Follow the trace

Evaluating physical and digital forensic findings given activity level propositions using Bayesian networks

Vink, M.

Publication date
2025

APPENDICES

A

SUPPLEMENTARY MATERIAL

Part II

Chapter 3
Fig. 3.1
Template network
PDF

Chapter 3
Fig 3.1
Template network
Hugin file

Part III

Chapter 4
Fig. 4.3
Case model
PDF

Chapter 5
Fig. 5.1
Template network
PDF

Chapters 4
and 5
Bayesian networks
Hugin files

Figure A.1: QR codes providing access to the Supplementary material that accompanies the published manuscripts. Note that the Bayesian network from Chapter 8 are not included, as that manuscript is still under review.

B

CHAPTER 2: FORENSIC EXAMPLES OF IDIOMS

This appendix provides concrete examples of the forensic idioms presented in Chapter 2. We only provide examples for the idioms where the variations lie in the node names (and not in the states, as in the case of the extended-likelihood ratio idiom).

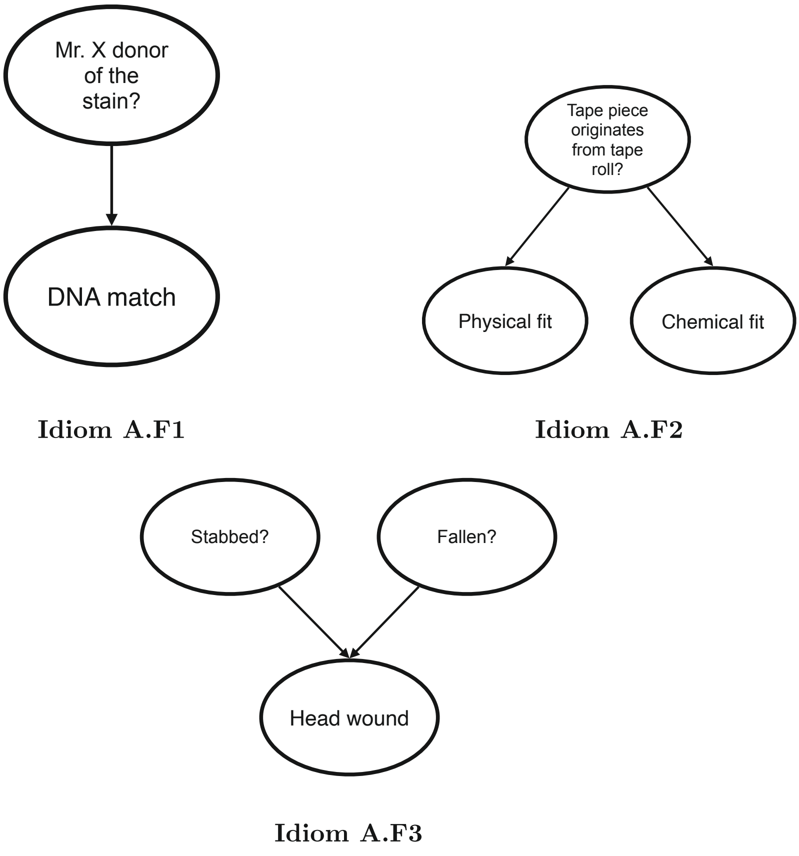
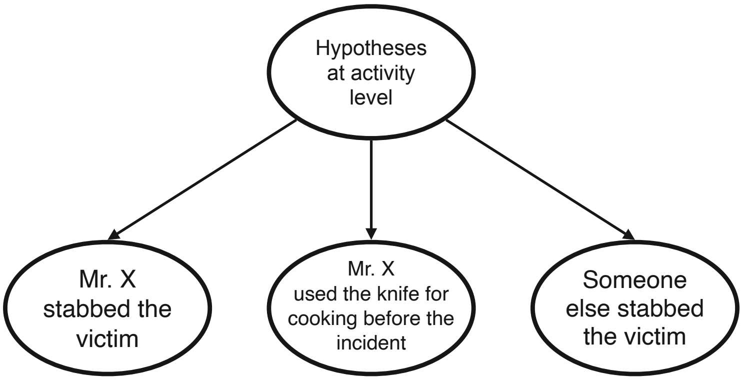
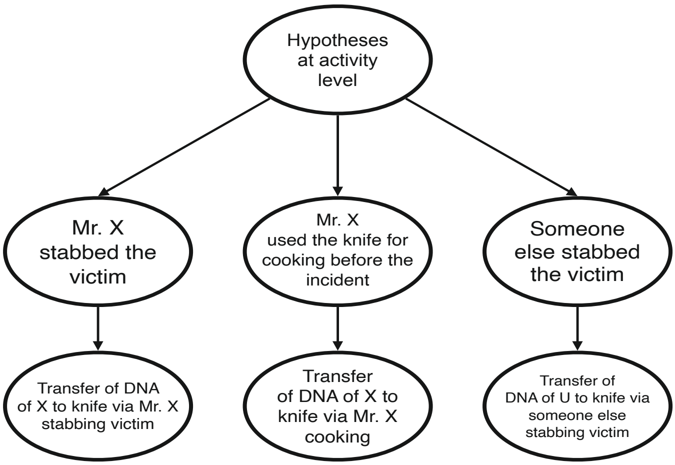


Figure B.1: Examples of the hypothesis-evidence idiom (upper left), the common cause idiom (upper right) and the common effect idiom (bottom).



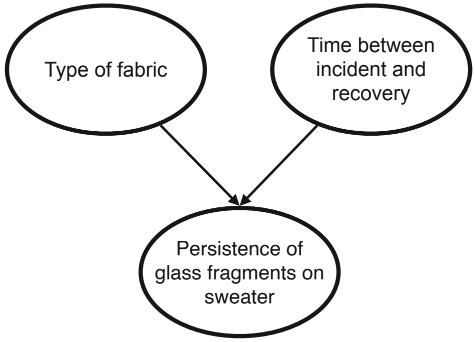
Idiom B.F1



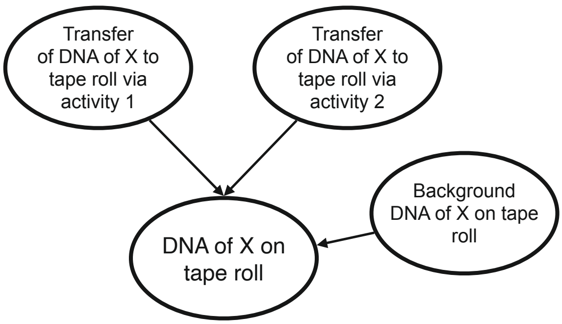
Idiom B.F2

Figure B.2: Examples of the hypothesis-to-activity idiom (above) and the hypothesis-to-transfer idiom (below).

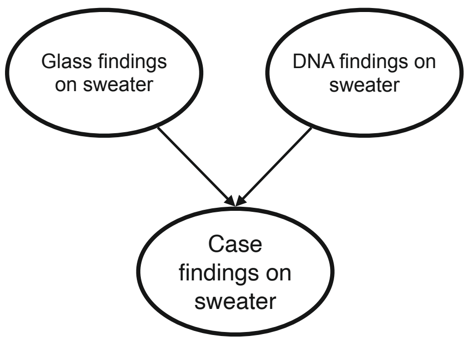
B



Idiom C.F2



Idiom C.F3



Idiom C.F4

Figure B.3: Examples of the influential-factors idiom (above), the trace-accumulation idiom (middle) and the case-findings idiom (below).

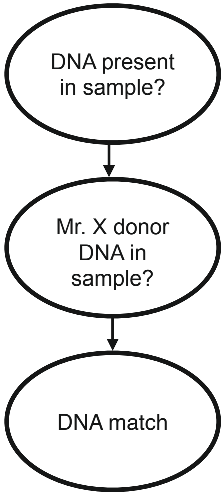


Figure B.4: Forensic example of **D.F3**, the hypothesis-precondition idiom.

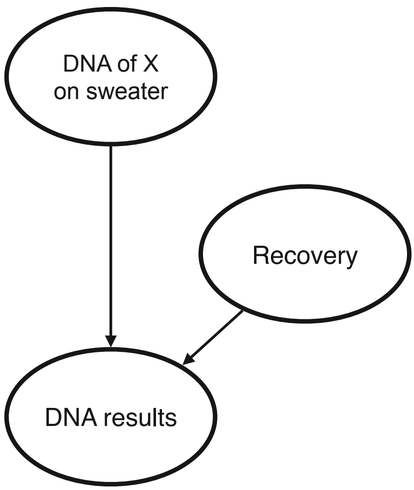


Figure B.5: Example of **E.F1**, the evidence-uncertainty idiom.

C

CHAPTER 6: MORE ON THE VALIDATION OF THE LR METHOD

Calibration

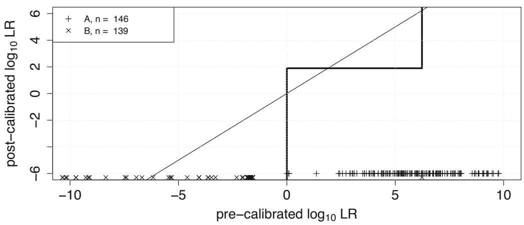
Pool Adjacent Violators (PAV) plots

Ideally, the output LRs are well-calibrated (i.e., consistent) and the evidential value increases with increasing LR. The Pool Adjacent Violators (PAV) algorithm enforces the LR method to adhere to this trait (Brümmer & Du Preez, 2006; Fawcett & Niculescu-Mizil, 2007; Vergeer *et al.*, 2021). In Chapter 6 we used Brummer & Preez' algorithm (2006) to calibrate the method-LRs. PAV transforms the method-LRs to well-calibrated LRs resulting in a monotonically increasing function. The algorithm first sorts all LRs in ascending order. Then all LRs are assigned a posterior probability of 1 when H_1 is true and 0 when H_2 is true. PAV searches this list of LRs for violators of the monotonicity principle. Violators are LRs whose probabilities decrease with respect to its adjacent - for example, a probability of 0 after a probability of 1. If the algorithm finds such violators, it pools them and replaces their probability estimates with the average of the pool's probabilities. This may result in new violators and the algorithm continues this iterative process until no violators appear. The result is a sequence of LRs and their non-decreasing posterior probabilities meaning increasing evidential value with LR-value. Bayes' rule is used to recover the calibrated LRs from the posterior probabilities. Violation of the monotonicity principle only appears if an LR occurs under both H_1 and H_2. Hence the transformation works on the overlap region of the LR distributions in Fig. 6.8 from Chapter 6 and all method-LRs outside this region are transformed to $-\infty$ or ∞. By plotting the PAV LRs against the method-LRs along the line $y = x$ one can study the calibration of the output LRs. Ideally the LRs lie along the line $y = x$, as it shows that the output LRs were well-calibrated. The smaller the deviation, the better the calibration. Vergeer *et al.* (2021) introduced a new quantitative measure for the cumulative deviation from the $y = x$ line: devPAV. Large devPAV-values indicate bad calibration and small devPAV-values represent good calibration.

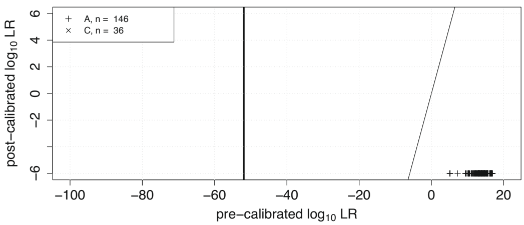
Fig. C.1 illustrates the PAV transform of the method LRs for the hypotheses set A,B (a), set A,C (b) and set B,C (c). The diagonal line represents the line $y = x$ and the one dimensional scatter plots represent the pre-calibrated method-LLRs. Almost none of the method-LLRs (except around 2 for set A,B) follow the line $y = x$.

There is only one misleading LR for set A,B, though a very strong one ($\sim 10^6$). Consequently, method-LRs between 3 and 6 differ from their post-calibrated LRs by at least one order of magnitude. The transformation becomes less resolute in the region between 0 and 2. The method-LRs for set A,C show bad calibration in general. The plot shows only one vertical line at -28. Every method-LLR below and above -52 is transformed to $-\infty$ or ∞ respectively. This is in agreement with the perfect discrimination properties of the model A,C. For set B,C, the overlap region lies roughly between -2.5 and 0.5. The method shows good calibration between -1.5 and 0.5 as there is less than one magnitude difference between the pre- and post-calibrated LRs. On the contrary, the method-LLRs below -1.5 are poorly calibrated because the PAV transform resulted in post-calibrated LRs of at least one order of magnitude less. It is remarkable that some of the positive method-LLRs are transformed to negative LLRs, meaning that LRs once in favor of H_1 are now in favor of H_2.

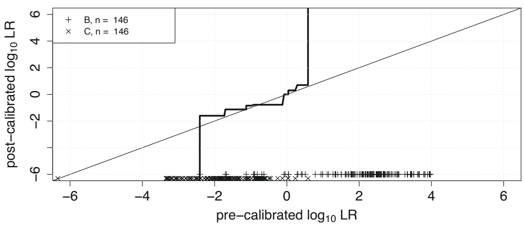
The devPAV equals 11 and 0.89 for model A,B and model B,C respectively. The devPAV for model A,C cannot be derived as the model shows perfect discrimination. A devPAV-value of 11 for model A,B reflects poor calibration and is explained by the strong misleading evidence given H_2. A value of 0.89 indicates better calibration, yet neither of the two models are below the rule-of-thumb value of 0.5 stipulated by Vergeer *et al.* (2021). This rule-of-thumb is based on LR methods with Gaussian distributions and not on LR methods with KDE distributions like our method. Hence, the rule-of-thumb value should be used carefully.



(a)



(b)



(c)

Figure C.1: PAV plots for set A,B (a), set A,C (b) and set B,C (c).

Lower and upper LR bound

Brümmer & Du Preez (2006) state that "any reasonable calibration should result in a score to log-likelihood-ratio mapping that has both an upper and a lower log-likelihood-ratio bound." (p. 266) Hence, the PAV LRs should also be finite outside the overlap region of the LR distributions. We have derived these finite lower and upper bound using two solutions: adding misleading evidence to the set of LRs, similar to Brümmer and Du Preez' procedure, and applying the Empirical Upper and Lower Bound LR (ELUB LR) method proposed by (Vergeer *et al.*, 2016).

The first solution consists of adding misleading evidence of one extremely large LR given H_2 and one extremely small LR given H_1[1]. The first is larger than the largest LR given H_1 and the latter is smaller than the smallest LR given H_2. By adding misleading evidence, the overlap region stretches over the entire range of LRs and so does the PAV transformation. As a result, the PAV LRs equal to $-\infty$ and ∞ are transformed to a finite lower and upper log-likelihood ratio bound. The lower and upper bound are highly data-driven: the larger the data set, the more extreme the LRs. Moreover, adding more misleading evidence results in a smaller LR range.

The Empirical Upper and Lower Bound LR (ELUB LR) method is another method to derive a minimum and maximum value of the method-LRs. Just as in the previous method, adding misleading evidence is the catalyst for determining the limits of the LR. ELUB LR, however, works with pre-calibrated LRs instead of post-calibrated LRs and aims to establish limits to unreasonable large LRs resulting from models with good discrimination properties (such as model A,C). Besides the discrimination properties of the method, the limits of the LR also depend on the size of the data sets.

Tableẽftable2 shows the lower and upper bound for the LRs of all three models derived using the two solutions. Both methods yield the largest LRs for model B,C even though model B,C shows poor discrimination compared to the other two models. These larger LRs are due to the larger sample size of LRs for set B,C ($n = 292$) compared to set A,B ($n = 285$) and set A,C ($n = 182$). The upper bounds of all three models lie within the same order

C

[1]Brümmer & Du Preez (2006) added four data points of misleading evidence to the set of LRs. Hence, our solution is similar but not equal to Brümmer and Du Preez' approach.

	Adding misleading evidence (Brümmer and Du Preez, 2006)	Empirical Upper and Lower Bound LR (ELUB LR) (Vergeer et al., 2016)
A,B	(0.007, 70)	(0.008, 67)
A,C	(0.007, 37)	(0.007, 36)
B,C	(0.03, 127)	(0.04, 96)

Table C.1: Lower and upper bound of post-calibrated PAV LRs (Brümmer & Du Preez, 2006) and of method-LRs (Vergeer *et al.*, 2016).

of magnitude. The LR bounds of both methods do not differ much and both methods seem equally conservative. A consensus on the best method to set an LR bound has not been reached yet and what to report in casework remains open for discussion.

Accuracy

Empirical Cross-Entropy (ECE) plots

An LR is accurate if it supports the true hypothesis and becomes more accurate as the degree of support increases. The Empirical Cross-Entropy (ECE) is a common measure for the accuracy of an LR method (Van Es *et al.*, 2017; Leegwater *et al.*, 2017; Meuwly *et al.*, 2017; Ramos *et al.*, 2013; Vergeer *et al.*, 2014). The ECE measures the average decision-cost of the method. The decision-cost is low if the LR (sufficiently) points in the direction of the true hypothesis. Hence, if the method-LRs are accurate, the ECE values will be low. The ECE can be expressed as:

$$\begin{aligned} ECE = &-\frac{P(H_1)}{N_1}\sum_{i:H_1-true} log_2\left(1+\frac{1}{LR_i\times\frac{P(H_1)}{P(H_1)}}\right) \\ &-\frac{P(H_2)}{N_2}\sum_{j:H_2-true} log_2\left(1+LR_j+\frac{P(H_1)}{P(H_2)}\right) \end{aligned} \tag{C.1}$$

with

$P(H_1)$, $P(H_2)$ = the prior probabilities of H_1 and H_2 respectively

N_1, N_2 = number of LRs given H_1 and H_2 respectively

LR_i, LR_j = LRs given H_1 and H_2 respectively.

Thus the ECE can be calculated with two sets of LRs and fixed prior probabilities. The performance of the method depends on the choice of prior odds and an ECE plot illustrates this dependency: it represents all ECE values as a function of (the 10log of) the prior odds. An ECE plot shows three ECE curves: the ECE curve of the LR method, the ECE curve of the LR method after the PAV transformation and the ECE curve of a reference method with LR = 1 always. The ECE takes values between 0 and 1 and if the ECE = 0 for all prior odds, the method will perfectly discriminate between the two hypotheses.

The metric log-likelihood-ratio cost ($Cllr$) is defined as the ECE at prior odds = 1. If the $Cllr$ = 1, the LR model is as non-informative as a model that always produces LRs equal to 1. An LR method is accurate if the $Cllr$ does not exceed the $Cllr$ of the reference method and the smaller the $Cllr$, the better. In addition, the performance of the system will generally be better after calibration and the ECE curve of the PAV LRs is therefore situated lower than the ECE curve of the LR method.

The ECE was calculated for model A,B; A,C and B,C using Eq. C.1 where the choice of H_1 and H_2 and the LR test sets differed for each model. Fig. C.2 demonstrates the ECE plots for set A,B (a), set A,C (b) and set B,C (c). The solid curves and dashed curves represent the ECE curves of the LR method and the ECE curves of the LR method after PAV transformation respectively. The dotted curve illustrates the ECE values of a reference LR method (LR = 1 always). The ECE curves of all three models are much lower than the reference curve. This means that all models are more informative than the reference method. For set A,B, the ECE values are larger for negative log10 prior odds. This indicates that the performance is worse for prior odds in favor of distance B than for prior odds in favor of distance A. This is due to the 8% misleading evidence given H_2. For set A,C, the ECE values for the output LRs before and after PAV transform equal 0 over all prior odds. This is expected as the LR distributions in Fig. 6.8 and the Tippett plots in Fig. 6.9 showed perfect discrimination. For set B,C, there is no drastic change in ECE after the PAV transformation. Moreover, log10 prior odds above 0 result in only slightly greater ECE values than log10 prior odds below 0, illustrating that the performance of the method is good for prior odds in favor of both H_1 and H_2. The LRs for method A,C are thus most accurate and the LRs for method B,C are least accurate, both before and after calibration.

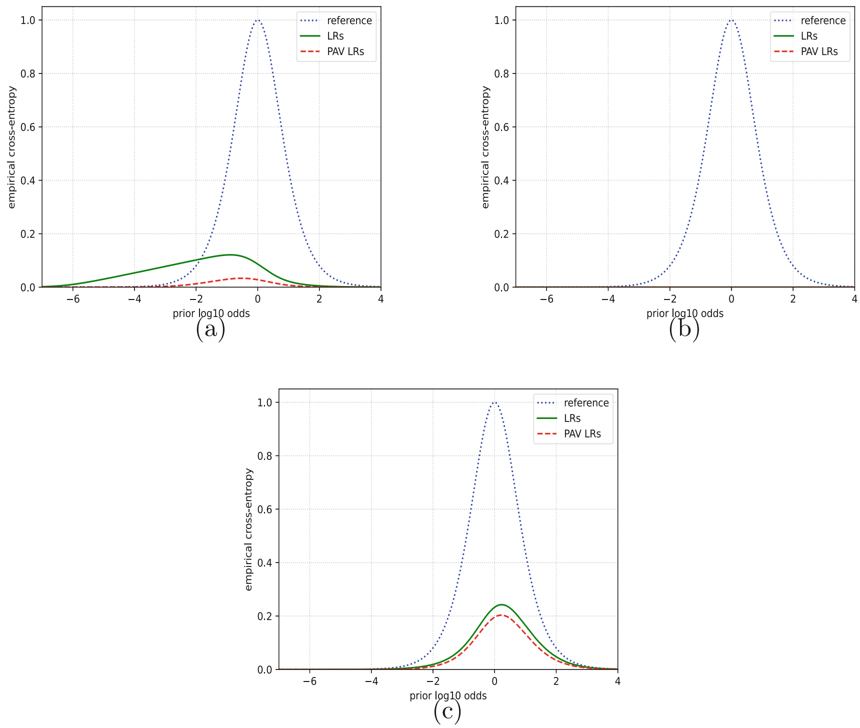


Figure C.2: ECE plots for set A,B (a), set A,C (b) and set B,C (c).

C

D

CHAPTER 8: PROBABILITY ASSIGNMENTS TROJAN HORSE DEFENSE CASES

Node	Probability		Assignment
01. Who knowingly downloaded the files to the computer?	P(H_1 | 39. Level of digital protection = Low)	p_{1a}	0.5
	P(H_2 | 39. Level of digital protection = Low)	$1 - p_{1a}$	0.5
	P(H_1 | 39. Level of digital protection = High)	p_{2b}	0.7
	P(H_2 | 39. Level of digital protection = High)	$1 - p_{2b}$	0.3
04 Unknown person U had remote access to the computer during time T	P(04. Unknown had.. = True | 39. Level of digital protection = Low, H_1 = True)	p_{4a}	0.1
	P(04. Unknown had.. = False | 39. Level of digital protection = Low, H_1 = True	$1 - p_{4a}$	0.9
	P(04. Unknown had.. = True | 39. Level of digital protection = High, H_1 = True)	p_{4b}	0.02
	P(04. Unknown had.. = False | 39. Level of digital protection = High, H_1 = True)	$1 - p_{4b}$	0.98
05. Who performed regular user activities on the computer around time T?	P(Mr. X | H_1)	p_{5a}	0.9
	P(Someone else | H_1)	p_{5b}	0.05
	P(No normal activities | H_1)	$1 - p_{5a} - p_{5b}$	0.05
	P(Mr. X | H_2)	p_{5c}	0.05
	P(Someone else | H_2)	p_{5d}	0.25
	P(No normal activities | H_2)	$1 - p_{5c} - p_{5d}$	0.7
06. Who performed file-ownership related activities after the download on September 8, 2023?	P(Mr. X | H_1)	p_{6a}	0.9
	P(Unknown person U | H_1)	p_{6b}	0.02
	P(No activities | H_1)	$1 - p_{6a} - p_{6b}$	0.08
	P(Mr. X | H_2)	p_{6c}	0.0
	P(Unknown person U | H_2)	p_{6d}	0.5
	P(No activities | H_2)	$1 - p_{6c} - p_{6d}$	0.5
17. Mr. X used his bank account for money transfer	P(17. Mr. X used.. = True | 07. Mr. X performed regular activities.. = True)	p_{17}	0.1
	P(17. Mr. X used.. = False | 07. Mr. X performed regular activities.. = True)	$1 - p_{17}$	0.9
18. Someone else used Mr. X's bank account for money transfer	P(18. Someone else used Mr. X's.. = True | 08. Someone else performed regular activities.. = True)	p_{18}	1.0×10^{-6}
	P(18. Someone else used Mr. X's.. = False | 08. Someone else performed regular activities.. = True)	$1 - p_{18}$	0.999999
19. Someone else used own bank account for money transfer	P(19. Someone else used own.. = True | 08. Someone else performed regular activities.. = True)	p_{19}	1.0×10^{-6}
	P(19. Someone else used own.. = False | 08. Someone else performed regular activities.. = True)	$1 - p_{19}$	0.99999
20. Mr. X used eMule application to search for the contested files	P(20. Mr. X used LogMeIn.. = True | 02. Mr. X knowingly downloading the files.. = True)	p_{20}	0.8
	P(20. Mr. X used LogMeIn.. = False | 02. Mr. X knowingly downloading the files.. = True)	$1 - p_{20}$	0.2
21. Mr. X copied the contested files to home directory	P(21. Mr. X copied.. = True | 02. Mr. X knowingly downloading the files.. = True)	p_{21}	0.9
	P(21. Mr. X copied.. = False | 02. Mr. X knowingly downloading the files.. = True)	$1 - p_{21}$	0.1
22. U used eMule application to search for the contested files	P(22. U used eMule.. = True | 03. Unknown person U knowingly downloading the files.. = True)	p_{22}	0.8
	P(22. U used eMule.. = False | 03. Unknown person U knowingly downloading the files.. = True)	$1 - p_{22}$	0.2
23. U copied the contested files to home directory	P(23. U copied.. = True | 03. Unknown person U knowingly downloading the files.. = True)	p_{23}	0.5
	P(23. U copied.. = False | 03. Unknown person U downloading the files.. = True)	$1 - p_{23}$	0.5
24. U used LogMeIn to control the computer	P(24. U used LogMeIn.. = True | 04. Unknown person U had remote access.. = True)	p_{24}	0.98
	P(24. U used LogMeIn.. = False | 04. Unknown person U had remote access.. = True)	$1 - p_{24}$	0.02
25. Mr. X used VLC to view the contested files	P(25. Mr. X used VLC.. = True | 07. Mr. X performing file-ownership.. = True)	p_{25}	0.95
	P(25. Mr. X used VLC.. = False | 07. Mr. X performing file-ownership.. = True)	$1 - p_{25}$	0.05
26. Mr. X used VLC to view the contested files	P(26. U used VLC.. = True | 08. Unknown person U performing file-ownership.. = True)	p_{26}	0.95
	P(26. U used VLC.. = False | 08. Unknown person U performing file-ownership.. = True)	$1 - p_{26}$	0.05

Table D.1: Probability assignments for p_n, where n represents the node number in the Bayesian network shown in Chapter 8, Fig. 8.7.

Node	Probability		Assignment
27. Generation and persistence of bank transfer traces...via Mr. X..	P(27. Generation.. = True | 17. Mr. X used his bank account.. = True)	p_{27}	0.98
	P(27. Generation.. = False | 17. Mr. X used his bank account.. = True)	$1 - p_{27}$	0.02
28. Generation and persistence of bank transfer traces...via someone else..	P(28. Generation.. = True | 18. U used Mr. X's bank account.. = True)	p_{28}	0.98
	P(28. Generation.. = False | 18. U used Mr. X's bank account.. = True)	$1 - p_{28}$	0.02
29. Generation and persistence of bank traces related to U...via someone else..	P(29. Generation.. = True | 18. U used Mr. X's bank account.. = True)	p_{29}	0.2
	P(29. Generation.. = False | 18. U used Mr. X's bank account.. = True)	$1 - p_{29}$	0.8
30. Generation and persistence of search history traces to the computer via Mr. X using..	P(30. Generation.. = True | 20. Mr. X used eMule application.. = True)	p_{30}	0.98
	P(30. Generation.. = False | 20. Mr. X used eMule application.. = True)	$1 - p_{30}$	0.02
31. Transfer and persistence of files to home directory via Mr. X copying..	P(31. Transfer.. = True | 21. Mr. X copied files to home directory = True)	p_{31}	1
	P(31. Transfer.. = False | 21. Mr. X copied files to home directory = True)	$1 - p_{31}$	0
32. Generation and persistence of search history traces to the computer via U using..	P(32. Generation.. = True | 22. U used eMule application.. = True)	p_{32}	0.98
	P(32. Generation.. = False | 22. U used eMule application.. = True)	$1 - p_{32}$	0.02
33. Transfer and persistence of files to home directory via U copying..	P(33. Transfer.. = True | 23. U copied files to home directory = True)	p_{33}	1
	P(33. Transfer.. = False | 23. U copied files to home directory = True)	$1 - p_{33}$	0
34. Generation and persistence of LogMeIn traces on computer via U using LogMeIn	P34. Generation = True | 24. U used LogMeIn to control.. = True)	p_{34}	0.98
	P(34. Generation.. = False | 24. U used LogMeIn to control.. = True)	$1 - p_{34}$	0.02
35. Generation and persistence of VLC traces on computer via Mr. X using VLC..	P(35. Generation.. = True | 25. Mr. X used VLC to view.. = True)	p_{35}	0.95
	P(35. Generation.. = False | 25. Mr. X used VLC to view.. = True)	$1 - p_{35}$	0.05
36. Generation and persistence of VLC traces on computer via U using VLC..	P(36. Generation.. = True | 26. U used VLC to view.. = True)	p_{36}	0.95
	P36. Generation.. = False | 26. U used VLC to view.. = True)	$1 - p_{36}$	0.05
39. Level of digital protection	P(39. Level of digital protection = High)	p_{39}	0.5
	P(39. Level of digital protection = Low)	$1 - p_{39}$	0.5

Table D.2: Probability assignments for p_n, where n represents the node number in the Bayesian network shown in Chapter 8, Fig. 8.7.

Follow the trace

Forensic scientists face questions ranging from “What is the source of this trace?” to more complex issues like “Who did what, when, where, and how?” These latter questions, known as “activity level questions,” become particularly challenging when they involve multiple types of traces, such as DNA traces and fibers. Evaluating the combined strength of evidence in these cases provides great value to the court but requires managing numerous probabilities arising from various uncertainties. Such complex casework requires structured probabilistic reasoning, and Bayesian networks prove to be valuable tools for this purpose.

By using fictive case examples throughout this work, we advocate the benefits of the LR framework and Bayesian networks for evaluating forensic findings given activity level propositions across various disciplines—including digital forensic science.

The result? A collection of “building blocks” representing general probabilistic forensic problems and two template Bayesian networks applicable to both mono- and interdisciplinary casework that map pathways from alleged activities to traces found on items of interest (i.e., “Follow the trace”). Additionally, two use cases in digital forensic science are presented: one involving iPhone Health app data and another addressing Trojan horse defense cases.

This work aims to serve as a valuable reference for forensic scientists in their casework and research while also providing useful insights for other professionals in the forensic and legal fields.