Introduction to Logical Reasoning for the Evaluation of Forensic Evidence

Geoffrey Stewart Morrison

\[ \frac{p(E|H_p)}{p(E|H_d)} \]

http://forensic-evaluation.net/
Abstract

In response to the 2009 US National Research Council (NRC) Report on Strengthening Forensic Science in the United States, the 2010 England & Wales Court of Appeal ruling in *R v T*, and the 2012 US National Institute of Standards and Technology and National Institute of Justice (NIST/NIJ) report on latent fingerprint analysis there is increasing pressure across all branches of forensic science to adopt a paradigm consisting of the following three elements:

1. a logically correct framework for the evaluation and interpretation of forensic evidence
2. approaches based on relevant data, quantitative measurements, and statistical models
3. empirical testing of the degree of validity and reliability of forensic-evaluation systems under conditions reflecting those of the case under investigation

This workshop provides an introduction to the likelihood-ratio framework as the first element in this paradigm, and begins to touch on the second element. There is a great deal of misunderstanding and confusion about the likelihood-ratio framework among lawyers, judges, and forensic scientists. The likelihood-ratio framework is about logic, not mathematics or databases, and it makes explicit the questions which must logically be addressed by the forensic scientist and considered by the trier of fact in assessing the work of the forensic scientist. This workshop explains the logic of the likelihood-ratio framework in a way which is accessible to a broad audience and which does not require any prior knowledge about the framework. It uses intuitive examples and audience-participation exercises to gradually build a fuller understanding of the likelihood-ratio framework. The workshop also includes discussion of common logical fallacies.
Imagine you are driving to the airport...
Imagine you are driving to the airport...
Imagine you are driving to the airport...
Imagine you are driving to the airport...

Initial probabilistic belief + evidence → Updated probabilistic belief

higher? or lower?
Imagine you are driving to the airport...

\[
\text{initial probabilistic belief} + \text{evidence} \rightarrow \text{updated probabilistic belief}
\]

Higher? or lower?
• This is Bayesian reasoning
  – It is about logic
  – It is not about mathematical formulae or databases
  – There is nothing complicated or unnatural about it
  – It is the logically correct way to think about many problems
Imagine you work at a shoe recycling depot...

- You pick up two shoes of the same size
  - Does the fact that they are of the same size mean they were worn by the same person?
  - Does the fact that they are of the same size mean that it is highly probable that they were worn by the same person?
Imagine you work at a shoe recycling depot...

- You pick up two shoes of the same size
  - Does the fact that they are of the same size mean they were worn by the same person?
  - Does the fact that they are of the same size mean that it is highly probable that they were worn by the same person?

- Both similarity and typicality matter
Imagine you are a forensic shoe comparison expert...

suspect’s shoe

crime-scene footprint
Imagine you are a forensic shoe comparison expert...

- The footprint at the crime scene is size 10

- The suspect’s shoe is size 10
  
  – What is the probability of the footprint at the crime scene being size 10 if it had been made by the suspect’s shoe? (similarity)

- Half the shoes at the recycling depot are size 10
  
  – What is the probability of the footprint at the crime scene being size 10 if it had been made by the someone else’s shoe? (typicality)
Imagine you are a forensic shoe comparison expert...

- The footprint at the crime scene is size 14
- The suspect’s shoe is size 14
  - What is the probability of the footprint at the crime scene being size 14 if it had been made by the suspect’s shoe? (similarity)
- 1% of the shoes at the recycling depot are size 14
  - What is the probability of the footprint at the crime scene being size 14 if it had been made by the someone else’s shoe? (typicality)
Imagine you are a forensic shoe comparison expert...

- The footprint at the crime scene and the suspect’s shoe are size 10

\[
\frac{\text{similarity}}{\text{typicality}} = \frac{1}{0.5} = 2
\]

you are twice as likely to get a size 10 footprint at the crime scene if it were produced by the suspect’s shoe than if it were produced by somebody else’s shoe.
Imagine you are a forensic shoe comparison expert...

- The footprint at the crime scene and the suspect’s shoe are size 14

\[
\text{similarity} / \text{typicality} = 1 / 0.01 = 100
\]

you are 100 times as likely to get a size 14 footprint at the crime scene if it were produced by the suspect’s shoe than if it were produced by somebody else’s shoe.
Imagine you are a forensic shoe comparison expert...

- size 10
  \[
  \text{similarity} / \text{typicality} = 1 / 0.5 = 2
  \]

- size 14
  \[
  \text{similarity} / \text{typicality} = 1 / 0.01 = 100
  \]

- If you didn’t have a database, could you have made subjective estimates at relative proportions of different shoe sizes in the population and applied the same logic to arrive at a conceptually similar answer?
¿Area?
similarity / typicality = likelihood ratio
Given that it is a cow, what is the probability of it having four legs?

\[ p(4 \text{ legs} \mid \text{cow}) = ? \]
Given that it has four legs, what is the probability that it is a cow?

\[ p( \text{cow} \mid 4 \text{ legs} ) = ? \]
Given two voice samples with acoustic properties $x_1$ and $x_2$, what is the probability that they were produced by the same speaker?

$$p(\text{same speaker} \mid \text{acoustic properties } x_1, x_2) = ?$$
\[ p( \text{same speaker} \mid \text{acoustic properties } x_1, x_2 ) = ? \]

\[ p( \text{same wearer} \mid \text{shoe size } x, \text{footprint size } x ) = ? \]

\[ p( \text{cow} \mid x \text{ legs} ) = ? \]
Bayes’ Theorem

\[
\textit{posterior odds} = \frac{p(\text{same speaker} \mid \text{acoustic properties } x_1, x_2)}{p(\text{different speaker} \mid \text{acoustic properties } x_1, x_2)}
\]

\[
= \frac{p(\text{acoustic properties } x_1, x_2 \mid \text{same speaker})}{p(\text{acoustic properties } x_1, x_2 \mid \text{different speaker})} \times \frac{p(\text{same speaker})}{p(\text{different speaker})}
\]

\textit{likelihood ratio}

\textit{prior odds}
Bayes’ Theorem

initial probabilistic belief + evidence → updated probabilistic belief

higher? or lower?
However !!!

The forensic scientist acting as an expert witness can NOT give the posterior probability. They can NOT give the probability that two speech samples were produced by the same speaker.
Why not?

- The forensic scientist does not know the prior probabilities.

- Considering all the evidence presented so as to determine the posterior probability of the prosecution hypothesis and whether it is true beyond a reasonable doubt (or on the balance of probabilities) is the task of the trier of fact (judge, panel of judges, or jury), not the task of the forensic scientist.

- The task of the forensic scientist is to present the strength of evidence with respect to the particular samples provided to them for analysis. They should not consider other evidence or information extraneous to their task.
Bayes’ Theorem

_posterior odds_

\[
\frac{p(\text{same speaker} | \text{acoustic properties } x_1, x_2)}{p(\text{different speaker} | \text{acoustic properties } x_1, x_2)}
\]

= 

\[
\frac{p(\text{acoustic properties } x_1, x_2 | \text{same speaker})}{p(\text{acoustic properties } x_1, x_2 | \text{different speaker})} \times \frac{p(\text{same speaker})}{p(\text{different speaker})}
\]

*likelihood ratio*

*prior odds*
Bayes’ Theorem

\[
p(\text{same speaker} | \text{acoustic properties } x_1, x_2 ) \quad \text{posterior odds}\n\]

\[
\frac{p(\text{same speaker} | \text{acoustic properties } x_1, x_2 )}{p(\text{different speaker} | \text{acoustic properties } x_1, x_2 )} = \quad \text{likelihood ratio}
\]

\[
\frac{p(\text{acoustic properties } x_1, x_2 | \text{same speaker})}{p(\text{acoustic properties } x_1, x_2 | \text{different speaker})} \times \frac{p(\text{same speaker})}{p(\text{different speaker})} \quad \text{prior odds}
\]
Bayes’ Theorem

\[
\frac{p(\text{same speaker} | \text{acoustic properties } x_1, x_2)}{p(\text{different speaker} | \text{acoustic properties } x_1, x_2)} = \frac{p(\text{acoustic properties } x_1, x_2 | \text{same speaker})}{p(\text{acoustic properties } x_1, x_2 | \text{different speaker})} \times \frac{p(\text{same speaker})}{p(\text{different speaker})}
\]
Likelihood Ratio

\[
\frac{p(\text{acoustic properties } x_1, x_2 \mid \text{same speaker})}{p(\text{acoustic properties } x_1, x_2 \mid \text{different speaker})}
\]

\[
\frac{p(\text{shoe size } x, \text{ footprint size } x \mid \text{same wearer})}{p(\text{shoe size } x, \text{ footprint size } x \mid \text{different wearer})}
\]

\[
\frac{p(\text{x legs} \mid \text{cow})}{p(\text{x legs} \mid \text{not a cow})}
\]

\[
\frac{p(\text{E} \mid \text{H}_{\text{prosecution}})}{p(\text{E} \mid \text{H}_{\text{defence}})}
\]
Example

- You would be **4 times more likely** to get the **acoustic properties of the voice on the offender recording if it were produced by the suspect** than if it were produced by some other speaker from the relevant population.

- Whatever the trier of fact’s belief as to the relative probabilities of the same-speaker versus the different-speaker hypotheses before being presented with the likelihood ratio, after being presented with the likelihood ratio their relative belief in the probability that the voices on the two recordings belong to the same speaker versus the probability that they belong to different speakers should be 4 times greater than it was before.

These examples are intended to illustrate the logic of the expression of the strength of the evidence in the likelihood-ratio framework. They are not meant to imply that jury members must assign precise numbers to their beliefs nor that they must apply mathematical formulae.
Example: The evidence is 4 time more likely given the same-speaker hypothesis than given the different-speaker hypothesis.

Before

- different: 1
- same: 1

If before you believed that the same-speaker and different-speaker hypotheses were equally probable.

After

- different: 1
- same: 4

Now you believe that the same-speaker hypothesis is 4 times more probable than the different-speaker hypothesis.

Multiply this weight by 4.
Example: The evidence is 4 time more likely given the same-speaker hypothesis than given the different-speaker hypothesis.

Before:

- If before you believed that the same-speaker hypothesis was 2 times more probable than the different-speaker hypotheses.

After:

- Now you believe that the same-speaker hypothesis is 8 times more probable than the different-speaker hypothesis.
Example: The evidence is 4 time more likely given the same-speaker hypothesis than given the different-speaker hypothesis.

Before

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

if before you believed that the different-speaker hypothesis was 2 times more probable than the same-speaker hypotheses

After

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

multiply this weight by 4

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

now you believe that the same-speaker hypothesis is 2 times more probable than the different-speaker hypothesis

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example: The evidence is 4 time more likely given the same-speaker hypothesis than given the different-speaker hypothesis.

If before you believed that the different-speaker hypothesis was 8 times more probable than the same-speaker hypotheses, now you believe that the different-speaker hypothesis is 2 times more probable than the same-speaker hypothesis.
Example

- You would be **4 times more likely** to get the **acoustic properties of the voice on the offender recording** if it were produced by the suspect than if it were produced by some other speaker from the relevant population.

- Whatever the trier of fact’s belief as to the relative probabilities of the same-speaker versus the different-speaker hypotheses before being presented with the likelihood ratio, after being presented with the likelihood ratio their relative belief in the probability that the voices on the two recordings belong to the same speaker versus the probability that they belong to different speakers should be 4 times greater than it was before.

These examples are intended to illustrate the logic of the expression of the strength of the evidence in the likelihood-ratio framework. They are not meant to imply that jury members must assign precise numbers to their beliefs nor that they must apply mathematical formulae.
Likelihood Ratio

\[
p( \text{acoustic properties } x_1, x_2 \mid \text{same speaker} )
\]
\[
p( \text{acoustic properties } x_1, x_2 \mid \text{different speaker} )
\]

\[
p( \text{shoe size } x, \text{footprint size } x \mid \text{same wearer})
\]
\[
p( \text{shoe size } x, \text{footprint size } x \mid \text{different wearer})
\]

\[
p( x \text{ legs} \mid \text{cow} )
\]
\[
p( x \text{ legs} \mid \text{not a cow} )
\]

\[
p( E \mid H_{\text{prosecution}} )
\]
\[
p( E \mid H_{\text{defence}} )
\]
Discrete data: bar graph

- **Proportion**
- **Legs**

- **Cows**
- **Not cows**

The bar graph shows the proportion of cows and not cows with different numbers of legs. The highest proportion of cows is with 4 legs, while the other categories have much lower proportions.
\[
p(\text{4 legs} \mid \text{cow}) \quad \frac{0.98}{0.49}
\]

\[
p(\text{4 legs} \mid \text{not a cow})
\]

The diagram shows a bar chart with the x-axis representing the number of legs (1 to 8) and the y-axis representing the proportion. The chart compares the proportions of cows and not cows for different numbers of legs, with 4 legs being the most common for both categories.
\[
p\left(\frac{4 \text{ legs} \mid \text{cow}}{4 \text{ legs} \mid \text{not a cow}}\right) = \frac{0.98}{0.49} = 2
\]
Imagine you are a forensic hair comparison expert...

- The hair at the crime scene is blond
- The suspect has blond hair
- What do you do?
Imagine you are a forensic hair comparison expert...

- The hair at the crime scene is blond
- The suspect has blond hair

\[
p(\text{blond hair at crime scene} | \text{suspect is source})
\]

\[
p(\text{blond hair at crime scene} | \text{someone else is source})
\]
Imagine you are a forensic hair comparison expert...

- The hair at the crime scene is blond
- The suspect has blond hair

\[
p( \text{blond hair at crime scene} \mid \text{suspect is source})
\]

\[
p( \text{blond hair at crime scene} \mid \text{someone else is source})
\]

- What is the relevant population?
Imagine you are a forensic hair comparison expert...

- The hair at the crime scene is blond
- The suspect has blond hair

\[
p(\text{blond hair at crime scene} \mid \text{suspect is source})
\]

\[
p(\text{blond hair at crime scene} \mid \text{someone else is source})
\]

- What is the relevant population?
  - Stockholm
  - Beijing
- You need to use a sample representative of the relevant population
A likelihood ratio is the answer to a **specific question** defined by the prosecution and the defence hypotheses.

- The **defence** hypothesis specifies the **relevant population**.

- The forensic scientist must make explicit the specific question they answered so that the trier of fact can:
  - understand the question
  - consider whether the question is an appropriate question
  - understand the answer
Prosecutor’s Fallacy

• Forensic Scientist:
  – One would be one thousand times more likely to obtain the acoustic properties of the voice on the intercepted telephone call had it been produced by the accused than if it had been produced by some other speaker from the relevant population.

• Prosecutor:
  – So, to simplify for the benefit of the jury if I may, what you are saying Doctor is that it is a thousand times more likely that the voice on the telephone intercept is that of the accused than that of any other speaker from the relevant population.
Prosecutor’s Fallacy

- Forensic Scientist:
  - One would be one thousand times more likely to obtain the acoustic properties of the voice on the intercepted telephone call had it been produced by the accused than if it had been produced by some other speaker from the relevant population.

\[
p\left( E \mid H_{\text{prosecution}} \right) \frac{p(E \mid H_{\text{prosecution}})}{p(E \mid H_{\text{defence}})}
\]

- Prosecutor:
  - So, to simplify for the benefit of the jury if I may, what you are saying Doctor is that it is a thousand times more likely that the voice on the telephone intercept is that of the accused than that of any other speaker from the relevant population.

\[
\frac{p(H_{\text{prosecution}} \mid E)}{p(H_{\text{defence}} \mid E)}
\]
Defence Attorney’s Fallacy

- Forensic Scientist:
  – One would be one thousand times more likely to obtain the measured properties of the partial latent finger mark had it been produced by the finger of the accused than if it had been produced by a finger of some other person.

- Defence Attorney:
  – So, given that there are approximately a million people in the region and assuming initially that any one of them could have left the finger mark, we begin with prior odds of one over one million, and the evidence which has been presented has resulted in posterior odds of one over one thousand. One over one thousand is a small number. Since it is one thousand times more likely that the finger mark was left by someone other than my client than that it was left by my client, I submit that this evidence fails to prove that my client left the finger mark and as such it should not be taken into consideration by the jury.

\[
prior \text{ odds} \times \text{ likelihood ratio} = posterior \text{ odds}
\]

\[
\frac{1}{1,000,000} \times 1,000 = \frac{1}{1,000}
\]
Large Number Fallacy

• Forensic Scientist:
  – One would be one billion times more likely to obtain the measured properties of the DNA sample from the crime scene had it come from the accused than had it come from some other person in the country.

• Trier of Fact:
  – One billion is a very large number. The DNA sample must have come from the accused. I can ignore other evidence which suggests that it did not come from him.
Discrete data: bar graph

\[
p\left( \frac{4 \text{ legs} \mid \text{cow}}{4 \text{ legs} \mid \text{not a cow}} \right) = 2
\]
Continuous data: histograms $\rightarrow$ probability density functions (PDFs)
Continuous data: histograms $\rightarrow$ probability density functions (PDFs)

(a) (b)

(c) (d)

$\mu = 100$

$\sigma = 30$
probability density

population model

suspect model
$LR = 0.021 / 0.005 = 4.02$
LRs are the past

- 1906 retrial of Alfred Dreyfus

- Jean-Gaston Darboux, Paul Émile Appell, Jules Henri Poincaré
LRs are the present

- Adopted as standard for evaluation of DNA evidence in mid 1990’s
LRs are the future

- Increasing support for the position that the likelihood-ratio framework is the logically correct framework for the evaluation of forensic evidence
LRs are the future

- Increasing support for the position that the likelihood-ratio framework is the logically correct framework for the evaluation of forensic evidence

- 2011 Position Statement
  - 2010 England & Wales Court of Appeal ruling *R v T*
  - 31 signatories
  - endorsed by ENFSI, 58 laboratories in 33 countries

LRs are the future

• Increasing support for the position that the likelihood-ratio framework is the logically correct framework for the evaluation of forensic evidence

• 2012 Response
  – Draft Australian Standard for interpretation of forensic evidence
  – 22 endorsers
  – 5 others submitted separate responses

LRs are the future

- Increasing support for the position that the likelihood-ratio framework is the logically correct framework for the evaluation of forensic evidence

- 2012 NIST/NIJ Report on latent fingerprint analysis
  - working group of 34 experts
  - others contributed

The New Paradigm for the Evaluation of Forensic Evidence

- **Use of the likelihood-ratio framework for the evaluation of forensic evidence**
  - logically correct
  - adopted for DNA in the mid 1990s

- **Use of quantitative measurements, databases representative of the relevant population, and statistical models**
  - transparent and replicable
  - relatively robust to cognitive bias

- **Empirical testing of validity and reliability under conditions reflecting those of the case at trial using test data drawn from the relevant population**
  - *Daubert*
  - 2009 National Research Council Report
Thank You

Geoffrey Stewart Morrison

http://forensic-evaluation.net/