Section 6: Reporting Likelihood Ratios

Components

- Hierarchy of propositions
- Formulating propositions
- Communicating LRs

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

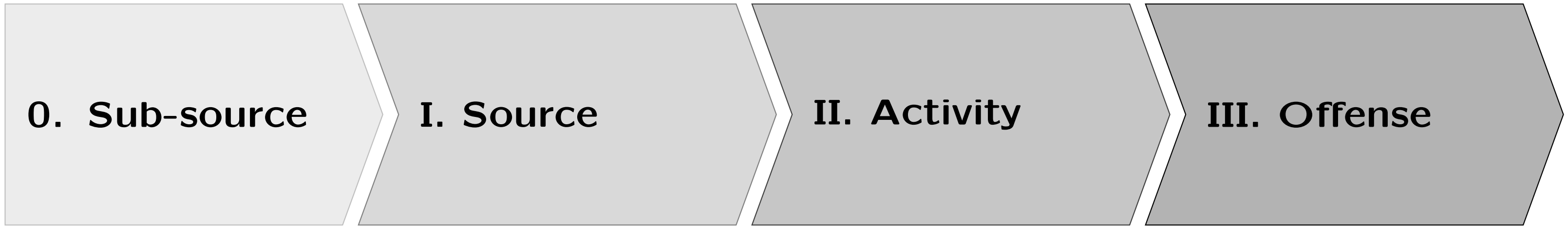
Evett & Cook (1998) established the following hierarchy of propositions:

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

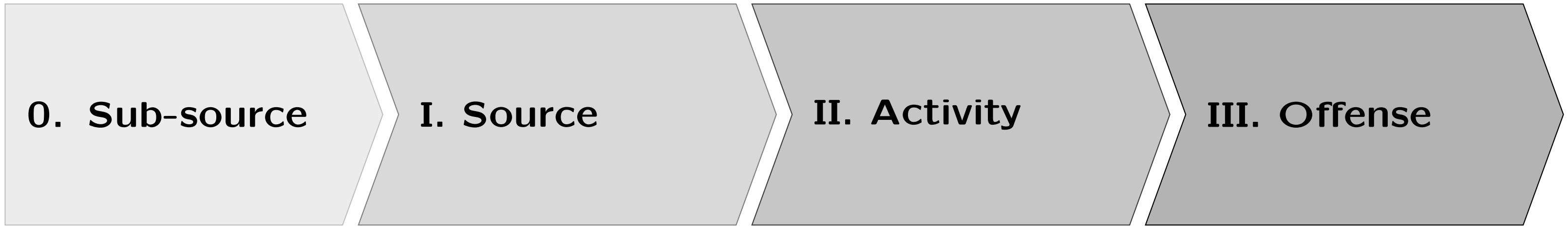
- The *offense* level deals with the ultimate issue of guilt/ innocence, which are outside the domain of the forensic scientist.
- The *activity* level associates a DNA profile or evidence source with the crime itself, and there may be occasions where a scientist can address this level.
- The *source* 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.

Hierarchy of Propositions



- Probabilistic genotyping is (usually) centered around sub-source level.

- Transition from sub-source to source or even activity level may be possible, e.g. by considering contamination, secondary transfer, timing, etc.

Setting Propositions

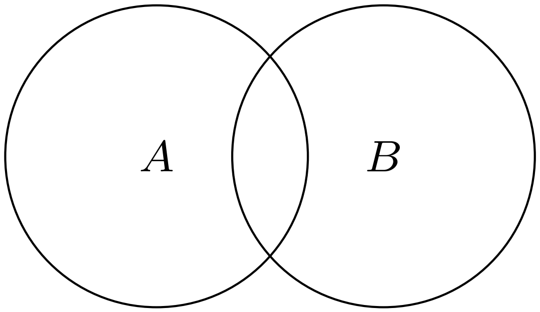
Some useful principles for setting hypotheses:

- 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.

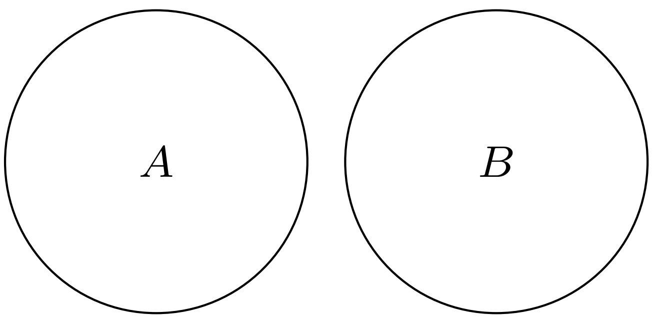
MECE Definition

Mutually exclusive

(i.e. non-overlapping)



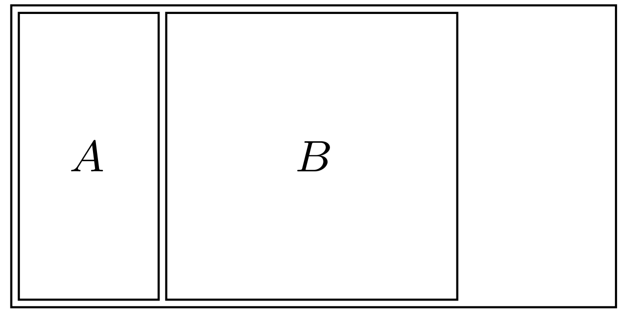
Not exclusive



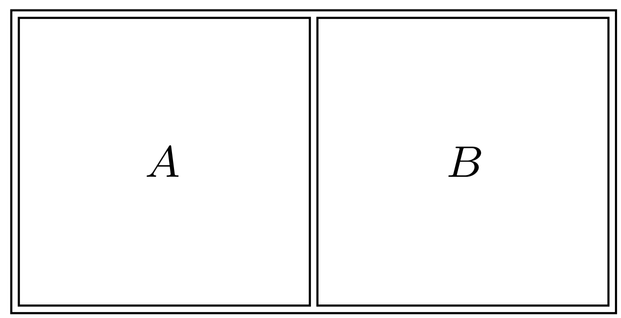
Exclusive

Collectively exhaustive

(i.e. covers all outcomes)



Not exhaustive



Exhaustive

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 - Example 1

An individual is discovered looking into a house one night. The police are called and find a single cigarette butt under the window where the incident occurred. No one in the family smokes. The police have a person of interest captured on a neighbor's CCTV.

A single-source profile is obtained from the cigarette butt and the reference profile of a person of interest (POI) matches.

Formulating Propositions - Example 1

An individual is discovered looking into a house one night. The police are called and find a single cigarette butt under the window where the incident occurred. No one in the family smokes. The police have a person of interest captured on a neighbor's CCTV.

A single-source profile is obtained from the cigarette butt and the reference profile of a person of interest (POI) matches.

H_p : The evidence came from the POI.
H_d : The evidence came from an unknown person.

Or, for simplicity:

H_p : POI
H_d : Unknown (U)

Formulating Propositions - Example 2

A complainant calls 911 to report a sexual assault in her home. She is taken to a hospital where an intimate swab is collected.

A POI is identified from the investigation and the obtained profile from the swab is fully explained by a mixture of the complainant (K) and the POI.

Formulating Propositions - Example 2

A complainant calls 911 to report a sexual assault in her home. She is taken to a hospital where an intimate swab is collected.

A POI is identified from the investigation and the obtained profile from the swab is fully explained by a mixture of the complainant (K) and the POI.

$$H_p: \quad \text{K} + \text{POI}$$
$$H_d: \quad \text{K} + \text{U}$$

Formulating Propositions - Example 3

A complainant is cut with a knife during an altercation. Based upon eyewitness testimony, a POI is identified.

A stain on the clothing of the POI is tested for blood, and a DNA profile is developed that is consisted with a mixture of the POI and the complainant.

Formulating Propositions - Example 3

A complainant is cut with a knife during an altercation. Based upon eyewitness testimony, a POI is identified.

A stain on the clothing of the POI is tested for blood, and a DNA profile is developed that is consisted with a mixture of the POI and the complainant.

$$H_p: \quad \text{POI} + \text{K}$$
$$H_d: \quad \text{POI} + \text{U}$$

Note how the direction of transfer provides important information.

Formulating Propositions - Example 4

Molotov cocktails have been thrown at random cars. An unexploded container is found in the street, and a 2 person mixture is developed from the evidence.

Two persons of interest are arrested.

Formulating Propositions - Example 4

Molotov cocktails have been thrown at random cars. An unexploded container is found in the street, and a 2 person mixture is developed from the evidence.

Two persons of interest are arrested.

$$\begin{aligned} H_p: &\quad \text{POI 1} + \text{POI 2} \\ H_{d1}: &\quad \text{POI 1} + \text{U} \\ H_{d2}: &\quad \text{POI 2} + \text{U} \\ H_{d3}: &\quad \text{2U} \end{aligned}$$

What if circumstances indicate that they cannot both be present?

Formulating Propositions - Example 5

A complainant walking through a city park is attacked from behind and is sexually assaulted on a blanket. She didn't get a good look at the perpetrator. The police recognize the blanket as possibly belonging to a vagrant known to live near the park.

A profile obtained from the blanket is fully explained by mixing of K and POI's DNA.

Formulating Propositions - Example 5

A complainant walking through a city park is attacked from behind and is sexually assaulted on a blanket. She didn't get a good look at the perpetrator. The police recognize the blanket as possibly belonging to a vagrant known to live near the park.

A profile obtained from the blanket is fully explained by mixing of K and POI's DNA.

$$
\begin{aligned}
H_p: &\quad \text{K + POI} \\
H_{d1}: &\quad \text{POI + U} \\
H_{d2}: &\quad \text{K + U} \\
H_{d3}: &\quad \text{2U}
\end{aligned}
$$

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.

Note that if K is a true source of the profile, but not considered under H_d, the LR will be larger than when assuming K as a known profile under both hypotheses. This is because K will explain many of the observed alleles (especially in case of being a major donor).

The Effect of Propositions on the LR

Consider a simple two-person mixture profile (e.g. contributors are unrelated, ignoring population structure, no drop-outs/drop-ins), where $G_C = ABCD$. Let K denote a known contributor with observed profile $G_K = CD$, and S the POI with profile $G_S = AB$.

- LR $= \frac{\text{Pr}(ABCD|H_p\text{: } K+S)}{\text{Pr}(ABCD|H_d\text{: } K+U)}$

- LR $= \frac{\text{Pr}(ABCD|H_p\text{: } K+S)}{\text{Pr}(ABCD|H_d\text{: } 2U)}$

- LR $= \frac{\text{Pr}(ABCD|H_p\text{: } S+U)}{\text{Pr}(ABCD|H_d\text{: } 2U)}$

What are the correct expressions for the LR?

The Effect of Propositions on the LR

Consider a simple two-person mixture profile (e.g. contributors are unrelated, ignoring population structure, no drop-outs/drop-ins), where $G_C = ABCD$. Let K denote a known contributor with observed profile $G_K = CD$, and S the POI with profile $G_S = AB$.

- LR $= \frac{\Pr(ABCD|H_p\colon\ K+S)}{\Pr(ABCD|H_d\colon\ K+U)} = \frac{1}{2p_A p_B}$

- LR $= \frac{\Pr(ABCD|H_p\colon\ K+S)}{\Pr(ABCD|H_d\colon\ 2U)} = \frac{1}{6{\cdot}4p_A p_B p_C p_D} = \frac{1}{24p_A p_B p_C p_D}$

- LR $= \frac{\Pr(ABCD|H_p\colon\ S+U)}{\Pr(ABCD|H_d\colon\ 2U)} = \frac{2p_C p_D}{6{\cdot}4p_A p_B p_C p_D} = \frac{1}{12p_A p_B}$

What are the LRs for $p_A = p_B = p_C = p_D = 0.1$?

The Effect of Propositions on the LR

Consider a simple two-person mixture profile (e.g. contributors are unrelated, ignoring population structure, no drop-outs/drop-ins), where $G_C = ABCD$. Let K denote a known contributor with observed profile $G_K = CD$, and S the POI with profile $G_S = AB$.

- LR $= \frac{\Pr(ABCD|H_p\colon K+S)}{\Pr(ABCD|H_d\colon K+U)} = \frac{1}{2p_Ap_B} = 50$

- LR $= \frac{\Pr(ABCD|H_p\colon K+S)}{\Pr(ABCD|H_d\colon 2U)} = \frac{1}{24p_Ap_Bp_Cp_D} = 417$

- LR $= \frac{\Pr(ABCD|H_p\colon S+U)}{\Pr(ABCD|H_d\colon 2U)} = \frac{1}{12p_Ap_B} = 8$

Formulating Propositions

What about the number of contributors?
This is an important component of mixture interpretation. Most approaches assume that the NoC is known.

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.

Formulating Propositions - NoC

- 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).

Formulating Propositions - NoC

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

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

where prior independence is assumed:

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

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 - NoC

- 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.

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 $\text{LR} > 1$ supports the prosecution hypothesis, meaning that the evidence is more likely if H_p is true than if H_d is true;
- an $\text{LR} < 1$ supports the defense hypothesis;
- an $\text{LR} = 1$ 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**
$1 <$ LR ≤ 10	Limited support (for H_p)
$10 <$ LR ≤ 100	Moderate support (for H_p)
$100 <$ LR $\leq 1\,000$	Moderately strong support (for H_p)
$1\,000 <$ LR $\leq 10\,000$	Strong support (for H_p)
$10\,000 <$ LR $\leq 1\,000\,000$	Very strong support (for H_p)
$1\,000\,000 <$ 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).

Verbal Scales

Should we report a verbal equivalent for the LR?

Verbal Scales

Should we report a verbal equivalent for the LR?

- **Yes**: The verbal scale is helpful for the jury to put the LR into perspective.

- **No**: The verbal scale is not the responsibility of the forensic scientist.

SWGDAM states that a verbal scale should always be accompanied by a qualitative statement.

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:

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:

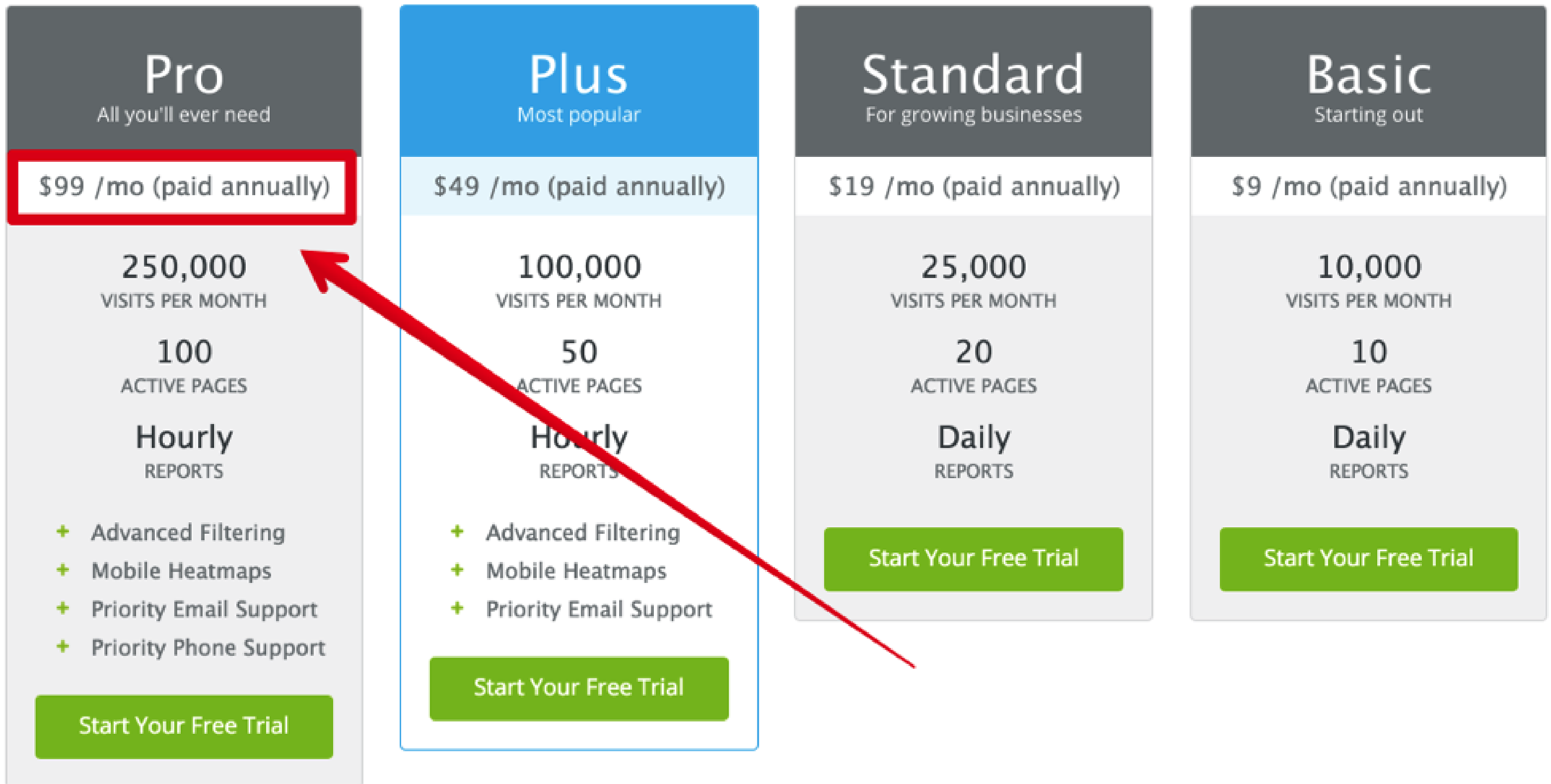
$$2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

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

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 5b



Case Study 6

- Of the women complaining of painful hardening of the breast, 1% have a malignant tumor: $\Pr(C) = 0.01$.

- The accuracy ($+$ or $-$) of a mammography is 90%: $\Pr(+|C) = \Pr(-|C') = 0.9$.

- Estimate $\Pr(C|+)$ to decide whether or not to order a biopsy.

Case Study 6

Most physicians estimate $\Pr(C|+) \approx 0.75$, while the correct answer is:

$$\Pr(C|+) = \frac{\Pr(+|C)\Pr(C)}{\Pr(+|C)\Pr(C) + \Pr(+|C')\Pr(C')} = 0.0833.$$

Representativeness leads people to neglect the base rate, by assessing a conditional probability by the 'degree of similarity' ($\Pr(A|B) \neq \Pr(B|A)$). This is known as the *base rate fallacy*.

Case Study 6

Using the odds form of Bayes' theorem:

$$\frac{\text{Pr}(C)}{\text{Pr}(C')} = \frac{1}{99}$$

$$\frac{\text{Pr}(+|C)}{\text{Pr}(+|C')} = \frac{0.9}{0.1} = 9$$

Even though the LR > 1, the prior odds (i.e. the base rate) is relatively small. The posterior odds are $\frac{1}{11}$, such that $\text{Pr}(C|+) = \frac{1}{12} = 0.0833$.

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, 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.

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

Prosecutor's Fallacy

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

$$\mathsf{Pr}(E|H_p) \neq \mathsf{Pr}(H_p|E),$$

or, alternatively,

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

Prosecutor's Fallacy - Exercise

- The evidence is much more likely if the DNA profile came from the suspect.
- The probability of this DNA profile if it came from someone else is very low.
- The probability that this DNA profile came from someone else is very low.
- The probability of someone else having this DNA profile is very low.
- The probability of someone else leaving DNA of this type is very low.
- The evidence strongly supports the hypothesis that the DNA profile came from the suspect.

Prosecutor's Fallacy - Exercise

- The evidence is much more likely if the DNA profile came from the suspect.
- The probability of this DNA profile if it came from someone else is very low.
- The probability that this DNA profile came from someone else is very low.
- The probability of someone else having this DNA profile is very low.
- The probability of someone else leaving DNA of this type is very low.
- The evidence strongly supports the hypothesis that the DNA profile came from the suspect.

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'.

Fallacies in Practice - Case Example

The *People v. Nelson* (CA) court's decision report contains the following statements:

> *"In 2002, investigators compared evidence from a 1976 murder scene with defendant's deoxyribonucleic acid (DNA) profile and identified him as a possible donor of that evidence. He was then tried for and convicted of that murder. The prosecution presented evidence that the odds that a random person unrelated to defendant from the population group that produced odds most favorable to him could have fit the profile of some of the crime scene evidence are one in 930 sextillion (93 followed by 22 zeros)."*

> *"Because the worlds total population is only about seven billion (seven followed by nine zeros), this evidence is tantamount to saying that defendant left the evidence at the crime scene."*

> *". . . We also conclude that the jury properly heard evidence that it was virtually impossible that anyone other than defendant could have left the evidence found at the crime scene."*

Fallacies in Practice - Case Example

The *People v. Nelson* (CA) court's decision report contains the following statements:

> *". . . Specifically, [the defendant] contends the evidence regarding the odds that the crime scene evidence could have come from some other person was inadmissible because the statistical method used to calculate those odds has not achieved general scientific acceptance under the standard stated in [. . .] People v. Kelly (1976) 17 Cal.3d 24 (sometimes referred to as the Kelly test)."*

> *". . . Defendant agrees that using the product rule to calculate the random match probability makes sense when comparing one suspects profile with the crime scene evidence because, as he explains, the random match probability "estimates the chance that any single, random person drawn from the relevant population would have the same DNA profile as that of the unknown person whose DNA was found at the crime scene.""*

> *". . . It is already settled that the product rule reliably shows the rarity of the profile in the relevant population. [. . .] To this extent, the product rule has already passed the Kelly test."*

Fallacies in Practice - Case Example

The *People v. Nelson* (CA) court's decision report contains the following statements:

> *"The Court of Appeal in this case and other courts that have considered this question have concluded that use of the product rule in a cold hit case is not the application of a new scientific technique subject to a further Kelly (or Kelly-like) test."*

> *"We agree. Jenkins explained its reasoning: "At the heart of this debate is a disagreement over the competing questions to be asked, not the methodologies used to answer those questions. [. . .] [T]here is no controversy in the relevant scientific community as to the accuracy of the various formulas. In other words, the math that underlies the calculations is not being questioned. [. . .] [T]he debate . . . is one of relevancy, not methodology . . . .""*

> *". . . The debate that exists is solely concerned with which number – rarity, database match probability, Balding-Donnelly, or some combination of the above is most relevant in signifying the importance of a cold hit. "*

Fallacies in Practice - Case Example

The *People v. Nelson* (CA) court's decision report contains the following statements:

> *"The database match probability ascertains the probability of a match from a given database. "But the database is not on trial. Only the defendant is". Thus, the question of how probable it is that the defendant, not the database, is the source of the crime scene DNA remains relevant. The rarity statistic addresses this question."*

> *"The fact that the match ultimately came about by means of a database search does not deprive the rarity statistic of all relevance. It remains relevant for the jury to learn how rare this particular DNA profile is within the relevant populations and hence how likely it is that someone other than defendant was the source of the crime scene evidence. Accordingly, the trial court correctly admitted the evidence, and the Court of Appeal correctly upheld that admission."*

Miscarriage of Justice - Case Example 1

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 - Case Example 1

- 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*.

Miscarriage of Justice - Case Example 2

The association fallacy assumes a dependency between two observation or events. The opposite version may also lead to errors, i.e. assuming independence where non exists.

Sally Clark was arrested and convicted for the murder of her two infant sons. In this case (UK, 1999) it was assumed that two sudden infant death syndrome (SIDS) deaths in a single family were independent events. A consulting pediatrician estimated the likelihood of a cot death as 1 in 8 500, and calculated the combined probability by squaring this number (i.e. yielding a likelihood of 1 in 73 million).

Miscarriage of Justice - Case Example 2

It was later found that her second son might have died from natural causes, and moreover, assuming independence of these events is unreasonable, due to possible underlying genetic causes:

$$P(A, B) = P(A|B)P(B) \neq P(A)P(B).$$

Sally Clark was released from prison after having served more than three years of her sentence.

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 362 post-conviction exonerations due to DNA testing as of January 2019;
- Incorrect identification by eyewitnesses was a factor in over 70% of wrongful convictions;
- Of those exonerated 70% are part of minority groups;

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