

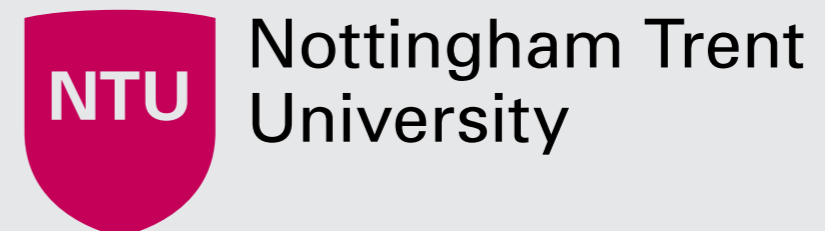
# Evaluative opinions for fingermark-fingerprint comparison

*Geoffrey Stewart Morrison*

Forensic Data Science Laboratory  
Aston University

*William Morris*

School of Science & Technology  
Nottingham Trent University



# Acknowledgment

- This research was supported by Research England's Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2024.

# Disclaimer

- All opinions expressed are those of the presenters and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the presenters are associated.

# Slides

- [http://geoff-morrison.net/#FCN\\_fingerprint\\_2023\\_03](http://geoff-morrison.net/#FCN_fingerprint_2023_03)

# Contents

- Who are we?
- What is a likelihood ratio?
- What is calibration?
- Regulator's appendix on evaluative opinions
- Quantitative-measurement & statistical-model approach
- Calibration of human perception & judgement
- Comparison and fusion of systems
- Verbal scales
- Conclusion

**Who are we?**

# Who are we?

- **Geoffrey Stewart Morrison**

- Director, Forensic Data Science Laboratory, Aston University

- forensic inference and statistics

- calibration and validation of likelihood-ratio systems

- forensic voice comparison, fired cartridge case comparison

- cell-site analysis, fingermark-fingerprint comparison

- Chair, Forensic Science Committee, British Standards Institution

- contributor to ISO21043 Forensic Science, particularly Part 4 Interpretation

- <http://geoff-morrison.net/>

- <http://forensic-data-science.net/>

# Who are we?

- **William Morris**

- Senior Lecturer in Forensic Science, Nottingham-Trent University
  - fingermark detection and fingerprint identification
  - footwear marks detection and identification
  - forensic inference covering fingerprint evidence
- Forensic Trainer, College of Policing (2014–2017)
- Entered on the National Register of Fingerprint Experts (2003–2017)
- <https://www.ntu.ac.uk/staff-profiles/science-technology/william-morris>

# Who are we?

- We are developing an R&D project on calculation, calibration, and validation of likelihood ratios for fingerprint-fingermark comparison
- Includes collection of a database of fingerprint and fingermark images
- We will be applying for funding
- We will be seeking collaboration with partner organization



**What is a likelihood ratio?**

# What is a likelihood ratio?

- Likelihood of obtaining the properties of the questioned-source fingermark and the known-source fingerprint if they both came from the same finger (**similarity**).

divided by

- Likelihood of obtaining the properties of the questioned-source fingermark and the known-source fingerprint if they came from two different fingers in the relevant population (**typicality**).

# What is Calibration?

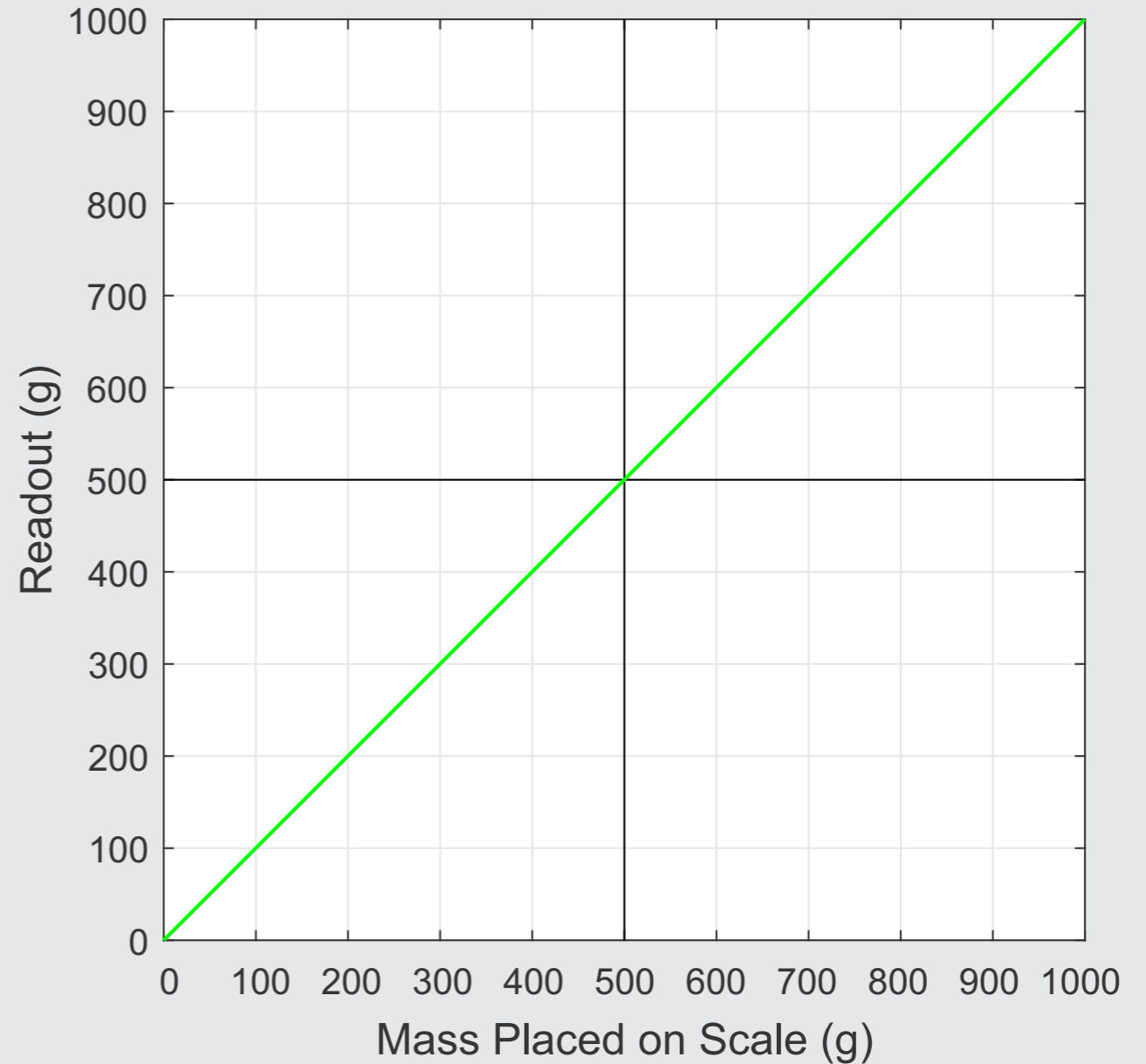
# What is Calibration?

- What is a well-calibrated set of scales?
- A set of scales for which:
  - The mass stated in the readout is the same as the mass placed on the scale



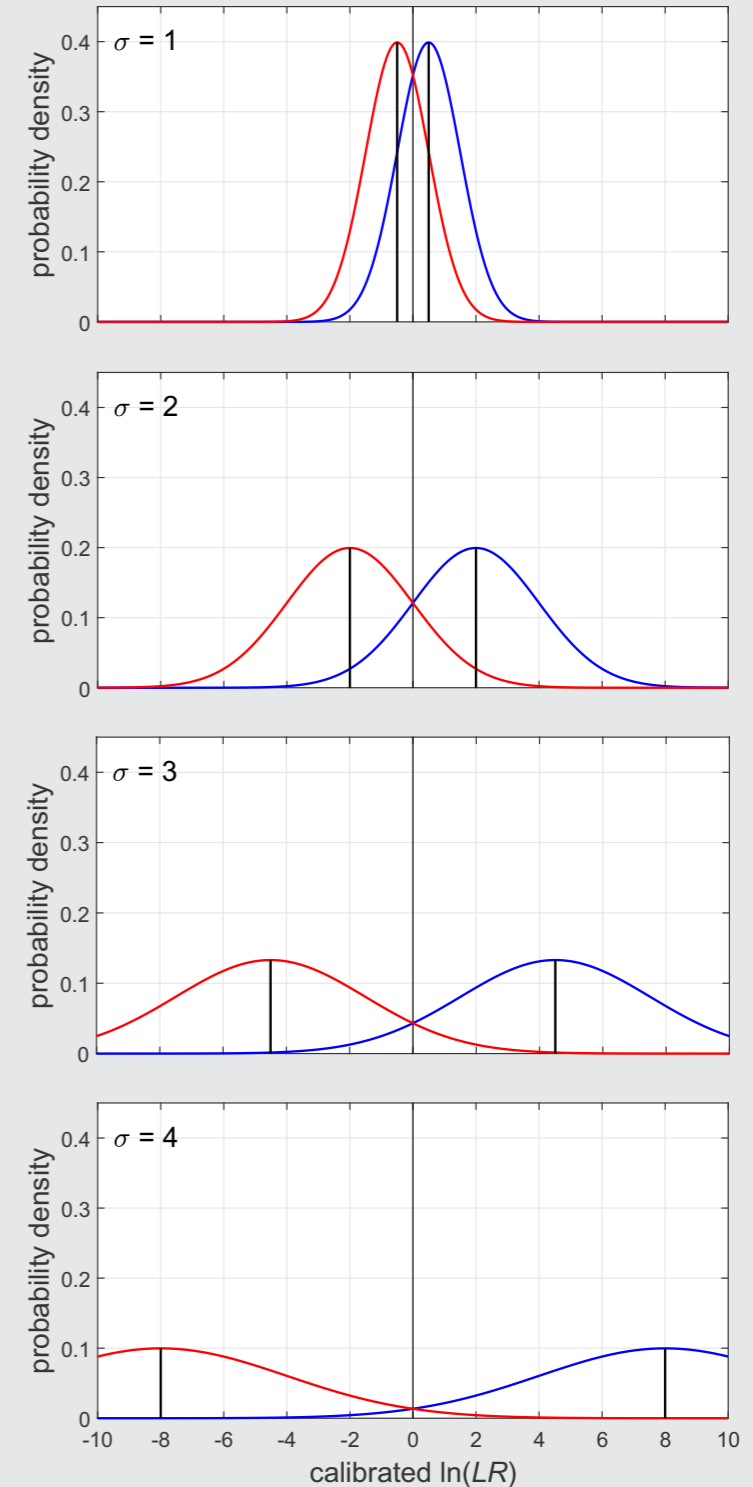
# What is Calibration?

- Calibration is the process of adjusting the set of scales so that its output is well calibrated.



# What is Calibration?

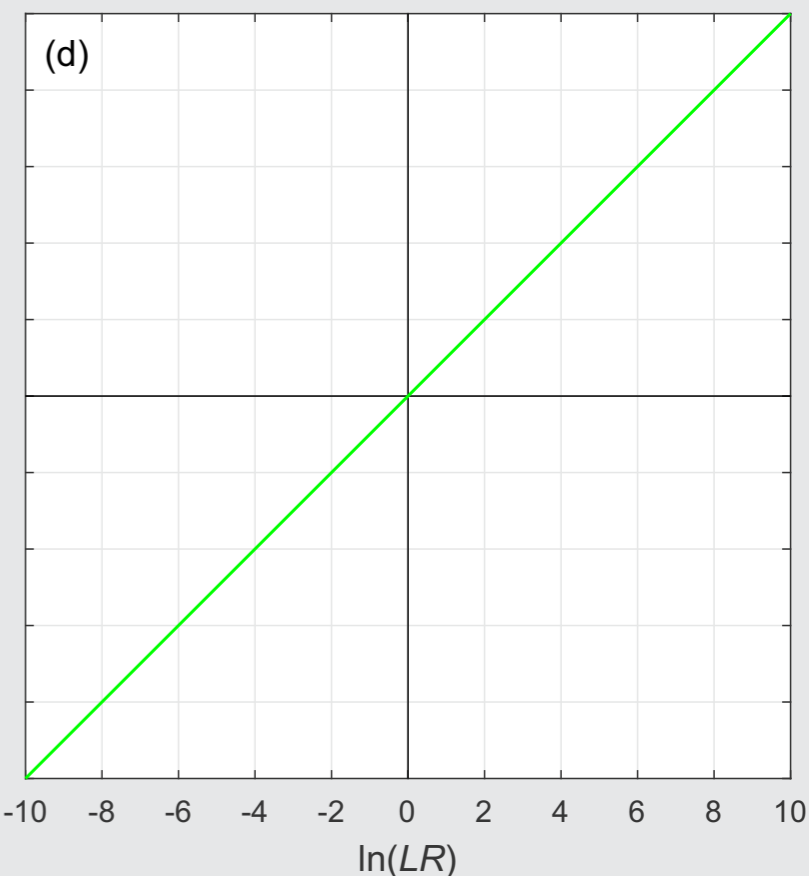
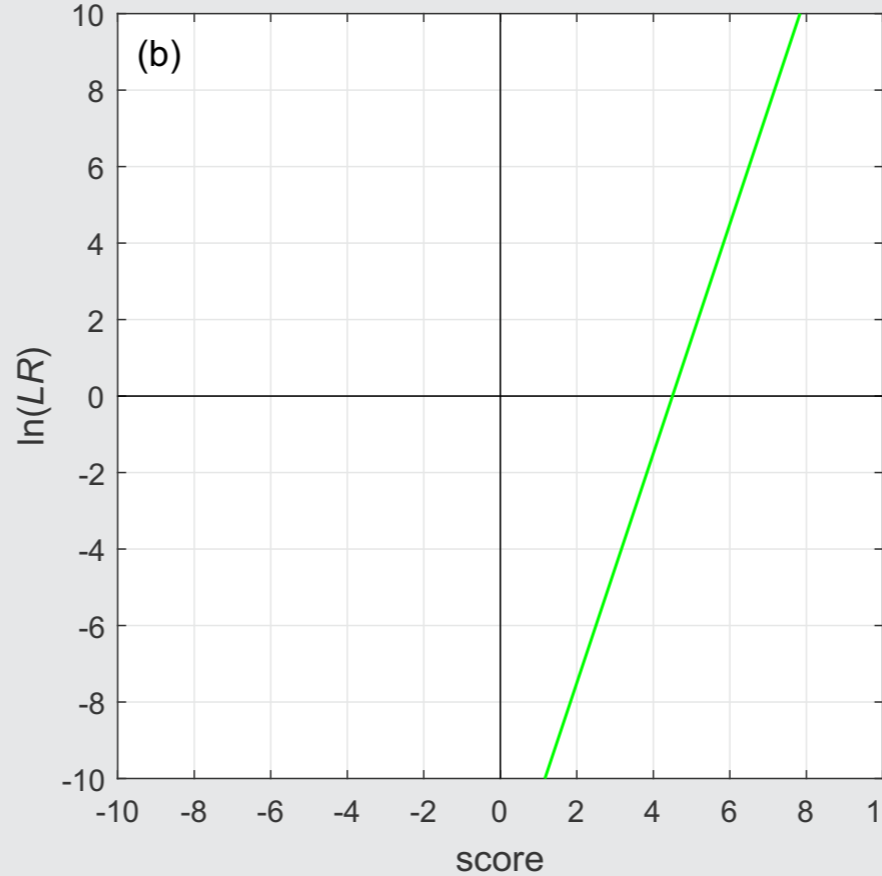
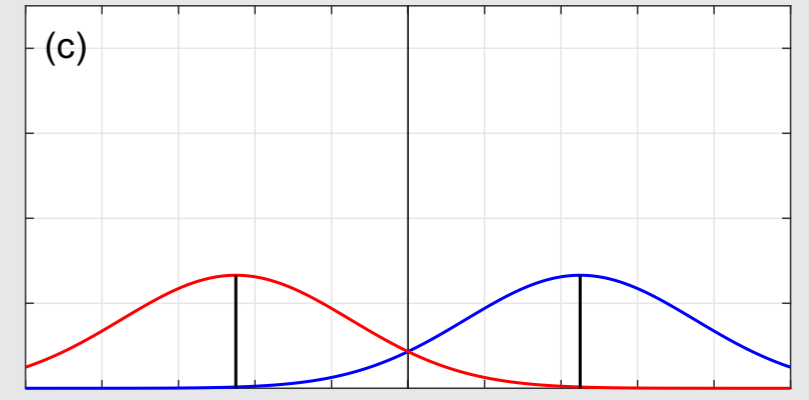
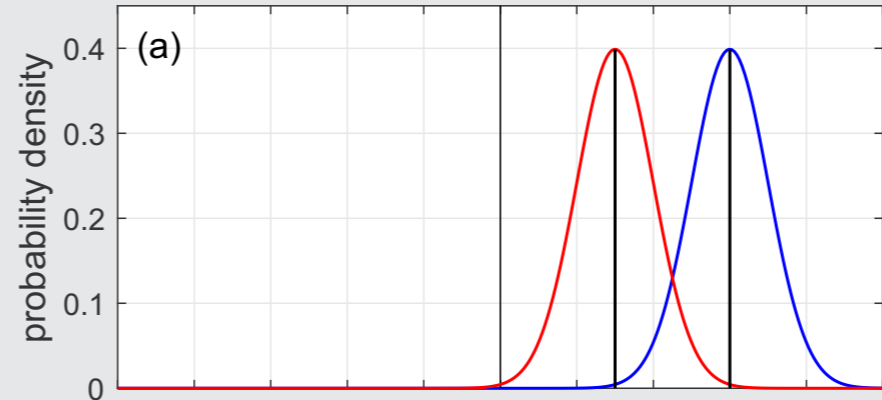
- What is a well-calibrated likelihood-ratio system?
- A system for which:
  - The likelihood ratio of the likelihood ratio is the likelihood ratio



# What is Calibration?

- Calibration is the process of adjusting the system so that its output is well calibrated, i.e., so that:

$$LR = \frac{f(LR | H_s)}{f(LR | H_d)}$$



# **Regulator's Appendix on Evaluative Opinions**



# Regulator's Appendix on Evaluative Opinions

- Forensic Science Regulator (2021). **Codes of practice and conduct: Development of evaluative opinions** (FSR-C-118 Issue 1). Birmingham, UK: Forensic Science Regulator. <https://www.gov.uk/government/publications/development-of-evaluative-opinions>
- Will not be included in the first version of the Statutory Codes
- But will be revised for inclusion in a later version

# Regulator's Appendix on Evaluative Opinions



“probabilities have been assigned on the basis of a data set of sufficient relevance, quality and size”



“probabilities have been assigned on the basis of structured data set(s) which are limited in their relevance, quality and/or size but are available for inspection by another expert”



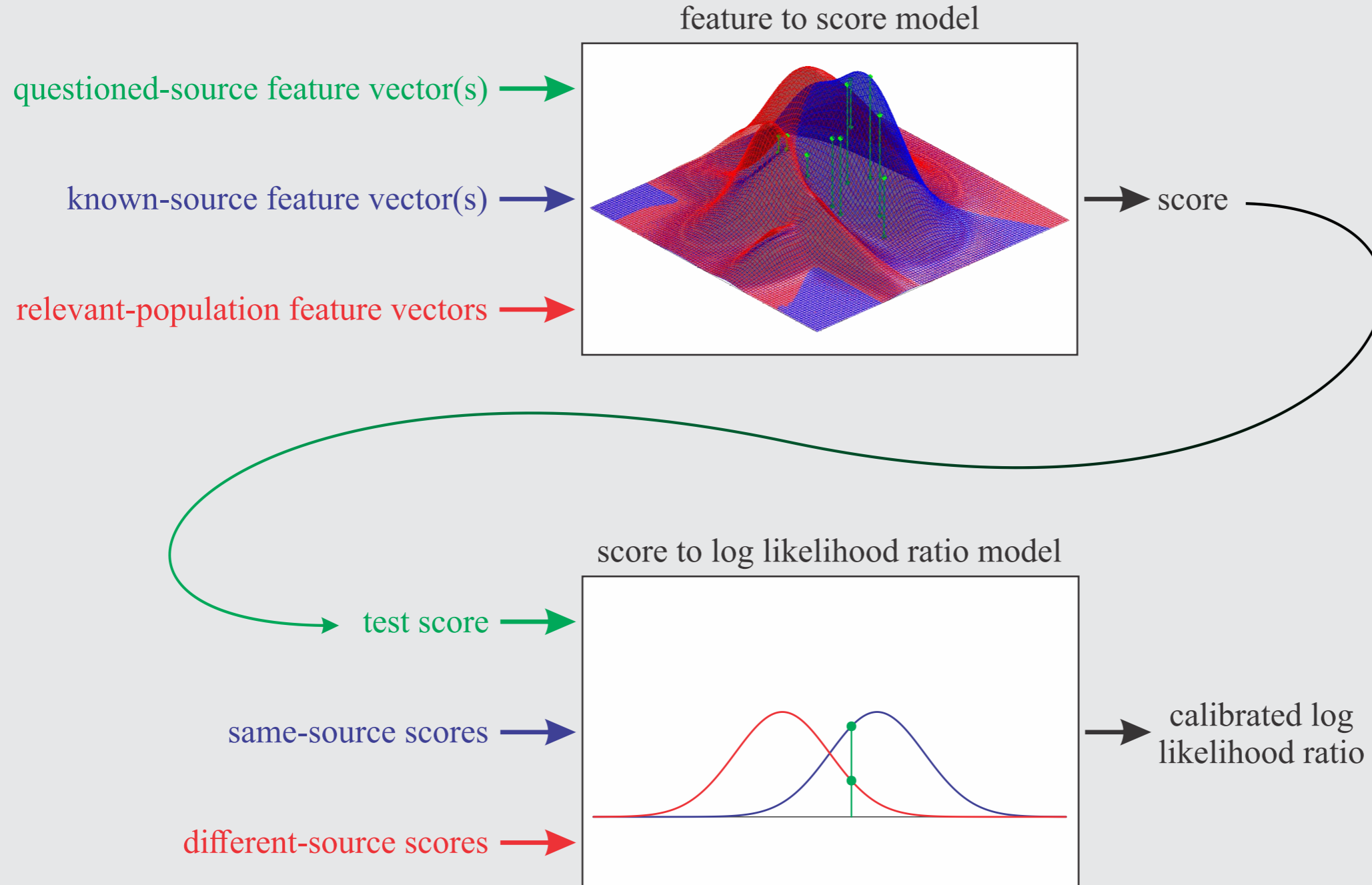
“probabilities have been assigned on the basis of unstructured observations from experience, which are not available for inspection by another expert”

# Regulator's Appendix on Evaluative Opinions



“probabilities have been assigned on the basis of a data set of sufficient relevance, quality and size”

# Regulator's Appendix on Evaluative Opinions



# Regulator's Appendix on Evaluative Opinions



- Important condition:
  - The data used for training the calibration model must:
    - represent the relevant population in the case
      - including there being enough data
    - reflect the conditions of the questioned-source and known-source items in the case
      - including any mismatches in conditions
  - If not, the system will be miscalibrated

# Regulator's Appendix on Evaluative Opinions



- Important condition:
  - The first model must output scores which are **uncalibrated log likelihood ratios**.  
They must take account of both:
    - the **similarity** between the questioned-source and the known-source items
    - their **typicality** with respect to the relevant population
  - Similarity-only scores cannot be used

# Regulator's Appendix on Evaluative Opinions



“probabilities have been assigned on the basis of structured data set(s) which are limited in their relevance, quality and/or size but are available for inspection by another expert”

# Regulator's Appendix on Evaluative Opinions



- “The validity of a structured data set (including any local data set) from previous casework, a ‘knowledge base’ ..., shall be calibrated regularly by conducting studies using ground truth data as described by Evett [22].”
- “Knowledge Base  
A structured database of information and assigned probabilities, ordered according to casework conditions. The knowledge base is calibrated through regular review of its content through experimentation under controlled conditions [22].”



# Regulator's Appendix on Evaluative Opinions



- “Calibration involves regular review of sections of the content by conducting experimentation using ground truth data under controlled conditions and comparing to relevant sections of the knowledge base.
- Such ground truth experimentation enables the knowledge base to be updated and expert opinions to be checked against a snapshot of known-source data.”

# Regulator's Appendix on Evaluative Opinions



[22] Evett I.W. (2015). **The logical foundations of forensic science: Towards reliable knowledge.** *Philosophical Transactions of the Royal Society B*, 370, article 20140263.

<http://dx.doi.org/10.1098/rstb.2014.0263>

- This is a high-level review paper.
- It does not provide detail about how to:
  - calibrate a knowledge base
  - use a knowledge base to assign probabilities in the context of a case

# Regulator's Appendix on Evaluative Opinions

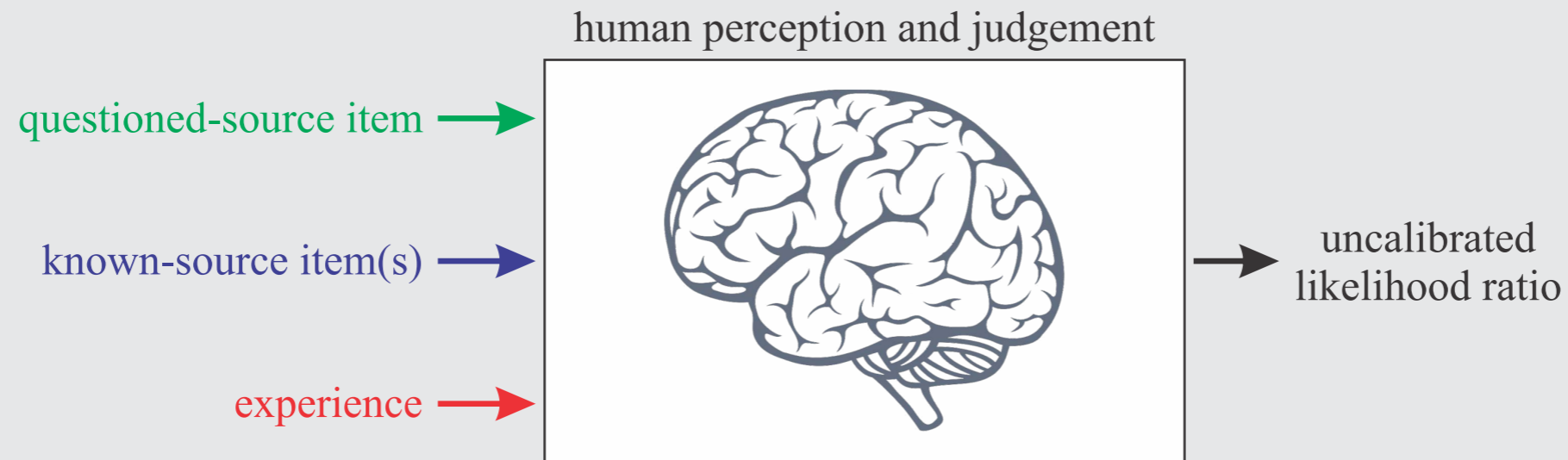


“probabilities have been assigned on the basis of unstructured observations from experience, which are not available for inspection by another expert”

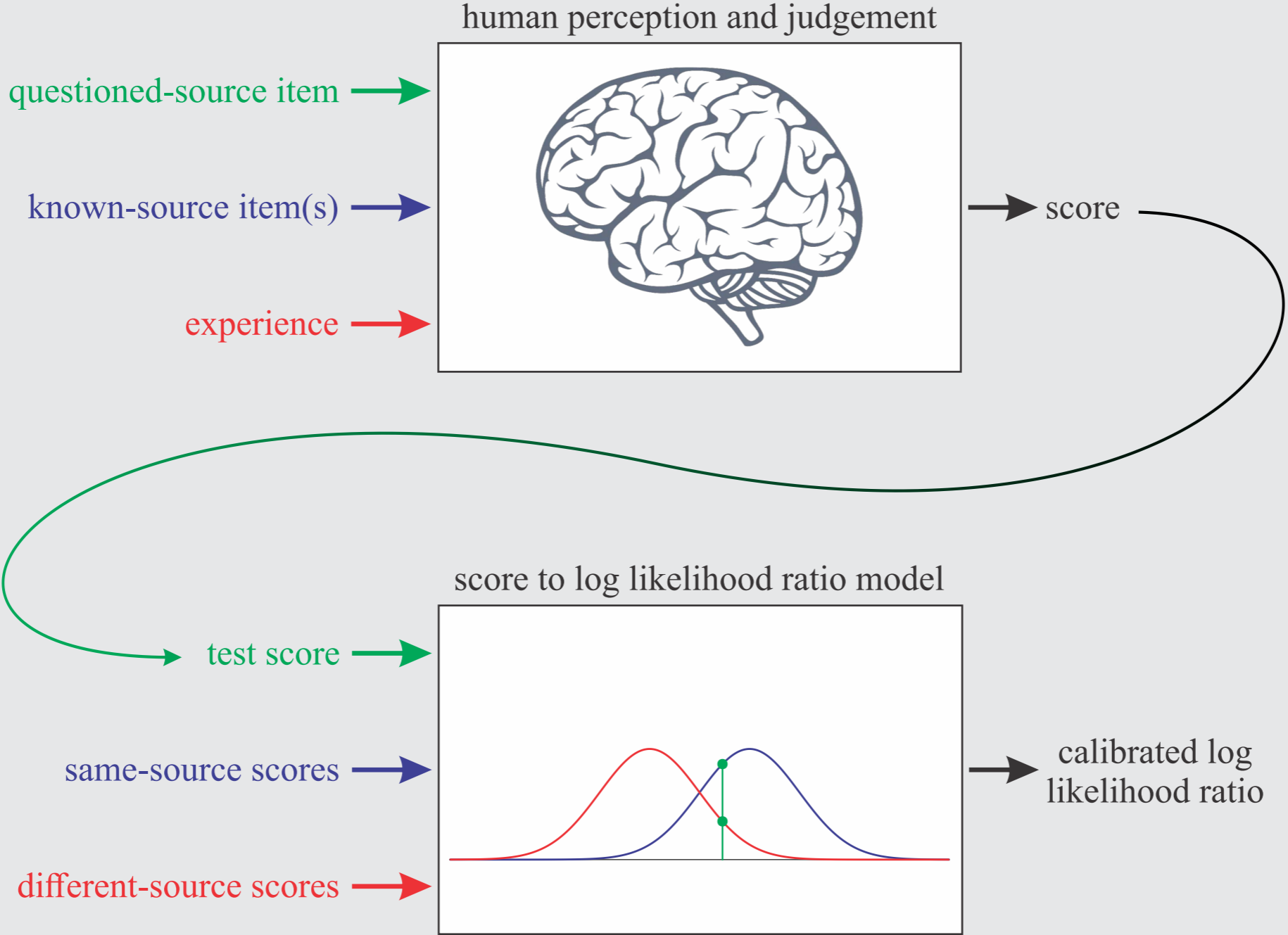
# Regulator's Appendix on Evaluative Opinions



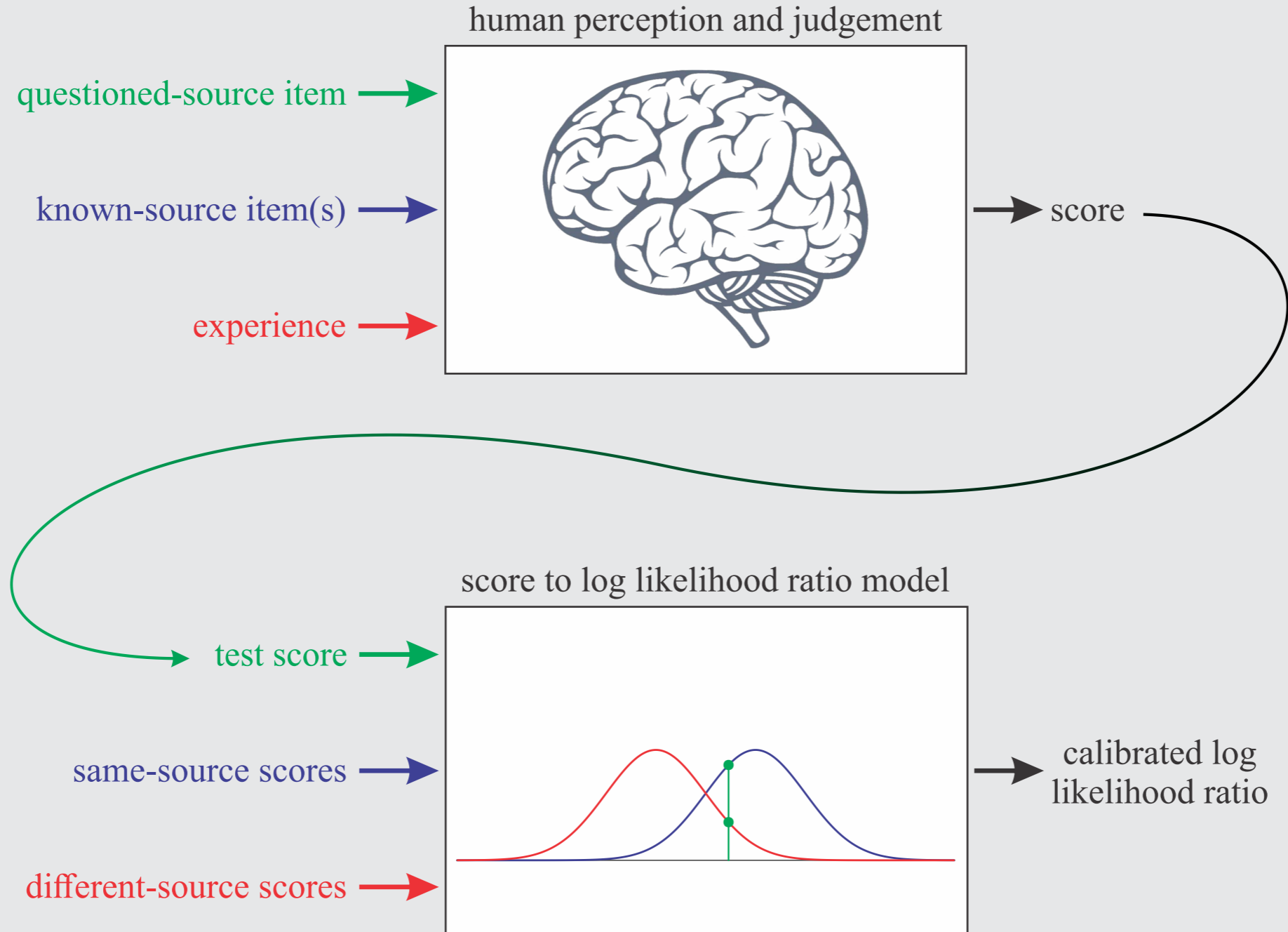
- “In instances where an expert is unable to demonstrate any ... calibration of their expertise, the commissioning party and the court shall be made aware that their opinion is uncalibrated.”



# Regulator's Appendix on Evaluative Opinions



# Regulator's Appendix on Evaluative Opinions



# Regulator's Appendix on Evaluative Opinions

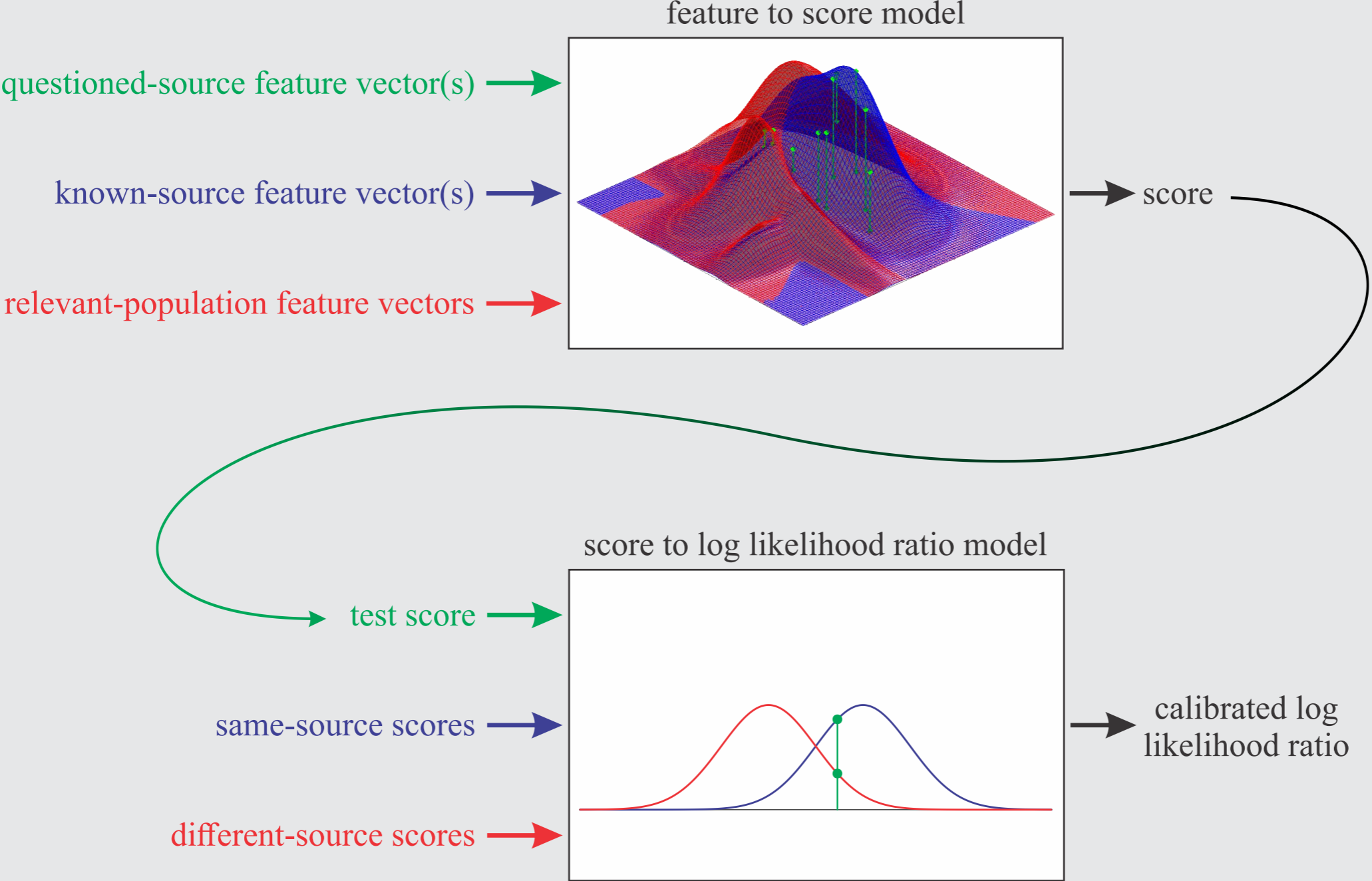


- Important condition:
  - The data used for training the calibration model must:
    - represent the relevant population in the case
      - including there being enough data
    - reflect the conditions of the questioned-source and known-source items in the case
      - including any mismatches in conditions
  - If not, the system will be miscalibrated

# **Quantitative-Measurement & Statistical-Model Approach**

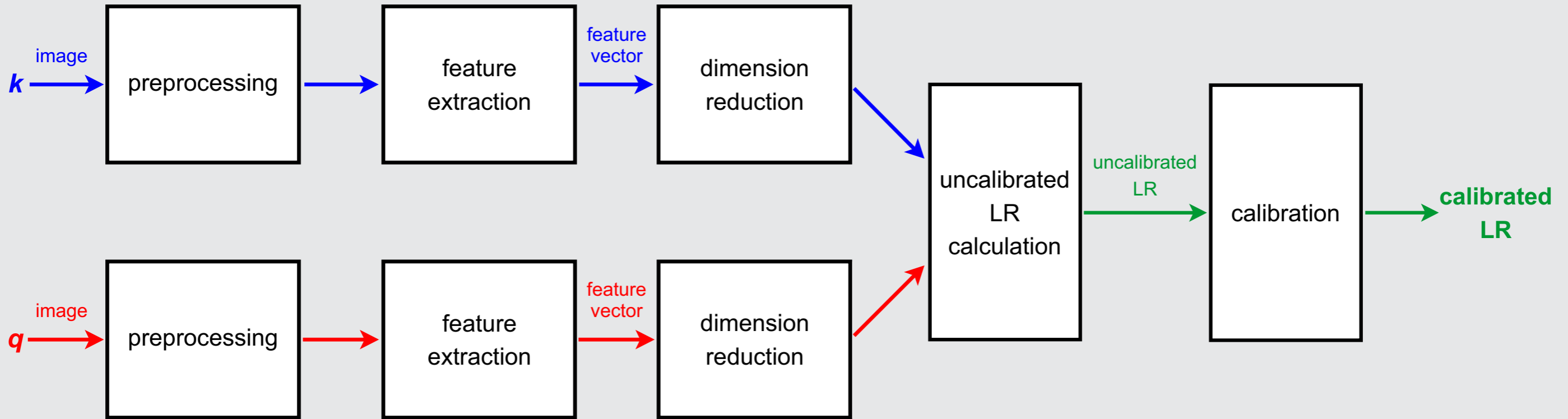


# Quantitative-Measurement & Statistical-Model Approach



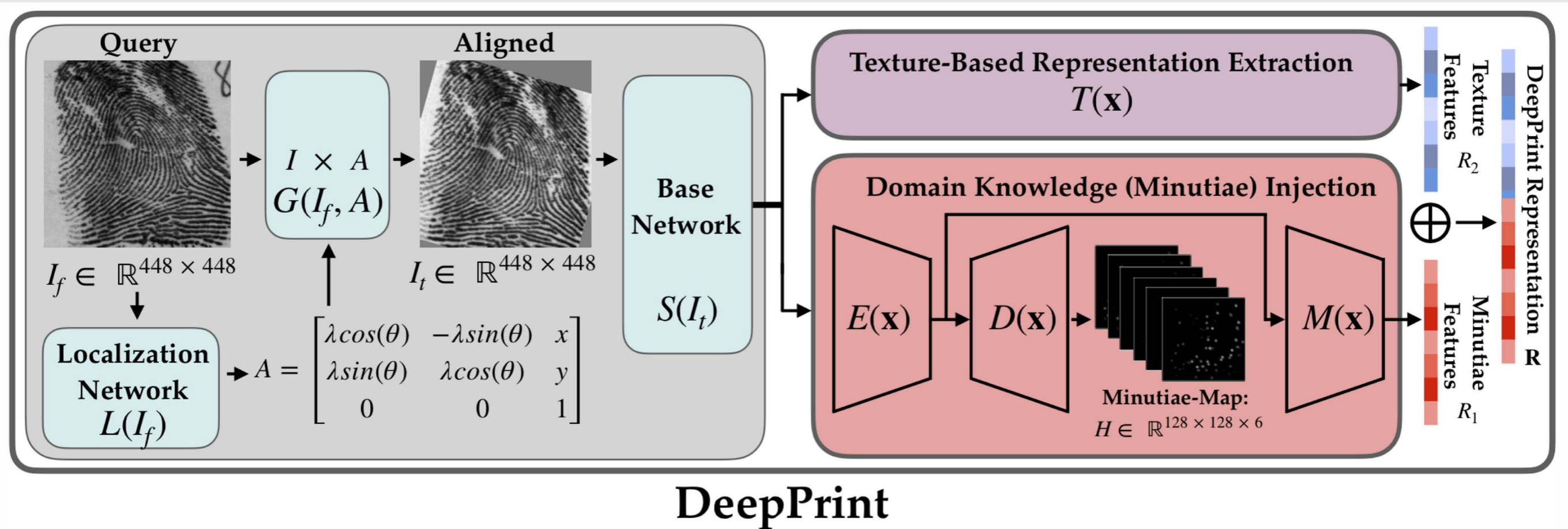
# Quantitative-Measurement & Statistical-Model Approach

- Pipeline based on state-of-the-art forensic voice comparison



# Quantitative-Measurement & Statistical-Model Approach

- Feature-extraction using a Deep-Neural-Network (DNN) embedding



- Engelsma J.J., Cao K., Jain A.K. (2021). Learning a fixed-length fingerprint representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 1981–1997. <https://doi.org/10.1109/TPAMI.2019.2961349>

# Quantitative-Measurement & Statistical-Model Approach

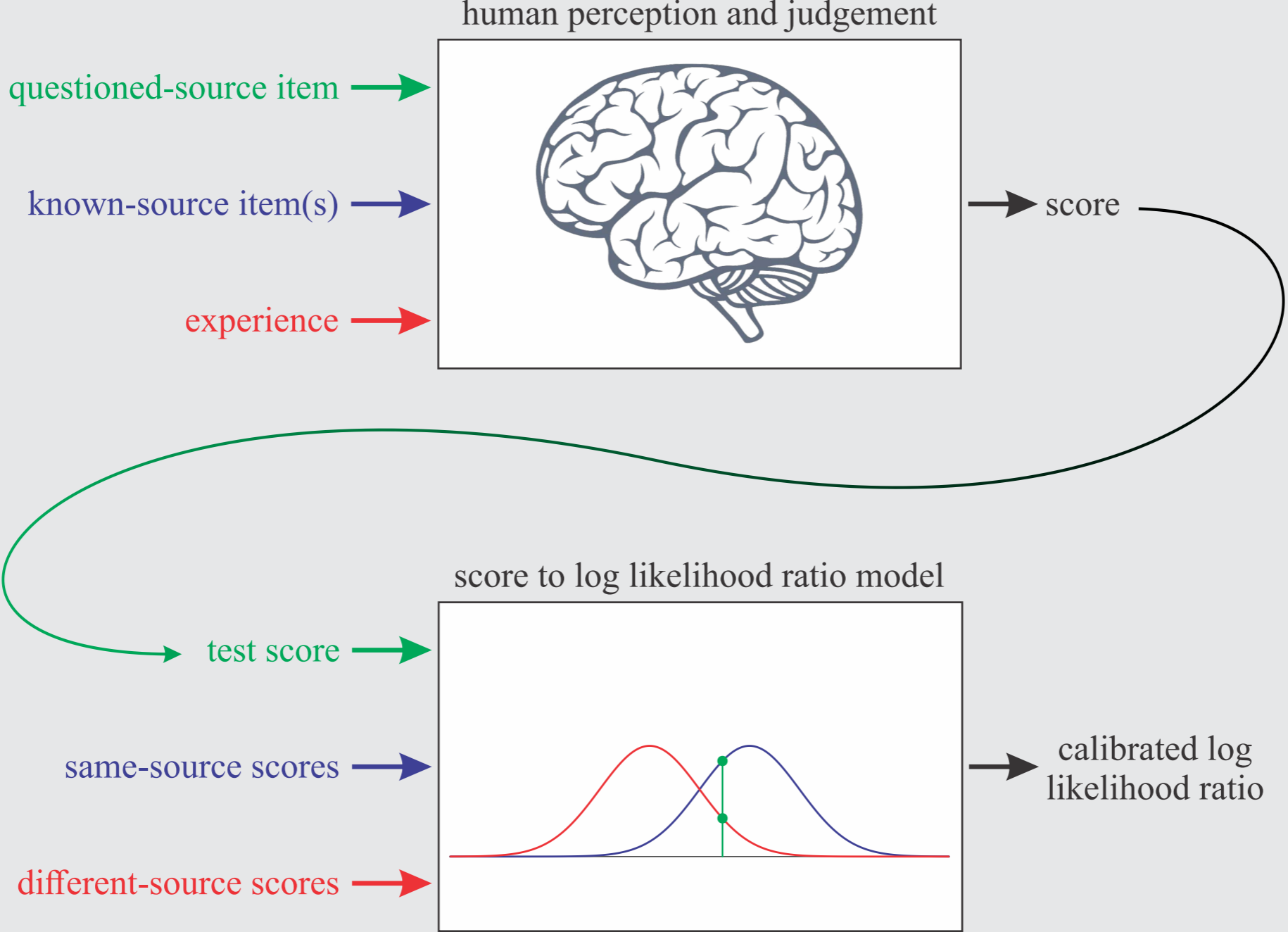
- The quantitative-measurement & statistical-model system is a tool used by a human expert.
- Human expertise is required:
  - Selection of appropriate data (fingermark and fingerprint images) for calibrating and validating the system under conditions reflecting those of the fingermark and fingerprint from the case.
  - Otherwise: Garbage in, garbage out.
  - Communicating the meaning of the output.

# Quantitative-Measurement & Statistical-Model Approach

- For training, calibration, and validation we need data:
  - Fingermarks and fingerprints from a large number of donors
  - Fingermarks representing conditions commonly encounter in casework
  - In each condition, a large number of fingermarks from each donor
- Model between-source and within-source variability in casework-relevant conditions

# Calibration of Human Perception & Judgement

# Calibration of Human-Perception & Judgement



# Calibration of Human-Perception & Judgement

- Human examiner output:
  - identification - inconclusive - exclusion
  - five/seven/nine-level ordinal scale
    - e.g., draft ASB 013 Standard for Friction Ridge Examination Conclusions
  - subjectively assigned numerical likelihood ratio
    - e.g., ENFSI Guideline for Evaluative Reporting in Forensic Science



# Calibration of Human-Perception & Judgement

$$\frac{p(\text{“identification”} \mid H_{\text{same-source}})}{p(\text{“identification”} \mid H_{\text{different-source}})}$$

# Calibration of Human-Perception & Judgement

- Morrison G.S. (2022). **A method to convert traditional fingerprint ACE / ACE-V outputs (“identification”, “inconclusive”, “exclusion”) to Bayes factors.** Unpublished Manuscript. [https://geoff-morrison.net/#ID\\_IN\\_EX\\_to\\_BF](https://geoff-morrison.net/#ID_IN_EX_to_BF)
- Busey T., Coon M. (2023). **Not all identification conclusions are equal: Quantifying the strength of fingerprint decisions.** *Forensic Science International*, 343, 111543. <https://doi.org/10.1016/j.forsciint.2022.111543>

# Calibration of Human-Perception & Judgement

- For calibration and validation we need data:
  - Fingermarks and fingerprints from a large number of donors
  - Fingermarks representing conditions commonly encounter in casework
  - In each condition, a large number of fingermarks from each donor

# Calibration of Human-Perception & Judgement

- The system that is actually to be used for the case is the system that must be calibrated and validated:
  - Individual examiner
  - Pair of examiners conducting ACE-V
- The system must be calibrated and validated under conditions reflecting each of the cases to which it is to be applied.

# Calibration of Human-Perception & Judgement

- In each condition, the system must provide responses to a large number of fingermark-fingerprint pairs.
- The size of the calibrated likelihood-ratio value (how far it is from 1) will be constrained by the number of fingermark-fingerprint pairs that the system provides responses for.
- We can use cross-validation, so the responses provided for calibration can also be used for validation.

# Comparison and Fusion of Systems

# Comparison and Fusion of Systems

- Have quantitative-measurement & statistical-model system and human perception & judgement system each provide responses to the same fingerprint-fingerprint pairs.
- Build a model that fuses the responses from the two systems.
  - the process is similar to calibration, but takes parallel input from two or more systems
- Compare the performance of the individual systems with each other and with the fused system.
  - log likelihood ratio cost ( $C_{lr}$ )
  - Tippett plots

# Verbal Scales



# Verbal Scales

- Association of Forensic Science Providers (2009). **Standards for the formulation of evaluative forensic science expert opinion.** *Science & Justice*, 49, 161–164.  
<http://dx.doi.org/10.1016/j.scijus.2009.07.004>
- Willis S.M., McKenna L., McDermott S., O'Donnell G., Barrett A., Rasmusson A., Nordgaard A., Berger C.E.H., Sjerps M.J., Lucena-Molina J.J., Zadora G., Aitken C.G.G., Lunt L., Champod C., Biedermann A., Hicks T.N., Taroni F. (2015). **ENFSI Guideline for Evaluative Reporting in Forensic Science.** [http://enfsi.eu/wp-content/uploads/2016/09/m1\\_guideline.pdf](http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf)
- AAFS Standards Board (2021). **ASB 013 Standard for Friction Ridge Examination Conclusions [draft].** [https://www.aafs.org/sites/default/files/media/documents/013\\_Std\\_Ballot02.pdf](https://www.aafs.org/sites/default/files/media/documents/013_Std_Ballot02.pdf)

# Verbal Scales

- Expression on verbal scales have no intrinsic meaning
  - moderately strong
- Calibrated numerical likelihood ratios do have intrinsic meaning
  - propositions must be clearly stated

# Conclusion

# Conclusion

- Collect relevant data
  - fingerprints and fingermark that reflect casework conditions
  - sufficient quantity to capture between-source and within-source variability
- Develop quantitative-measurement & statistical-model systems that are calibrated and validated under casework conditions
- Calibrate and validated human-perception & judgement under casework conditions

*Thank You*