistic reasoning (i.e. fallacies).

Case Study 1

Quickly read/say the colors of the word:

Case Study 1

Quickly read/say the colors of the word:

RED
ORANGE
YELLOW
GREEN
BLUE
PURPLE

Case Study 1

Quickly read/say the colors of the word:

Case Study 1

Quickly read/say the colors of the word:

RED
ORANGE
YELLOW
GREEN
BLUE
PURPLE

Case Study 1

Quickly read/say the colors of the word:

RED
ORANGE
YELLOW
GREEN
BLUE
PURPLE

Automatic cognitive processes are unintentional and involuntary, and occur outside awareness, probably controlling us more than we want to admit.

Case Study 2

Which option has the most paths? What is the difference?

Option A	Option B
XXXXXXXX	XX
XXXXXXXX	XX
XXXXXXXX	XX
	XX
	XX
	XX
	XX
	XX
	XX

Case Study 2

Option A	Option B
XXXXXXXX	XX
XXXXXXXX	XX
XXXXXXXX	XX
	XX
	XX
	XX
	XX
	XX
	XX

The number of paths is the same for both options:

$$8^3 = 2^9 = 512$$

In a study (Tversky and Kahneman) 85% of respondents found more paths in option A (median: 40) than in option B (median: 18).

This is an example of *availability heuristic*, i.e. the likelihood of an event is estimated as the ease with which examples of such events can be retrieved from memory.

Case Study 3

An unusual disease is expected to kill 600 people. Two alternative programs to combat the disease have been proposed:

- If program A is adopted, 200 people will be saved.
- If program B is adopted, there is a 1/3 chance that all 600 people will be saved and a 2/3 chance that nobody will be saved.

Which program would you choose?

Case Study 3

An unusual disease is expected to kill 600 people. Two alternative programs to combat the disease have been proposed:

- If program C is adopted, 400 people will die.
- If program D is adopted, there is a 1/3 chance that nobody will die and a 2/3 chance that all 600 people will die.

Which program would you choose?

Case Study 3

All four programs have the same expected outcome: 200 people will live, 400 will die.

When framed in terms of gains, 72% choose program A (risk-averse). When framed in terms of losses, 78% choose program D (risk-taking).

Certain gain is preferred over possible gain, while possible loss is preferred over certain loss.

This is an example of the *framing effect*.

Case Study 4

Four cards, each with a letter on one side and a number on the other, are placed on a table. The following hypothesis is proposed:

Every card that has a D on one side has a 3 on the other.

D	K	3	7

Which card(s) need to be turned over to determine whether the hypothesis is true?

Case Study 4

Hypothesis: *Every card that has a D on one side has a 3 on the other.*

D	K	3	7

The correct answer is D and 7. Selecting D and 3 is indicative of *confirmation bias*, i.e. the tendency to search for or interpret information in a way that confirms one's preexisting beliefs or hypotheses, but $\Pr(3|D) \neq \Pr(D|3)$.

Case Study 5

Estimate the number resulting from the following expression:

$$2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

Case Study 5

Estimate the number resulting from the following expression:

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2$$

Case Study 5

Estimate the number resulting from the following expression:

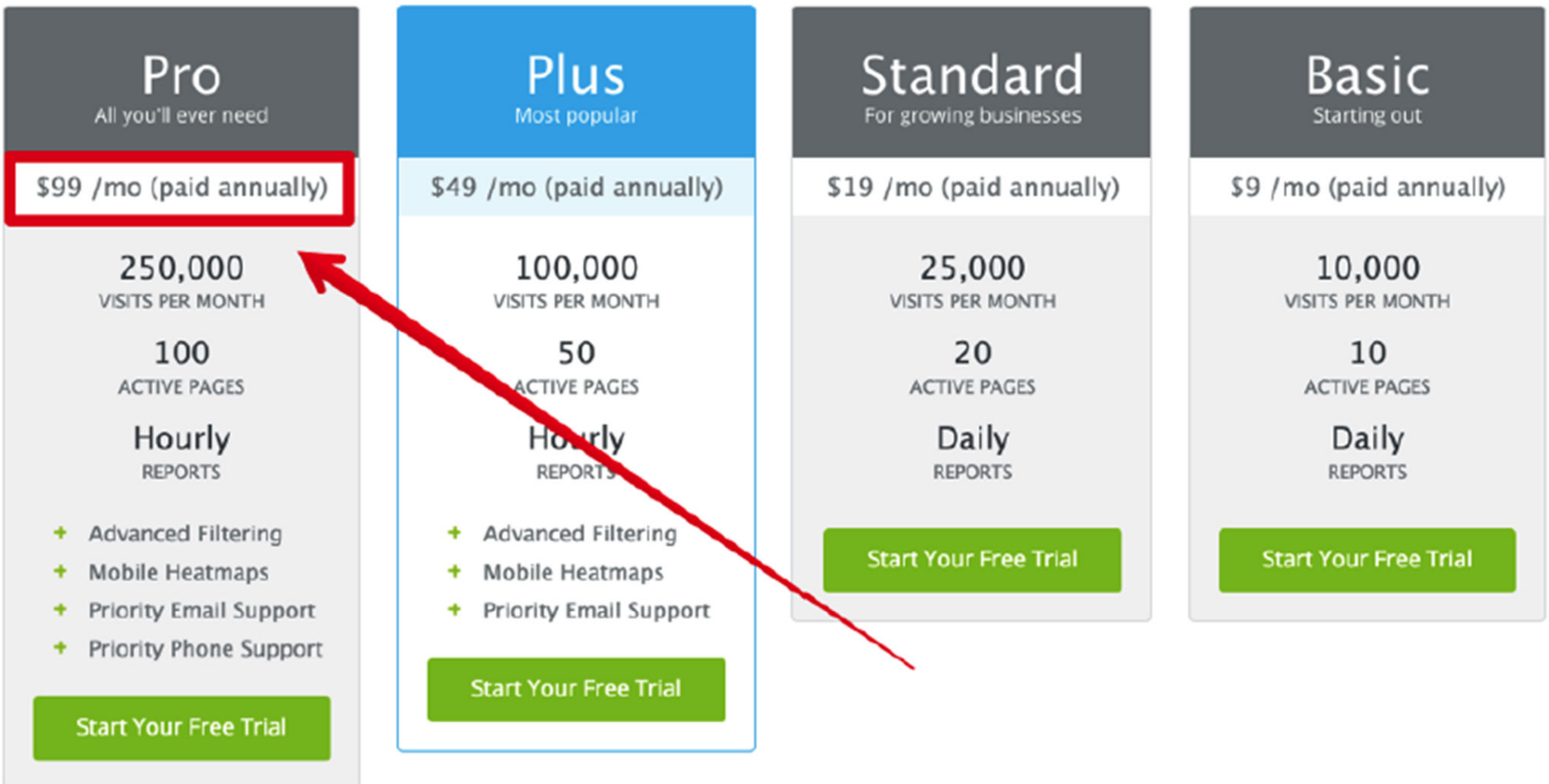
$$2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2$$

Subjects gave a median estimate of 512 in the first case, while the second case had a median of 2 250. The true answer is of course $8! = 40\,320$.

This is an example of *anchoring*, i.e. estimates may depend too much on an initial number.

Case Study 5


Pro
All you'll ever need
$99 /mo (paid annually)
250,000
VISITS PER MONTH
100
ACTIVE PAGES
Hourly
REPORTS
Advanced Filtering
Mobile Heatmaps
Priority Email Support
Priority Phone Support
Start Your Free Trial
Plus
Most popular
$49 /mo (paid annually)
100,000
VISITS PER MONTH
50
ACTIVE PAGES
Hourly
REPORTS
Advanced Filtering
Mobile Heatmaps
Priority Email Support
Start Your Free Trial
Standard
For growing businesses
$19 /mo (paid annually)
25,000
VISITS PER MONTH
20
ACTIVE PAGES
Daily
REPORTS
Start Your Free Trial
Basic
Starting out
$9 /mo (paid annually)
10,000
VISITS PER MONTH
10
ACTIVE PAGES
Daily
REPORTS
Start Your Free Trial

Bias in Forensic Science

- *Attractiveness bias*: Attractive criminals get lower sentences.

- *Target/suspect driven bias*: Using a reference profile to resolve drop-outs.

- *Base rate expectation*: Routinely pairing of examiners and reviewers, high verification rates.

- *Anchoring*: A dice throw influencing sentencing decisions[1].

[1] Playing Dice With Criminal Sentences (Englich et al., 2006).

Bias in Forensic Science

Cognitive bias (i.e. unintentional bias) affects forensic decision-making:

- Biases lead to differences between and within (forensic) experts;
- Bias doesn't necessarily translate into an error in interpretation;
- But cognitive contamination should be avoided just as physical contamination.

This, relatively new, area is often called *cognitive forensics*.

Avoiding Bias

The first step in avoiding cognitive bias is *awareness*: appreciate that it exists, and identify where it resides and affects interpretation, through training and education.

Awareness is necessary, but is insufficient to reduce cognitive bias and contamination: active steps must be taken as mere will power does not control bias.

Several methods have been proposed that can help manage bias sources, such as *Linear Sequential Unmasking*[1].

[1] Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias (Dror, 2017).

Bias in Forensic Science

What about probabilistic genotyping software?

- Interpretation software can reduce variation in interpretation among examiners.
- It does *not* make interpretation bias free;
- Subjectivity is also involved in software development (and underlying modeling).
- Different software can show LRs varying over 10 logs for the same DNA profiles.

Fallacies

Biases can lead to potential fallacies in the courtroom, and may even lead to a miscarriage of justice[1].

- Prosecutor's fallacy
- Defendant's fallacy
- Uniqueness fallacy
- Association fallacy

[1] See also Misleading DNA Evidence (Gill, 2014).

Prosecutor's Fallacy

One of the most common errors is to transpose the conditional:

$$\text{Pr}(A|B) \neq \text{Pr}(B|A),$$

e.g. saying that there is a very high probability that an animal has four legs *if* it is an elephant, is not the same as the probability that an animal is an elephant *if* it has four legs.

$$\text{Pr}(4 \text{ legs} \mid \text{Elephant}) \neq \text{Pr}(\text{Elephant} \mid 4 \text{ legs}).$$

Prosecutor's Fallacy

This example may seem obvious, but it's often not so easy in court proceedings:

$$\Pr(E|H_p) \neq \Pr(H_p|E),$$

or, alternatively,

$$\begin{aligned}\Pr(\text{Evidence} \mid \text{Proposition}) &\neq \Pr(\text{Proposition} \mid \text{Evidence}) \\ &\neq \Pr(\text{Proposition})\end{aligned}$$

Prosecutor's Fallacy

- Subtle misstatements can lead (and have led) to misunderstandings.

- Forensic scientists should be (and are trained to be) very careful about the wording of probability statements.

Defendant's Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 1 000. The defendant's fallacy is a logical error that usually favors the defendant:

- The city where the crime occurred has population size 100 000;
- So there are 100 people with a matching profile;
- This means that $\Pr(H_p|E)$ is only 1 in 100 or 1%.

Defendant's Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 1 000. The defendant's fallacy is a logical error that usually favors the defendant:

- The city where the crime occurred has population size 100 000;
- So there are 100 people with a matching profile;
- This means that $\Pr(H_p|E)$ is only 1 in 100 or 1%.

Defendant's Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 1 000. The defendant's fallacy is a logical error that usually favors the defendant:

- The city where the crime occurred has population size 100 000;
- So we *expect* 100 people with a matching profile;
- $\Pr(H_p|E)$ is 1 in 100 or 1% *only* if each of these individuals has the same prior probability.

Uniqueness Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 100 000. The uniqueness fallacy argues:

- The city where the crime occurred has population size 100 000;
- So there is only one individual with a matching profile;
- This means that this DNA profile is unique in this city and must come from the suspect.

Uniqueness Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 100 000. The uniqueness fallacy argues:

- The city where the crime occurred has population size 100 000;
- So there is only one individual with a matching profile;
- This means that this DNA profile is unique in this city and must come from the suspect.

Uniqueness Fallacy

Suppose $\Pr(E|H_d)$ is reported as 1 in 100 000.

- The city where the crime occurred has population size 100 000;
- So we *expect* 1 *other* individual with a matching profile;
- This usually also incorporates the belief that DNA profiles yield unique identification, which is untrue in light of LTDNA, often leading to complex mixtures and partial profiles (and ignores relatives, coancestry and phenomena such as drop-in).

Association Fallacy

An association fallacy occurs when a probability statement is transposed from one scale of the hierarchy of propositions to a higher level.

This is usually a result from assuming that there is a dependency between two observations or events, e.g.:

- Statements about evidence samples (sub-source) that are interpreted as the 'evidence being more likely if the suspect is the *source* of the crime stain';

- Or even on *activity* level as 'the evidence is more likely if the suspect left the crime stain'.

Miscarriage of Justice - Example

Adam Scott was arrested, accused of rape and incarcerated on the basis of a DNA profile match, which was eventually traced back to a contamination incident.

> *"It is estimated that the chance of obtaining matching DNA components if the DNA came from someone else unrelated to Adam Scott is approximately one in 1 billion. In my opinion the DNA matching that of Adam Scott has most likely originated from semen. [. . .] In my opinion these findings are what I would expect if Adam Scott had some form of sexual activity with [the victim]. In order to assess the overall findings in this case I have therefore considered the following propositions:*
>
> - *Adam Scott had vaginal intercourse with [the victim]*
> - *Adam Scott has never been to Manchester and does not know [the victim]"*

Source: Misleading DNA Evidence (Gill, 2014).

Miscarriage of Justice - Example

- The perpetrator DNA was absent (hidden perpetrator effect and false inclusion error).
- The DNA match was falsely associated with the presence of sperm (association fallacy).
- The 'presence' of sperm was associated with sexual intercourse (association fallacy).
- Exculpatory evidence was ignored (base rate fallacy and confirmation bias).

Different biases/effects resulted in a *compounded error* or *snowball effect*.

The Innocence Project

The Innocence Project was founded in 1992 as a non-profit legal organization committed to exonerating wrongly convicted people. The work focuses on cases in which DNA evidence is available to be tested or retested.

- There have been 367 post-conviction exonerations due to DNA testing to date (July, 2020);
- Incorrect identification by eyewitnesses was a factor in around 70% of wrongful convictions;
- Of those exonerated 70% are part of minority groups;

Source: `https://www.innocenceproject.org`.