Combining forensic evidence

de Zoete, J.C.

Link to publication

Citation for published version (APA):
de Zoete, J. C. (2016). Combining forensic evidence

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
In this thesis I consider the evaluation of a combination of different pieces of evidence in a legal and a forensic context. Evaluation of forensic evidence is the main topic of a research area called forensic statistics. In forensic statistics, the likelihood ratio framework is regarded as the standard for evaluating evidence. In legal practice, it is common that two competing propositions are presented to the trier of fact. The trier of fact needs to establish whether the proposition presented by the prosecution can be proven to the extent that there could be no reasonable doubt in the mind of a reasonable person that the defendant is guilty. The presented evidence should rule out any reasonable doubt. The likelihood ratio framework is based on probabilistic inference by applying Bayes' Theorem. It allows for the transition from prior belief regarding the presented propositions to posterior beliefs. This transition is based on the conditional probabilities to observe the evidence given the propositions, the likelihood ratio.

In forensic casework, it is common that multiple pieces of evidence that need to be evaluated in terms of their support regarding the presented propositions are available. The most straightforward way of doing this for a forensic expert is by presenting a separate likelihood ratio for each individual piece of evidence. However, when doing so, one needs to be confident that the individual reports are optimally combined by the trier of fact. When forensic experts believe that their knowledge regarding the dependency structure between pieces of evidence is lost by presenting the likelihood ratios separately, one should strive to combine this evidence before it is sent to the trier of fact. Especially in situations where the pieces of evidence are of the same type (e.g., two shoe marks), one usually cannot regard them as conditionally independent observations and a combined evaluation is needed to prevent unnecessary misconceptions.
Combining Forensic Evidence

Jacob Coenraad de Zoete
Combining Forensic Evidence

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in
de Agnietenkapel
op 22 november 2016, te 14:00 uur

door

Jacob Coenraad de Zoete

geboren te Hoorn, Nederland
Promotiecommissie

Promotores:
Prof. dr. M. J. Sjerps
*Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam

Prof. dr. R. W. J. Meester
*Department of Mathematics
*Vrije Universiteit Amsterdam

Overige leden:
Prof. dr. M. R. H. Mandjes
*Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam

Prof. dr. A. D. Kloosterman
*Institute for Biodiversity and Ecosystem Dynamics
Universiteit van Amsterdam

Dr. A. J. van Es
*Korteweg-de Vries Institute for Mathematics
Universiteit van Amsterdam

Prof. dr. J. M. Curran
*Department of Statistics
*University of Auckland

Prof. dr. D. A. Lagnado
*Causal Cognition Lab
*University College London

Faculteit der Natuurwetenschappen, Wiskunde en Informatica
This dissertation is submitted in partial fulfillment of the requirements for conferral of the degree of Doctor of Philosophy at the University of Amsterdam.

The printing of this thesis was supported by the Co van Ledden Hulsebosch Centrum, Amsterdam Center for Forensic Sciences and Medicine.

This research was carried out under the objectives of the Combining Evidence project in the Forensic Science programme financed by the Netherlands Organisation for Scientific Research (NWO) with Grant Number 727.011.007.
1 Introduction
  1.1 Motivation .......................................................... 1
  1.2 The likelihood ratio framework .................................. 2
    1.2.1 Combining evidence ........................................... 3
    1.2.2 Bayesian networks ............................................ 4
    1.2.3 Hierarchy of propositions ................................. 7
  1.3 Evidence .......................................................... 7
    1.3.1 DNA .......................................................... 8
    1.3.2 Other forensic evidence .................................... 10
  1.4 Outline of contents .............................................. 12
    1.4.1 Chapter 2: Combining two types of DNA profiles from the same stain ........................................... 12
    1.4.2 Chapter 3: Assessing the relevance of evidence - DNA profile from a stain obtained from adhesive tape ....... 13
    1.4.3 Chapter 4: Assessing the relevance of evidence - categorical methods for RNA profile interpretation ................. 14
    1.4.4 Chapter 5: Assessing the relevance of evidence - probabilistic methods for RNA profile interpretation .............. 14
    1.4.5 Chapter 6: Combining cell type results with DNA profiles .................................................. 14
    1.4.6 Chapter 7: Comparing crime linkage from a probabilistic and a Dutch legal perspective ........................... 15
    1.4.7 Chapter 8: A Bayesian network approach for evaluating evidence in crime linkage scenarios - one offender .......... 15
    1.4.8 Chapter 9: A Bayesian network approach for evaluating evidence in crime linkage scenarios - multiple offenders .... 16

2 The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample .......... 19
  2.1 Introduction .......................................................... 19
  2.2 The database independence test - method and result ........... 22
  2.3 The simulation - method ............................................ 23
  2.4 The simulation - results ........................................... 24
  2.5 Conclusion and Discussion ....................................... 27

3 The interpretation of traces found on adhesive tapes 35
  3.1 Introduction .......................................................... 35
  3.2 The Bayesian network .............................................. 37
    3.2.1 Transfer ........................................................ 38
    3.2.2 Persistence ..................................................... 40
    3.2.3 Composition of the trace and manner of deposition ........ 41
    3.2.4 Recovery ....................................................... 42
    3.2.5 DNA analysis .................................................. 42
    3.2.6 The position of the sampling location on the original roll 43
    3.2.7 Sensitivity analyses .......................................... 51
CONTENTS

3.3 Discussion and conclusion ........................................... 53

4 Categorical methods for the interpretation of RNA profiles as cell type evidence and their limitations ........................................... 57
  4.1 Introduction ........................................................... 57
  4.2 Method overview ...................................................... 58
    4.2.1 Basic method .................................................. 58
    4.2.2 Relative expression method .................................. 58
    4.2.3 $n/2$ method ................................................... 59
    4.2.4 Marker value and scoring threshold method ................. 59
  4.3 Discussion and conclusion .......................................... 60

5 A probabilistic approach for the interpretation of RNA profiles as cell type evidence .................................................. 63
  5.1 Introduction ........................................................... 64
  5.2 Literature overview of RNA interpretation methods ............... 65
    5.2.1 Multiplex mRNA profiling for the identification of body fluids 65
    5.2.2 $n/2$ scoring method .......................................... 66
    5.2.3 Marker value and scoring threshold method ................... 67
    5.2.4 Other methods ................................................. 70
    5.2.5 Summary ....................................................... 70
  5.3 Material and methods ............................................... 72
  5.4 Probabilistic methods .............................................. 72
    5.4.1 Bayesian network/ naïve Bayes’ approach .................... 73
    5.4.2 Multinomial logistic regression ................................ 78
  5.5 Method comparison ................................................. 79
    5.5.1 Data from Lindenbergh et al. [12] .......................... 79
    5.5.2 Data from Roeder et al. [17] ................................. 82
  5.6 Conclusion and discussion ........................................... 86
  5.A Overview of the used marker data ................................ 91
  5.B $p$-values overview ................................................ 93

6 Cell type determination and association with the DNA donor ............ 95
  6.1 Introduction ........................................................... 96
  6.2 Literature overview ................................................ 97
    6.2.1 The prostate specific antigen test .......................... 98
    6.2.2 Effect of sample type, time and temperature ................ 98
    6.2.3 Cross reactivity of the PSA test ............................. 99
  6.3 Mock case example .................................................. 102
    6.3.1 Estimating conditional probabilities for classical cell type
determination tests using literature .................................. 103
    6.3.2 Integration of LR mix with the BN for mixture calculations 105
    6.3.3 Results ......................................................... 107
    6.3.4 Sensitivity analysis ........................................... 108
  6.4 Discussion ........................................................... 111
    6.4.1 An example without positive, gender associated, cell type tests111
6.4.2 Dealing with prior probabilities ......................................... 114
6.4.3 Additional information regarding cell type test ..................... 114
6.5 Conclusion ........................................................................ 115
6.A Literature overview sensitivity and specificity of cell type tests ... 119
6.B States of the configuration nodes ........................................... 124

7 Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten 125
7.1 Inleiding ........................................................................ 125
7.2 Juridische context; dubbele bevestiging als beginsel van strafrechtelijk bewijs ......................................................... 126
  7.2.1 Het formele perspectief: hoe komt de rechterlijke overtuiging tot stand? ................................................................. 126
  7.2.2 Het materiële perspectief: op welke gronden berust het rechterlijke oordeel? .......................................................... 129
7.3 Kanstheoretische context; modellen en Bayesiaanse netwerken ...... 131
7.4 Schakelbewijs .................................................................... 136
  7.4.1 Juridische praktijk ............................................................ 136
  7.4.2 Kanstheoretisch ............................................................. 138
  7.4.3 Balans ........................................................................ 139
7.5 Verantwoording van het rechterlijke oordeel: de modus operandi als conditio sine qua non ...................................................... 139
  7.5.1 Juridische praktijk ............................................................ 140
  7.5.2 Kanstheoretisch ............................................................. 141
  7.5.3 Balans ........................................................................ 142
7.6 Motivering van de bewezenverklaring ...................................... 143
  7.6.1 Moet voor elk feit afzonderlijk worden vastgesteld dat een strafbaar feit begaan is? ......................................................... 143
  7.6.2 Moet schakelbewijs verankerd worden? .......................... 145
  7.6.3 Telkens één schakel: een waarborg tegen overwaardering van bewijs? ................................................................. 149
  7.6.4 Conclusie .................................................................... 151

8 Modelling crime linkage with Bayesian networks 155
8.1 Introduction ........................................................................ 155
8.2 Literature and state of the art on crime linkage ......................... 157
  8.2.1 Literature on how to identify linked cases ....................... 157
  8.2.2 Literature on modelling crime linkage ............................ 157
8.3 Using Bayesian networks when there are two linked crimes ....... 159
  8.3.1 The basic assumptions ................................................... 159
  8.3.2 Similarity evidence ....................................................... 160
  8.3.3 “Dependent” evidence .................................................... 163
8.4 Three linked cases ................................................................ 165
  8.4.1 Assumptions about the evidence ..................................... 165
  8.4.2 Evidence proving innocence in the third case ................. 167
8.5 Discussion ......................................................................... 170
8.A The number of scenarios given the number of cases . . . . . 172

9 Evaluating evidence in crime linkage scenarios with multiple offenders 175
  9.1 Introduction ................................................................. 175
  9.2 Literature overview ..................................................... 176
    9.2.1 Identifying linked crimes ........................................ 176
    9.2.2 Consistency and behaviour of groups in serial crime ..... 178
  9.3 Modelling crime linkage with multiple offenders ............ 178
    9.3.1 Different situations ................................................ 180
    9.3.2 Alternative hypotheses: offender configurations ......... 184
    9.3.3 Bayesian networks representing the different situations . 185
  9.4 Computations and results ............................................. 190
    9.4.1 Posterior probabilities for different situations .......... 190
    9.4.2 Same offender evidence .......................................... 193
  9.5 Discussion ............................................................... 195
    9.5.1 Alternative distribution of evidence ......................... 195
  9.A Offender configurations .............................................. 199

Summary 201

Samenvatting 205

Overview of author contributions 209

Acknowledgements 213
In this thesis I consider the evaluation of a combination of different pieces of evidence in a legal and a forensic context. Evaluation of forensic evidence is the main topic of a research area called forensic statistics. Below, I give a basic introduction to the field of forensic statistics and the different types of combining evidence most relevant for this thesis, followed by a brief introduction on Bayesian networks which are used as tools for the evaluation of evidence in many chapters of this thesis. I present different types of evidence in Section 1.3. Lastly, I present an outline of the thesis, introducing each of the chapters, the link between them and their relation with the overall topic in detail.

1.1 Motivation

The unus-testis, nullus testis principle\(^1\) states that uncorroborated evidence of a single witness will be discounted. In other words, *one witness is no witness*. Hence, having multiple observations is an explicit desire in legal practice. In a situation where corroborating evidence is available, the trier of fact needs to consider how to evaluate the statement of the witness, *given the corroborating evidence*. Hence, he or she has to assess the dependency relation between the two observations. For example, when the corroborating evidence is another witness statement that agrees with the first, one would feel more certain when the second witness is an unknown to the first because this suggests a certain level of independence between them. Requiring multiple observations in order to reach a verdict suggests that one can make more reliable decisions. However, the dependency structure between these observations is crucial when evaluating the strength of these observations combined.

Apart from witness statements, forensic evidence supporting claims made by either the prosecution or the defense can and often will be vital in reaching a verdict. Typically forensic evidence addresses a sub-question instead of the ultimate question on which this verdict is based. For example, (1) a DNA profile on a crime scene is informative for the sub-question regarding who was possibly present at the crime scene and (2) a test supporting the presence of blood on a t-shirt found

\(^1\)Art. 342 of the Dutch Code of Criminal Procedure
1. Introduction

at the crime scene may increase the belief that someone was assaulted. In such a situation, these sub-questions need to be combined to address the more relevant question, Did the suspect assault someone? Combining different pieces of evidence can be the key to getting a better understanding of what happened at the crime scene.

When considering how different pieces of evidence influence each other and the question of interest, it is necessary to consider several different situations. Pieces of evidence relevant for the same question (i.e. who was present at the crime scene?) are easier to combine in terms of evaluation of the evidence than when they are relevant for different questions (i.e. who was present at the crime scene? and what happened at the crime scene?). In this thesis, methods for combining evidence are presented for these different situations, using common scenarios from forensic practice (Chapter 2, 3 and 8). Apart from different situations, the dependency relation between the pieces of evidence can substantially influence the combined evidential value. Assessing this dependency relation and how this affects questions relevant in trials is discussed in Chapters 2, 4 and 5. Lastly, how mathematical models can aid legal practitioners in their reasoning is explored in Chapter 7 and 9.

The proposed methods in this thesis are presented using specific situations from forensic/legal practice. However, many of them can serve as a blueprint for different (forensic) scenarios. For example, evaluating evidence in crime linkage scenarios (Chapter 8) is from a probabilistic point of view very similar to evaluating evidence in one crime when uncertainty exists regarding whether there were different sources of separate pieces of evidence. Most importantly, this thesis hopefully helps us remind that combining evidence to assess the most likely cause of events, although necessary in many cases, can be very complicated. Probabilistic models can be particularly useful to structure the problem, identify influential parameters and pitfalls and to evaluate the evidence.

1.2 The likelihood ratio framework

The likelihood ratio framework is regarded internationally as the standard for evaluating evidence [23, 1, 3], although it is not uncriticized, see [15]. In legal practice, it is common that two competing hypotheses are presented to the trier of fact. The trier of fact needs to establish whether the proposition being presented by the prosecution (\(H_p\)) can be proven to the extent that there could be no “reasonable doubt” in the mind of a “reasonable person” that the defendant is guilty. The presented evidence (\(E\)) should rule out any reasonable doubt. The likelihood ratio framework is based on probabilistic inference by applying Bayes’ Theorem (Equation 1.1). It allows for the transition from prior (initial) to posterior (final) probabilities on the basis of evidence.

\[
\Pr(H_p|E) = \frac{\Pr(E|H_p)}{\Pr(E)} \cdot \Pr(H_p).
\]  

(1.1)

To assess the value of the evidence, the proposition presented by the defense is necessary. The odds form of Bayes’ theorem (Equation 1.2) splits the calculation
1.2. The likelihood ratio framework

into two parts. The odds in favor of the proposition of the prosecution \( H_p \) based on the observed evidence \( E \) is equal to the likelihood ratio times the prior odds.

\[
\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \cdot \frac{\Pr(H_p)}{\Pr(H_d)}
\]

In the likelihood ratio framework, a forensic expert evaluates the value of the forensic evidence in terms of conditional probabilities. What is the probability of observing the evidence \( E \) given the hypotheses presented by the prosecution \( (H_p) \) and the defense \( (H_d) \). The evidential value is reported as a likelihood ratio. The likelihood ratio summarizes to what extent an observation \( E \) supports one hypothesis over another. A likelihood ratio greater than one indicates that the observation supports \( H_p \) over \( H_d \). A likelihood ratio less than one corresponds with evidence that supports \( H_d \) over \( H_p \). When the likelihood ratio is equal to one, the evidence does not support one hypothesis over the other.

The trier of fact considers how probable the hypotheses are, before considering the evidence, based on background information, and, possibly, other (non-forensic) evidence. This background information could, for example, contain the number of potential offenders. The prior probabilities, combined with the likelihood ratio determine the posterior odds. Hence, the likelihood ratio framework approach is a workflow where the forensic expert assigns conditional probabilities based on data and/or knowledge to determine the likelihood ratio, and the trier of fact considers the ‘non-forensic’ evidence to determine the prior odds.

1.2.1 Combining evidence

For multiple pieces of evidence, \( E_1 \) and \( E_2 \), one could report the likelihood ratio of all these pieces individually, \( \frac{\Pr(E_1|H_p)}{\Pr(E_1|H_d)} \) and \( \frac{\Pr(E_2|H_p)}{\Pr(E_2|H_d)} \). In situations where the evidence covers multiple domains (e.g. a fingerprint and a DNA profile) this is common. However, when doing so, one needs to be confident that the individual reports are optimally combined by the trier of fact. When forensic experts believe that their knowledge regarding the dependency structure between pieces of evidence is lost by presenting the likelihood ratios separately, one should strive to combine this evidence before it is sent to the trier of fact. Especially in situations where the evidential pieces are of the same type (e.g. two shoe marks showing a similar sole pattern), one usually cannot regard them as conditionally independent observations and a combined evaluation is needed to prevent unnecessary misconceptions.

Evaluating forensic evidence (e.g. finger marks, DNA profiles) requires knowledge of the associated field (e.g. biometrics, genetics) to evaluate the corresponding likelihoods. For combining evidence, one needs to be able to identify the dependency structure of the problem and understand the consequences following from this structure. A stochastic approach using probabilistic models allows us to do both. Bayesian networks can be used to represent the probabilistic relationships between observations and to perform the necessary calculations. Furthermore,
they provide a graphical overview of the assumptions regarding the conditional dependency structure. Bayesian networks are used as probabilistic models in Chapter 3, 4, 5, 8 and 9.

1.2.2 Bayesian networks

A Bayesian network is a graphical representation of the dependency structure between a set of random variables. Random variables are represented with nodes, conditional dependencies with directed edges between nodes. For discrete variables, each node consists of a set of mutually exclusive states which represent the set of possible values of the random variable. A very basic Bayesian network is given in Figure 1.1. Nodes without incoming edges are called parent nodes. Nodes with incoming edges are child nodes. Apart from the graphical representation and definitions of the nodes and their states, (conditional) probability tables are necessary for the complete representation of the problem. Parent nodes have a prior distribution, child nodes a conditional probability table. For the example from Figure 1.1, this corresponds with the tasks of respectively the trier of fact and the forensic specialist. Examples of (conditional) probability tables for the Bayesian network for Figure 1.1 are given in Table 1.1 and 1.2.

<table>
<thead>
<tr>
<th>hypothesis</th>
<th>prosecution hypothesis</th>
<th>defense hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>match</td>
<td>$r$</td>
<td>$s$</td>
</tr>
<tr>
<td>no match</td>
<td>$1 - r$</td>
<td>$1 - s$</td>
</tr>
</tbody>
</table>

Table 1.1: Probability table for the hypotheses node

<table>
<thead>
<tr>
<th>hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>prosecution hypothesis</td>
</tr>
<tr>
<td>match</td>
</tr>
<tr>
<td>no match</td>
</tr>
</tbody>
</table>

Table 1.2: Probability table for the evidence node

The prior odds follow from the probability table of the hypotheses node.

$$\frac{\Pr(H_p)}{\Pr(H_d)} = \frac{p}{q}.$$
1.2. The likelihood ratio framework

The likelihood ratio can be obtained from the conditional probability table of the evidence node, Table 1.2.

\[
\frac{\Pr(E = \text{match}|H_p)}{\Pr(E = \text{match}|H_d)} = \frac{r}{s}
\]

Software packages like Hugin [12], AgenaRisk [16], Genie [4] or the R [18] package gRain [14] can assist in computing posterior odds/probabilities based on inserted evidence. However, for this example, the posterior odds can easily be computed manually.

\[
\frac{\Pr(H_p|E = \text{match})}{\Pr(H_d|E = \text{match})} = \frac{p \cdot r}{q \cdot s}.
\]

For more complex Bayesian networks, i.e. networks with more nodes and edges, capturing the influence of observations on states of interest manually becomes practically unfeasible. With the aid of software, which basically repeatedly applies Bayes’ theorem, one is still able to examine the impact of observations on other random variables.

When combining evidence in forensic practice, different situations should be recognized. Situations can be distinguished based on different dependency structures, i.e. different Bayesian networks. In Figure 1.2, a Bayesian network representation is presented for a situation where two conditionally independent pieces of evidence are relevant for the same set of hypotheses, e.g., a DNA profile and a finger mark that both suggest that the suspect was the donor of the mark.

![Figure 1.2: Bayesian network for the combined evaluation of two conditionally independent pieces of evidence relevant for the same hypothesis pair](image)

When the evidence is conditionally dependent, given the hypothesis the structure from Figure 1.3 is needed to evaluate the evidence. The edge between evidence 1 and evidence 2 represents the conditional dependency relation between them. Evidence 2 depends on the state of evidence 1, given the hypotheses. For example, when evidence 1 and evidence 2 represent shoe marks and it is given that an unknown person, whose shoe size is unknown, made both marks (hypothesis node), the observations evidence 1 and evidence 2 are dependent. If the first shoe mark is of size 12 so will the second shoe mark. Such a structure prevents that evidence is ‘counted twice’ in the evaluation. In Chapter 2 it is examined whether certain DNA evidence should be modeled as in Figure 1.2 compared to Figure 1.3.

In Figure 1.4, the situation where the different pieces of evidence are relevant for different sub-hypotheses, e.g. a DNA profile relevant for who was present at the
crime scene and a cell type test relevant for what happened at the crime scene, is presented as a Bayesian network structure.

Such a structure is used in Chapter 4. The dashed line between evidence 1 and evidence 2 is only present when they are conditionally dependent observations. In Chapter 8 and Chapter 9 this structure is used for the evaluation of evidence in crime linkage scenarios. Both conditionally dependent and independent evidence is evaluated in these chapters.

Lastly, Figure 1.5 shows a Bayesian network for evaluating evidence relevant for distinguishing between the hypotheses of interest and evidence relevant for the relevance of the first evidence. For example when a fingermark suggests that a suspect touched an object on the crime scene, but other evidence disputes that the object was on the crime scene during the crime, the latter disputes the relevance of the object. A structure regarding the relevance of evidence is presented in Chapter 3. In this chapter, the discriminative value of a matching DNA profile regarding two activity related hypotheses is determined based on the position of the crime stain on a set of pieces of tape.

Bayesian networks that can assist in evaluating common situations in forensic practice have been developed [8, 21]. Most of these consider propositions that dispute the source of material found on the crime scene. Recently, efforts have been made to develop Bayesian networks that assist in evaluating propositions that are of an higher level in the hierarchy of propositions [22, 20].
1.2.3 Hierarchy of propositions

In 1998, Cook et al. [9] introduced the concept of a ‘Hierarchy of propositions’. This hierarchy consists of three levels, (1) source, (2) activity and (3) offense level. Examples of propositions for these levels are given in Table 1.3. Generally, the higher the level, the more informative the answer will be to the trier of fact. However, assigning conditional probabilities for observing the evidence given the propositions is generally more demanding for propositions of a higher level. For example, for situation A in Table 1.3, determining the likelihood ratio for the source level requires a database with features of glass fragments of various types of glass and a probabilistic model that assigns probabilities to observe a certain glass fragment given the source. At the activity level, more information is needed to evaluate the evidence. For example, how likely is it that glass fragments transfer when smashing a window, what is the probability that these glass fragments are still present during the examination and if the glass fragments are still present, how likely is it that they are recovered by the forensic expert (transfer, persistence and recovery). Although more data is needed to evaluate evidence on activity level, the likelihood ratio is more informative, especially in situations where the source of the material is not disputed, but the activity that caused the transfer of material is. Such a situation is discussed in Chapter 3.

Determining the likelihood ratio for the offense level proposition pair would often require legal judgments. This is beyond the area of expertise of a forensic expert. Hence, forensic experts will usually not evaluate evidence under offense level propositions.

1.3 Evidence

In this section, different types of forensic and non-forensic evidence are introduced. In the individual chapters, many of these are used in examples that clarify the proposed methods. Forensic evidence is evidence obtained by scientific methods and the associated evidential value is usually determined by a forensic expert. Non-scientific evidence considers evidence like eye witnesses and behavioural similarities. The evaluation of non-scientific evidence is within the domain of the trier of fact. Here, DNA evidence is separately introduced as forensic evidence. First of all, because it is present in the majority of the chapters of this thesis. Furthermore, DNA evidence is well suited for numerical evaluation within the likelihood ratio
1. Introduction

<table>
<thead>
<tr>
<th>level</th>
<th>generic</th>
<th>examples</th>
</tr>
</thead>
</table>
| III   | offense | Mr A committed the burglary  
Another person committed the burglary  
Mr B raped Ms Y  
Some other man raped Ms Y  
Mr C assaulted Mr Z  
Mr C had nothing to do with the assault on Mr Z |
| II    | Activity| Mr A is the man who smashed window X  
Mr A was not present when window X was smashed  
Mr B had sexual intercourse with Ms Y  
Some other man had sexual intercourse with Ms Y  
Mr C is the man who kicked Mr Z in the head  
Mr C was not present at the kicking of Mr Z |
| I     | Source  | The glass fragments came from window X  
They came from some other broken glass object  
The semen came from Mr B  
The semen came from some other man  
The blood on Mr C’s clothing came from Mr Z  
The blood on Mr C’s clothing came from an unknown person |

Table 1.3: Examples of the hierarchy of propositions.²

framework. A lot of work has been done to construct necessary databases to evaluate the discriminative value of a particular DNA profile. Additionally, statistical methods have been developed to assign conditional probabilities for increasingly difficult scenarios [7, 11, 13].

Thus, DNA has been an important focus of forensic statistics research. It is also one of the most important types of forensic evidence in terms of the number of cases in which it is used.

1.3.1 DNA

DNA (Deoxyribonucleic acid) is the genetic code that is contained in the nuclei of all human cells. In each nucleus the DNA is distributed over 46 chromosomes, that occur in pairs. From each pair of chromosomes, one is inherited from your father, the other from your mother. These 23 chromosomal pairs can be divided into two groups; there are 22 ‘autosomal’ pairs and one pair of sex chromosomes.

The sex chromosomes are notated with either two X’s or one X and one Y, where two XX represents a woman, and XY represents a man. The remaining autosomal DNA doesn’t contain information on whether the source of the DNA is a man or a woman. DNA contains hypervariable areas that contain repetitive small

²Reprinted from A hierarchy of propositions: deciding which level to address in casework, 38 (4), R. Cook, I.W. Evett ,G. Jackson, P.J. Jones, J.A. Lambert, 231-239, Copyright (2016), with permission from Elsevier
1.3. Evidence

parts of DNA code, e.g., (ATCG-ATCG-ATCG-ATCG-...). For unrelated people one can be almost certain that they have a different number of repeating units in some of these areas. These features make them very attractive in matching DNA from a person to a trace. These repetitive parts are called ‘STRs’, Short Tandem Repeats. The location of the hypervariable areas where these short tandem repeats occurs is called the locus (pl. loci). An allele is the number of times that the repetitive part is occurring on a locus. For example, for ATCG-ATCG-ATCG-ATCG, this is 4. Because chromosomes occur in pairs, a DNA profile can be represented with pairs of alleles on different loci. An example is given in Table 1.4.

<table>
<thead>
<tr>
<th>Locus</th>
<th>alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>15  16</td>
</tr>
<tr>
<td>VWA</td>
<td>17  18</td>
</tr>
<tr>
<td>D16S539</td>
<td>10  12</td>
</tr>
<tr>
<td>D2S1338</td>
<td>17  23</td>
</tr>
</tbody>
</table>

Table 1.4: Autosomal DNA profile

Autosomal DNA profile Table 1.4 is an example of an autosomal DNA profile, i.e. based on loci from the autosomal chromosome pairs. The number of loci in a DNA profile depends on the DNA kit that was used to produce the DNA profile, where 15 or 22 loci are currently common. As previously stated, one can be almost certain that unrelated people have different alleles for at least one locus. For related people, one should be more careful. The autosomal DNA profile of each person is a result of the combination of the autosomal DNA profiles of his/her parents. On each locus, one allele is paternally and the other maternally inherited, both selected with probability 1/2, which is known as Mendelian inheritance. An example is given in Figure 1.6.

The evidential value of an autosomal DNA profile is summarized in a random match probability. This value represents the probability that a ‘random man’ would have the same autosomal DNA profile as that of the sample of interest. Due to the locations of the loci, mostly on different chromosomes or quite far apart, the alleles observed on a locus are usually regarded as independent from the alleles on another locus. A database containing the DNA profiles of a subset of a population can be used to estimate the frequencies of different alleles. See [3] for statistical subtleties when the frequencies of the alleles are unknown and are estimated from a population sample. By multiplying the frequencies corresponding to the alleles of a DNA profile, the random match probability is obtained. When one has two different alleles on a locus (heterozygote), this frequency is multiplied with a factor 2 because the difference between, e.g., 10/12 and 12/10 is not visible on a DNA profile. This factor is left out when one allele is present twice (homozygote). A smaller random match probability corresponds with a more discriminative DNA profile.
1. Introduction

**Y-chromosomal DNA profiles**  A Y-chromosomal DNA profile is a profile only of the Y-chromosome. Therefore, it is only possible to obtain a Y-chromosomal DNA profile from a man. Furthermore, if we ignore the possibility of a (rare) copying mistake (a mutation), a person’s Y-chromosomal DNA profile is a direct copy of the Y-chromosomal DNA profile of his father. An example of this is given in Figure 1.7.

Today, Y-chromosomal DNA profiles usually cover 7-22 loci. Because Y-chromosomal DNA profiles are paternally inherited, Y-chromosomal DNA profiles usually bear substantially less discriminative value than autosomal DNA profiles. However, in crime stains with a male and a female donor, the interpretation of the autosomal DNA profile can be difficult, especially in situations where unequal amounts of material were contributed. This is not uncommon in sexual offense cases. In these cases, a Y-chromosomal DNA profile can be essential as evidence regarding the male donor. Due to the way Y-chromosomal DNA profiles are inherited, acquiring a robust random match probability is substantially more difficult than for autosomal DNA profiles. For this purpose, many different methods have been proposed [2, 6, 19].

1.3.2 Other forensic evidence

Other forensic evidence, such as fingermarks, speech evidence or the chemical composition of illicit drugs follows a similar general interpretation framework as DNA
1.3. Evidence

Figure 1.7: Heritage of Y-chromosomal DNA profiles

Evidence. The characteristics of the evidence are evaluated given two competing hypotheses. Usually, for the defense hypothesis, a sample from a representative population is needed to assess the frequencies of the relevant characteristics. Likewise, for the hypotheses of the prosecution, multiple samples are needed from the suspected source in order to evaluate the distribution of these characteristics within this source, i.e. the intra individual variation. Furthermore, statistical models for estimating the conditional probabilities of interest are fundamentally different for discrete and continuous data. For discrete data, like DNA profiles\(^3\), the frequency of the characteristics of the evidence within a relevant population is used to determine the conditional probability under the defense hypothesis. For continuous data, more sophisticated statistical methods have been developed \([17, 5, 10]\). Also, the concept of a match should be reconsidered for continuous data. For discrete data, like a Y-chromosomal DNA profile, a match represents two profiles that show the same alleles on all loci. For continuous data, like the chemical compositions of ignitable liquids, two samples from the same source will not result in a match in the sense that the compositions will be exactly the same. Generally, these samples will result in very similar compositions. Hence, the label ‘match’ is problematic for continuous data.

\(^3\)Incorporating the peak heights of alleles as data of the DNA profile, instead of just the alleles would result in continuous data.
1. Introduction

1.4 Outline of contents

The thesis is based on the following papers:


which form the basis for Chapters 2 - 9 of this thesis, respectively. The following section provides an outline of the different chapters. The underlying relationship between the chapters and the link between the chapters and the overall topic are discussed. A summary is given at the end of this thesis. The chapters consider aspects from specific forensic domains. For example, combining evidence concerning who donated a stain and evidence concerning what type of material was donated. However, the presented methods can often be applied to similar problems from other (forensic) domains.

1.4.1 Chapter 2: Combining two types of DNA profiles from the same stain

There are multiple types of DNA profiles that can be obtained from a stain. An autosomal DNA profile is the most common and usually bears the most discriminative power. A Y-chromosomal DNA profile can only be obtained when at least one male donor contributed to the stain. The Y-chromosome is inherited from father
1.4. Outline of contents

1.4.1 Chapter 2: Theoretical foundations of evidence assessment

A piece of evidence that ‘matches’ the characteristics of a suspect increases the belief that there is a link between the suspect and the item/location from where the evidence was obtained. However, there can be numerous innocent reasons to observe the matching characteristics. In situations where both the prosecution and defense state that the suspect is the source of the material, but dispute how it got there, the focus shifts to evaluating this question, i.e. a hypothesis pair on the activity level.

In Chapter 3, a Bayesian network approach for a situation where a DNA profile matching the DNA profile of a suspect is obtained from a stain found on a piece of tape used in a criminal offense is presented. The defense states that the suspect owned a roll of tape that was stolen from him. His DNA material was present on that roll from before it was stolen which explains why his material was found on the crime scene. The prosecution hypothesis states that the suspect used the tape during the offense. Hence, the ‘relevance’ of the DNA evidence is disputed and the hypotheses are at the activity level. With a process called ‘physical end matching’ one can reconstruct the most likely original order of the pieces of tape found on the crime scene. Under the defense hypothesis, it is likely that the DNA material was originally on the outside layer of the roll, under the prosecution hypothesis, the DNA profile can originate from a stain anywhere on the tape. The Chapter...
introduces a Bayesian network to evaluate the relevance of the DNA material for distinguishing between the two hypotheses, based on the most likely position on the piece of tape on which the DNA material was found.

1.4.3 Chapter 4: Assessing the relevance of evidence - categorical methods for RNA profile interpretation

In this chapter categorical methods for the evaluation of RNA profiles as cell type evidence are discussed. RNA profiles, like DNA profiles, consist of a collection of markers/loci. However, where DNA profiles are informative for answering the question ‘who is the donor of the stain’, RNA profiles can be used for examining what type of material was donated. In forensic casework, the type of material can be crucial with regards to the relevance of a DNA profile. In sexual offense cases, a DNA match with the suspect becomes substantially more incriminating when it is likely that the DNA material came from semen cells compared to when it came from skin cells. Hence, the combined evaluation of RNA and DNA profiles can result in a likelihood ratio on the activity level. Several methods for interpreting the results of RNA analysis from the literature are discussed. These methods are categorical methods, meaning that they result in a categorical statement, i.e. the sample contains blood. Drawbacks of such non-probabilistic conclusions are examined. In Chapter 5 probabilistic approaches are presented.

1.4.4 Chapter 5: Assessing the relevance of evidence - probabilistic methods for RNA profile interpretation

In Chapter 5, as an alternative to categorical methods, two probabilistic methods, multinomial logistic regression and naïve Bayes, for RNA profile interpretation are proposed. Two data sets are used to compare the results of these probabilistic methods with two commonly used categorical methods. It is shown that these methods outperform categorical methods in terms of the frequency of correct classifications. Furthermore, contrary to the categorical methods, probabilistic methods give the opportunity to report the uncertainty about the cell type. Limitations of such probabilistic methods and the feasibility of such methods for samples containing a mixture of cell types, which are far more common in practice, are discussed.

1.4.5 Chapter 6: Combining cell type results with DNA profiles

Apart from RNA profiles, there exist more commonly applied ‘traditional’ cell type tests that can be used for examining the most likely cell type present in a sample. Such tests, like a PSA test for seminal fluid, have been tested thoroughly on a broad set of cell types. Compared to RNA profiles, there is substantially more data available regarding the sensitivity and specificity of these tests. However, generally, these tests are specific for one cell type whereas RNA profiles can be informative for multiple cell types. Hence, it is possible that it is necessary to perform multiple of these tests for one crime stain. In Chapter 6, a literature overview on the
sensitivity and specificity of three common tests (PSA, RSID semen and RSID saliva) is presented. The data obtained from the literature overview is used to estimate conditional probabilities that can be used in a Bayesian network. This Bayesian network is used to combine the question ‘who donated the material’ with ‘what kind of material was donated’ for stains with multiple donors. A sensitivity analysis is performed examining the added value of performing additional cell type tests. A software package for analyzing the evidential value of a DNA profile is implemented in the Bayesian network. Recommendations for implementing such a structure in practice are given.

1.4.6 Chapter 7: Comparing crime linkage from a probabilistic and a Dutch legal perspective

In Chapters 2 to 6 the evidential values of observations within the same crime have been examined. Chapter 7, which is written in Dutch, examines ‘schakelbewijs’ which is the term used within Dutch legal practice when several cases of which it is believed that they have been committed by the same offender are linked to reach a verdict, even in cases in which the evidence would be insufficient when regarding the case individually. The underlying idea is that, when there is ‘sufficient’ evidence to reach a verdict in the first crime, and the modus operandi of that crime is very similar to a second crime, the prior belief that the same suspect is the offender in this second crime is stronger. Hence, less evidence is needed to reach a verdict in this second crime. The logic behind such a ‘schakelbewijs’ methodology is approached from both a Dutch legal and a probabilistic point of view. These two are compared, leading to insights on how probabilistic reasoning can aid in schakelbewijs casework. The goal of this chapter is that by providing these two different views, people with a legal and people with a mathematical background can learn from each other. The chapter identifies several issues that hopefully lead to discussion in the legal community.

1.4.7 Chapter 8: A Bayesian network approach for evaluating evidence in crime linkage scenarios - one offender

Chapter 8 extends the probabilistic analysis from Chapter 7 by modelling crime linkage with Bayesian networks. A general structure is introduced that allows the interpretation of evidence in simplified crime linkage situations, in which it is assumed that each crime only had one offender. This simplified model allows for the examination of the validity of intuitive reasoning. Most importantly, it is highlighted that crime linkage is a double edged sword, it may be used to prove guilt but also innocence. An important practical consequence of this is that one cannot drop a crime from a series of crimes of which it is believed that they have a common offender because there is exculpatory evidence regarding the suspect. This can result in a substantial overestimation of the evidence in the remaining cases.
1.4.8 Chapter 9: A Bayesian network approach for evaluating evidence in crime linkage scenarios - multiple offenders

In Chapter 9, the methodology of Chapter 8 is extended to situations with multiple offenders. Several situations are distinguished for multiple offender crime linkage problems. For a simplified example, the influence of the assumed situation in terms of posterior probabilities is examined. Due to the complexity of such multiple offender crime linkage problems, probabilistic models in which the dependency structure of observations can be specified are valuable reasoning tools. The importance of understanding the implications of making assumptions regarding the number of offenders and/or whether evidence can be grouped is shown using a mock case example.

Bibliography


The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

Abstract

When a Y-chromosomal and a (partial) autosomal DNA profile are obtained from one crime sample, and both profiles match the suspect’s profiles, we would like to know the combined evidential value. To calculate the likelihood ratio of observing the autosomal and Y-chromosomal DNA profiles combined, we need to know the conditional random match probability of the observed autosomal DNA profile, given the Y-chromosomal match. We examine this conditional probability in two ways: (1) with a database containing 2085 men, and (2) using a simulation model. We conclude that if the Y-chromosomal DNA profiles match, we can still regard the autosomal DNA profile as independent from the Y-chromosomal DNA profile if the matching person is not a descendant of the father of the donor of the (crime) sample. The evidential value can in that case be computed by multiplying the random match probabilities of the individual profiles.

2.1 Introduction

Suppose that a crime stain is found at a crime scene, and that two DNA profiles are obtained from it: a Y-chromosomal DNA profile and a partial autosomal DNA profile. Furthermore, no indication of a mixture is observed so we assume that the two profiles originate from the same person. When a suspect is identified that matches both profiles, the question arises how to assess the combined evidential
The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

Following the likelihood ratio approach described in e.g. Evett and Weir [15], we define the hypotheses

\[ H_p: \text{the suspect is the donor of the crime stain} \]

and

\[ H_d: \text{some other man is the donor of the crime stain} \].

Usually, this other man is assumed to be unrelated to the suspect. We will not make this assumption here. The evidential value of the match is then expressed as the likelihood ratio (LR) of the observations (the autosomal and Y-chromosomal DNA profiles of the stain and of the suspect), considering these hypotheses. It is easy to show that this LR simplifies to \( 1 / (p \cdot q) \), where \( p \) equals the probability that the observed Y-chromosomal profile of the stain matches that of the suspect, and \( q \) equals the conditional probability that the observed autosomal DNA profile of the stain matches that of the suspect, given the Y-chromosomal match.

Combining the evidential value of the DNA profiles by simply multiplying the Y-chromosomal and autosomal random match probabilities can only be done when the profiles can be assumed to be independent, which is not a priori obvious. Indeed, the observation that two men share the same Y-chromosome makes it more likely that they are related, and this makes it more likely that their autosomal profiles will match.

If the population of alternative suspects contains no (or only distant) relatives of the suspect, we can assume independence (conditional on \( H_d \)) between the autosomal and Y-chromosomal DNA profiles and \( q \) simplifies to the "standard" random autosomal match probability. If the population of alternative suspects contains relatives of the suspect but we know the number of relatives as well as their degree of relatedness, we can also compute the likelihood ratio (by using the Weight-of-Evidence Formula from [6], see also Anderson and Weir [5] and Bright, Curran and Buckleton [9]). Unfortunately, in practice we usually do not know the number of relatives, nor their order of relatedness, and in this paper we will investigate this situation.

There are several papers in the literature which discuss the combination of autosomal and Y-chromosomal DNA profiles. Amorim [2] states that combining the evidential value obtained from lineage markers (such as Y-chromosome) with that resulting from individuality markers (autosomal) is difficult, if not impossible. Amorim advises to change the prosecutor’s hypothesis from The suspect is the source of the stain to The suspect or somebody from their lineage is the source of the stain and to report the analysis from the two types of evidential sources separately, along with the distinct theoretical and statistical frameworks underlying them. In that way, it will always be possible to join the two types of evidence, but the assumption of non-involvement of relatives will have to be explicitly accepted by the court in the case under judgment. However, in [11], Weir et al. state that there seems to be neither a logical nor a legal basis for changing the prosecution hypothesis like Amorim suggests. Moreover, they remark that if the possibility of mutation is neglected, then the likelihoods of observing the profiles given the two hypotheses would be identical for lineage markers anyway.

Walsh, Redd and Hammer [19] give several complications in computing Y-autosomal joint match probabilities. Apart from the fact that two individuals sharing the same Y haplotype are likely to be more closely related than two random individuals from the populations, they remark that Y chromosome haplotypes
may be highly informative as to which sub-population an individual belongs, and
this in turn may change the autosomal allele frequencies used to compute the au-
zosomal match probabilities. They performed an independence test on autosomal
and Y STRs on 16 populations. Their results did not give any reason to reject
independence of Y and autosomal markers. They recommend to compute the joint
Y-autosomal matching probability by computing the autosomal match probability
as the product of single-locus genotype frequencies. These are corrected by using
the sampling formula of Balding and Nichols, [7], with $\theta = 0.04$. This can be done
for several populations, and finally the maximum of the resulting random match
probabilities is multiplied with the estimated matching probability for the Y hap-
lotype to obtain the joint random match probability. This recommendation is a
conservative approach to computing the joint Y-autosomal matching probability.
It is based on Walsh et al.’s calculations showing that when population structure
is already present in the autosomes, the additional effect due to conditioning on
the Y is small.

Budowle et al. [13] also performed an independence test on autosomal and Y
STRs, for three sampled populations of unrelated males from Texas. The test did
not give any reason to reject that the frequencies of autosomal and Y STR profiles
can be combined using the product rule.

None of these papers, however, solves our problem. In most cases, the method is
not applicable for a population containing many relatives. In some of the papers, a
database is used to investigate dependence between autosomal and Y-chromosomal
DNA profiles. The databases used are quite small, and when the conclusion of the
database research is that there is no reason to assume dependence we cannot be
sure whether there actually exists dependence in the population. The power of
the statistical test is very important here, see Validating Databases in [10]. The
suggestion of using $\theta = 0.04$ on the population with the maximum Y match prob-
ability has similarities with the ceiling principle, and may be a source of criticism
[16].

In this paper we examine the dependence between matching autosomal and
matching Y-chromosomal DNA profiles in two different ways. First we investigate
dependence in a data set of 2085 Dutch males (blood donors who volunteered
for forensic research, de Kniijff and Sijen, in preparation) gathered by The Foren-
sic Laboratory for DNA Research of the Leiden University Medical Center. The
data shows significant departures from independence. However, this can be due to
several causes which we explore. Secondly, we investigate the impact of the depen-
dence with a general simulation model which gives us the opportunity to assess
the consequences for forensic case work. It turns out that for relatives that do not
have the father of the suspect as an ancestor, the dependence is negligible, and we
can safely compute the random match probability for the two combined profiles by
multiplying the random match probabilities of the individual profiles. A separate
issue is how to compute the random match probabilities of Y-chromosomal DNA
profiles in the first place. We will briefly discuss this problem in Section 2.5.
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

2.2 The database independence test - method and result

In this section we investigate the dependence between autosomal and Y-chromosomal markers in a database containing 2085 men. Our database has information on 23 autosomal markers and 17 Y-chromosomal loci. With 2085 men, we can compare \( \binom{2085}{2} = 2,172,570 \) pairs of men. For each of these pairs of men and for each autosomal marker, we consider the number of autosomal alleles that match (0, 1 or 2). Furthermore, we consider whether or not their Y-chromosomal DNA profiles match (on 17 loci). Thus, we can make 23 contingency tables of 2 by 3. An example of such a table is given in Table 2.1.

<table>
<thead>
<tr>
<th>autosomal alleles</th>
<th>0 matching</th>
<th>1 matching</th>
<th>2 matching</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>no matching Y-haplotype</td>
<td>1447795</td>
<td>685267</td>
<td>39067</td>
<td>2172129</td>
</tr>
<tr>
<td>matching Y-haplotype</td>
<td>281</td>
<td>149</td>
<td>11</td>
<td>441</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1448076</td>
<td>685416</td>
<td>39078</td>
<td>2172570</td>
</tr>
</tbody>
</table>

Table 2.1: Contingency table for autosomal locus D1S1656

We calculated the chi-square test statistic for each contingency table. Since we are comparing all possible pairs of men, the observations in the contingency table are not independent (a match between man 1 and 2 and a match between man 1 and 3 means there will be a match between man 1 and 3). Therefore, similar to [19, 12, 18], the distribution of the chi-square test statistic, assuming independence, was determined by a permutation test. We constructed new databases containing 2085 men by making permutations of the auto and Y-chromosomal DNA profiles in our database.

This way, we found 5 autosomal loci (D2S1338, SE33, TH01, D12S391, D13S317) with approximate \( p \)-values less or equal to 0.05 of which 1 was below 0.01 (SE33). The probability of finding at least 5 \( p \)-values less or equal 0.05 in 23 experiments assuming that the observations are independent is approximately 1 in 200, hence, we have reason to believe that there may be dependence between the markers. However, there are several explanations for the low \( p \)-values: (1) there actually exists dependence between autosomal and Y-chromosomal markers, or (2) the database contains relatives. Due to the way the database was constructed (using blood donors), it is not unlikely that the latter is the case. Especially on loci with high allelic variation (SE33), relatives in the database can have major influence on observing dependence between autosomal and Y-chromosomal DNA profiles. Indeed, The Forensic Laboratory for DNA Research of Leiden University Medical Center confirmed that there are at least three father-son pairs. It is not unlikely that there are more relatives in the database.

In Figure 2.1 we give boxplots with the number of matching autosomal alleles for all the pairs of men given the number of matching Y-chromosomal loci. We see that the boxplots are similar for 0-14 matching Y-chromosomal loci. When the number of matching Y-chromosomal loci is larger than 14 we start to find
large outliers. The red line represents the mean number of matching autosomal alleles as a function of the number of matching Y-chromosomal loci. We see that it is approximately constant except for completely matching Y-chromosomal DNA profiles (17 loci). Again, it is possible that the large outliers represent pairs of closely related men, where in some cases there is a mutation on the Y-chromosome, but it is also possible that something else is causing the outliers.

![Figure 2.1](image)

**Figure 2.1:** Boxplots for the number of autosomal alleles that match between pairs of men, given the number of matching Y-chromosomal loci

The relevant forensic question is: what is the combined evidential value of matching Y-chromosomal and autosomal DNA profiles? Simple multiplication is not straightforward, because two men with matching Y-chromosomal DNA profiles may be more likely to have matching autosomal DNA profiles. The database provides some evidence to support the preceding statement but it is not very informative. The number of autosomal matches in a population of people that may be related to the suspect is something that can be easily investigated with a simulation model.

### 2.3 The simulation - method

We are interested in the combined evidential value of matching Y-chromosomal and autosomal DNA profiles when the population of alternative suspects contains an unknown number of relatives of the suspect. A conservative assumption (defined as an assumption that is in favor of the suspect) is that all men in this population are related to the suspect. If we assume no mutations (which is again a conservative assumption) then all these men share the Y-chromosome of the suspect. The autosomal match probability for an individual in this population is an upper bound for the autosomal match probability in more realistic alternative populations. Hence, our goal is to derive the autosomal match probability in a population consisting of relatives of the suspect. We will compare this with
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

the standard autosomal random match probability in an unrelated population. If the difference is negligible we can compute the likelihood ratio of the two profiles combined by multiplying the likelihood ratios of the individual profiles.

We construct such populations by simulating family trees that relate a given number \( n \) of men to the suspect. We assume that all the men in this tree share the same Y-chromosome. To simulate family trees, we assume that the number of sons each man fathers is binomially distributed with mean 1.3. This is slightly conservative in view of the world average 1.24 [1]. We take the last person that is added to the tree, the \( n \)th living person, as our suspect. This has no effect on our conclusion because we will focus on members of the population that are not descendants of the father of the suspect.

By assigning the suspect a randomly drawn (partial) autosomal DNA profile and by assuming Mendelian inheritance, we can assign (partial) autosomal DNA profiles to all the men in the family tree. The (partial) autosomal DNA profiles of the mothers in the tree are randomly drawn. The allele frequencies used are computed using the same database, containing 2085 men, as in Section 2.2. A detailed description of the simulation model can be found in the supplementary material. This can be found in the online version of this paper\(^1\).

We simulated the autosomal DNA profile assignment multiple times over the same tree, always starting with the same (partial) autosomal DNA profile for the suspect, to estimate the distribution of the number of matching profiles as well as the match probability in the family tree. This is done with a large number of different family trees. We compared the distributions of the number of matching profiles of each individual tree with each other. Furthermore, to investigate the influence of relatedness on the number of matches, we compare the distributions for the family trees with the distribution of an unrelated population (which is a binomial distribution).

2.4 The simulation - results

We carried out a large number of simulations. We present a restricted number of results here; more results can be found in the supplementary material\(^1\). The parameter settings of the situations that are presented here are given in Table 2.2. For each simulation, at least 100 different family trees were simulated. This way, we can examine the influence of the shape of the family tree on the results. We are interested in the differences between the distribution of the number of matching profiles in a related population (family tree) and in an unrelated population (binomial distribution).

Generally, we see that they differ most in the tails of the distributions: It is more likely to see a larger number of matching profiles as well as smaller number of matching profiles in related populations than in unrelated populations. This is to be expected since in a family tree, there is correlation between the autosomal DNA profiles.

\(^1\)doi:10.1007/s00414-014-0971-7
2.4. The simulation - results

<table>
<thead>
<tr>
<th>number of men</th>
<th>$\mathbb{E}[\text{sons}]$</th>
<th>number of autosomal loci</th>
<th>random match probability</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1.29</td>
<td>3</td>
<td>$1.75 \cdot 10^{-3}$</td>
<td>2.2</td>
</tr>
<tr>
<td>100</td>
<td>1.29</td>
<td>2</td>
<td>$2.69 \cdot 10^{-2}$</td>
<td>2.3</td>
</tr>
<tr>
<td>10</td>
<td>1.29</td>
<td>2</td>
<td>$2.69 \cdot 10^{-2}$</td>
<td>2.4</td>
</tr>
<tr>
<td>1000</td>
<td>2.58</td>
<td>3</td>
<td>$1.75 \cdot 10^{-3}$</td>
<td>2.5</td>
</tr>
<tr>
<td>1000</td>
<td>1.29</td>
<td>6</td>
<td>$5.18 \cdot 10^{-7}$</td>
<td>2.6</td>
</tr>
<tr>
<td>1000</td>
<td>1.29</td>
<td>1</td>
<td>$6.58 \cdot 10^{-4}$</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Table 2.2: Overview of the parameters we used in the simulations

For each situation in Table 2.2, we selected a tree whose distribution deviated most in the mean and tails with the binomial distribution corresponding to an unrelated population. The tree with the distribution that deviated most from the distribution for an unrelated population in the first situation given in Table 2.2 is given in blue in Figure 2.2. With these parameters, the expected number of matching profiles in an unrelated population would be 1.75. In this tree, the expected number of matching profiles is 2.35. It turns out that this difference is mainly due to the close relatives of the suspect. If we only consider the men that do not have the father of the suspect as an ancestor (so we do not omit the profiles of the father and the brothers of the suspect) then we get the distribution given in green in Figure 2.2. The mean number of matching profiles in this family tree, where one branch is “cut off”, is 1.77. This makes sense, and is typical for all the trees we simulated.

This suggests that it may be an option to change the alternative hypothesis that we consider from some other man is the donor of the crime stain to some other man, who is not a descendant of the father of the suspect, is the donor of the crime stain. This means that the focus shifts to the family tree with this branch cut off.

To compare the distributions, we compare the mean and the 95th percentile of the distribution of the trees and the binomial distribution, see Table 2.3. The mean and 95th percentile of the related population distributions grow compared to 95th percentile of unrelated populations when we decrease number of men in the alternative population. A smaller number of men in the alternative population is similar to increasing the proportion of men that is closely related to the suspect. A tree relating a billion men has a negligible number of men who are closely related to the suspect. The majority of men in such a tree have autosomal DNA profiles that can be regarded as independent of the autosomal DNA profile of the suspect, since their common ancestor is very distant.

Increasing the expected number of sons a man gets has a similar effect. The proportion of closely related men increases when the expected number of sons increases. The number of generations that is needed to relate a fixed number of men decreases. However, the effect of increasing the expected number of sons or decreasing the number of men in the alternative population is small. The mean number of matching autosomal as well and the 95th percentile of the distributions
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

![Graph showing the pdf of the number of matching autosomal DNA profiles](image)

**Figure 2.2:** pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability 0.0017, $E[\text{sons}] = 1.29$ and a partial autosomal DNA profile on 3 loci as the suspect profile

...do increase, but the difference is small, see Table 2.3. It is important to note that both the situation where the alternative suspect population consists of 10 men and the situation where the expected number of sons a man gets is 3.87 are unrealistic. In the simulations that represent realistic situations (i.e. a normal expected number of sons, realistic number of men in the alternative population) the difference between the mean number of matching autosomal DNA profiles in an unrelated and in a related population (without the men in the father branch of the tree) are very small. As said, the differences in distribution occur in the tails of the distribution, as can be seen in the 95th percentiles, but these 95th percentiles do not substantially differ.

More results and a sensitivity analysis on the different parameters can be found
2.5 Conclusion and Discussion

A test for independence for Y-chromosomal DNA profiles and autosomal loci on the Dutch database containing 2085 men yielded a significant result on 5 of the 23 autosomal loci. However, there are multiple possible explanations for this. For instance, it can be due to close relatives in the database.

Figure 2.3: pdf of the number of matching autosomal DNA profiles in a tree relating 100 men, random match probability 0.0269, E[sons] = 1.29 and a partial autosomal DNA profile on 2 loci as the suspect profile

in the supplementary material, which can be found at in the online version of this paper\(^2\).

\(^2\)doi:10.1007/s00414-014-0971-7
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

Our simulation model showed that although the distribution of the number of matching profiles for an unrelated and a related population are not the same, the expected numbers of matching profiles are very close to each other. Especially when we exclude some of the close family of the suspect (everybody that has the father of the suspect as an ancestor), our results indicate that the difference between the expected number of matching profiles in a related and an unrelated population is very small. The differences in the 95th percentiles of the distributions are more pronounced, but they remain small. The results show that we can regard autosomal and Y-chromosomal DNA profiles as independent when we exclude the relatives of the suspect that have the father of the suspect as an ancestor.

We suggest to compute the likelihood ratio, in a case where an autosomal
and a Y-chromosomal DNA profile are obtained from the same stain, using the hypotheses $H_p$: the suspect is the donor of the crime stain and $H_d$: some other man, who is not a descendant of the father of the suspect, is the donor of the crime stain. When doing so, the likelihood ratio can be computed by multiplying the individual likelihood ratios of the autosomal and Y-chromosomal DNA profiles.

As mentioned, in [19], Walsh et al. suggest to use $\theta = 0.04$ in the sampling formula of Balding and Nichols [7] to estimate the random match probability of the autosomal DNA profile. This value is multiplied with the estimated matching probability of the Y-chromosomal DNA profile to obtain the joint match probability. This approach will give larger joint random match probabilities than our approach. This is because Walsh et al. consider the alternative donor as a random
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

![Figure 2.6: pdf of the number of matching autosomal DNA profiles in a tree relating 1000 men, random match probability \(8.52 \cdot 10^{-5}\), \(E[\text{sons}] = 1.29\) and a partial autosomal DNA profile on 6 loci as the suspect profile](image)

We are aware of the scientific discussion on how to compute the random match probability of a Y-chromosomal DNA profile [8, 14, 3, 4, 17]. Whatever the outcome of this discussion may be, the results of this paper can still be used to combine the evidential value of the two profiles.
This paper focuses on the combined evidential value of the two DNA profiles. However, the conclusions from this paper are also relevant in a wider context; in reporting autosomal matches. Currently, these are reported with an alternative hypothesis stating that the crime stain was left by an unknown unrelated person. In a situation where a crime was committed in a small village, it is likely that the alternative suspect population contains many relatives of the suspect. In such a case, the conclusion that it is a million times more likely to find the DNA match when the suspect is the donor of the crime stain than when some unrelated unknown person was the donor has little influence. The judge has no idea what the actual likelihood ratio is since the alternative suspect population consists of relatives of the suspect. Now, we can use the conclusions from our simulation.
2. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

<table>
<thead>
<tr>
<th>number of men</th>
<th>E[sons]</th>
<th>number of loci</th>
<th>random match prob.</th>
<th>unrelated mean 95th perc.</th>
<th>related; all family trees mean 95th perc.</th>
<th>related; max over all trees mean 95th perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>1.29</td>
<td>3</td>
<td>1.75E-3</td>
<td>1.75 4</td>
<td>1.78 5</td>
<td>1.91 5</td>
</tr>
<tr>
<td>100</td>
<td>1.29</td>
<td>2</td>
<td>2.69E-2</td>
<td>2.69 6</td>
<td>2.79 7</td>
<td>3.05 8</td>
</tr>
<tr>
<td>10</td>
<td>1.29</td>
<td>2</td>
<td>2.69E-2</td>
<td>0.27 1</td>
<td>0.35 2</td>
<td>0.44 3</td>
</tr>
<tr>
<td>1000</td>
<td>2.58</td>
<td>3</td>
<td>1.75E-3</td>
<td>1.75 4</td>
<td>1.85 5</td>
<td>1.97 6</td>
</tr>
<tr>
<td>1000</td>
<td>1.29</td>
<td>6</td>
<td>5.18E-7</td>
<td>0.08 1</td>
<td>0.09 1</td>
<td>0.10 1</td>
</tr>
<tr>
<td>1000</td>
<td>1.29</td>
<td>1</td>
<td>6.58E-4</td>
<td>0.66 2</td>
<td>0.51 2</td>
<td>0.63 3</td>
</tr>
</tbody>
</table>

*Table 2.3:* Mean and 95th percentile in related and unrelated populations for the different simulations where the father branch is cut from the family tree

model and suggest a new alternative hypothesis $H_a$ some other person, who is not a descendant of the father of the suspect, is the donor of the crime stain and we can compute the likelihood ratio by multiplying the allele frequencies of the alleles in the DNA profile of the crime stain. However, we need to know the gender of the donor to do so.

**Bibliography**


BIBLIOGRAPHY


The interpretation of traces found on adhesive tapes

Abstract

In violent crimes, adhesive tapes such as duct tape are often used by perpetrators, for example to tie up a victim. In the forensic examination of such tapes many different types of traces can be found, such as finger marks and human biological traces. These traces are first interpreted at source level. However, even when it is certain that a trace was donated by the suspect this does not necessarily mean that he donated the trace while taping the victim, as he could have, for example, used the tape roll from which the pieces came previous to the crime. Therefore, the trace can also be interpreted at activity level. For this, factors such as transfer, persistence and recovery, as well as the position of the trace as it would have been on the original roll have to be taken into consideration. In this study, we have developed a Bayesian network which can aid the forensic practitioner in his interpretation. From a sensitivity analysis, we have concluded that it would be most desirable to set up further studies to determine the most likely positions of DNA on tape rolls if there has only been innocent contact.

3.1 Introduction

Perpetrators often use adhesive tape when committing violent crimes such as homicides, robberies, sexual assaults and terrorist attacks. Duct tapes are, for example, used to tie up a victim or to bind together parts of an improvised explosive device. Due to its adhesive qualities, many different types of traces can be found on tape pieces, like human biological traces, finger marks or fibres from garments. These traces can be analysed by forensic practitioners and, subsequently, compared to
The interpretation of traces found on adhesive tapes suspected sources. For example, a DNA profile obtained from a crime stain can be compared to the DNA profile obtained from a suspect.

However, even when it is absolutely certain that the suspect was the donor of the trace, this does not automatically mean that he\textsuperscript{1} donated the trace while taping the victim or object\textsuperscript{2}. It could, for example, be that he had contact with the tape roll previous to the crime. Source attribution alone cannot be used to resolve these issues. One level higher in the hierarchy of propositions ([5]) would be to interpret the evidence at activity level. Here, forensic practitioners want to determine how likely it is to see the evidence (under the case specific conditions) if it came there when the suspect taped the object compared to when an unknown person taped the object (i.e. when the suspect only used the tape for innocent purposes).

The factors which are used for the interpretation at activity level of traces found on adhesive tapes (namely transfer, persistence, recovery\textsuperscript{3} and the position of the trace as it would have been on the original tape roll) can subjectively be combined in a collaborative manner by forensic practitioners from multiple disciplines to come to a conclusion at activity level. However, a more transparent and uniform approach making use of Bayesian networks may be preferable ([16]). A Bayesian network is a mathematical tool in which all the relevant variables used in interpretation and the dependencies between these variables can be charted. Here, variables are represented by nodes, and the dependencies between the variables are represented by directed edges. Each node has a certain number of states. Conditional probability tables have to be provided for each node given its parent nodes\textsuperscript{4}.

In order to perform calculations in a Bayesian network, all of the subjective probability assignments which forensic practitioners might use subconsciously in their interpretation have to be made explicit. Now, the variables which most influence their conclusions can be determined by performing sensitivity analyses. Further studies can then be set up to collect data to support the probabilities used in these nodes.

In this article, a Bayesian network is presented which can assist the forensic practitioner in the interpretation of a single human biological trace. Many of the terms that are used can be found in [15]. However, upfront some important terms will be explicitly stated.

1. A **human biological trace** (or **trace** for short) is human DNA that is present on a specific location on a secured\textsuperscript{5} piece of evidence (such as a tape

---

\textsuperscript{1}Throughout this article, male pronouns are used to refer to individuals to provide for a more fluent reading.

\textsuperscript{2}The word victim may imply that a crime was committed. As only interpretation at activity level is considered (and not interpretation at crime level), the use of the word “victim” is avoided. Instead, throughout the remainder of this article, the word “object” is used to refer to any taped object, such as an improvised explosive device or a person/victim.

\textsuperscript{3}For a more thorough discussion on the role of transfer, persistence and recovery for the interpretation at activity level in cases involving minimal quantities of DNA, see [6].

\textsuperscript{4}A node A is called a parent node of node B if there is a directed edge from node A to node B.

\textsuperscript{5}A “secured” piece of evidence is a piece of evidence that lies safely stored at the forensic laboratory and is ready for examination.
3.2 The Bayesian network

The developed network for the interpretation at activity level of a single human biological trace that was found on adhesive tape is presented in Figure 3.1. The factors that have to be taken into consideration by forensic practitioners when
they interpret the results of their analyses at activity level can be divided in three
groups; the process by which the trace is formed (Sections 3.2.1, 3.2.2 and 3.2.3),
the origin of the DNA (Sections 3.2.4 and 3.2.5), and the manner of deposition
(Section 3.2.6). Per factor, the ways in which the factor can be modelled in a
Bayesian network are outlined and it is discussed which way of modelling has our
preference. Nodes are indicated by CAPITALS and states by italicics.

3.2.1 Transfer

With transfer, the transfer of DNA from a certain person to a certain area on the
tape pieces (which will later turn out to be the sampling location) is meant. Transfer
depends on a number of factors, including (but not limited to) the duration,
intensity and frequency of contact.

Three different kinds of transfer scenarios are distinguished: (1) transfer from
the perpetrator\(^6\) to the sampling location while taping the object, (2) background
transfer from the suspect to the sampling location, and (3) background transfer
from unknown(s) to the sampling location. Here, it is assumed that there is only
a single perpetrator. As these three types of transfer are not mutually exclusive,
a mixture of DNA can be formed. DNA has transferred to the sampling location as background if it came there due to some innocent reason (i.e. if the DNA
was not transferred from the perpetrator to the sampling location while taping
the object). For instance, one can think of background material of a person being
present because that person touched the tape at some moment before or after the
crime, or because material was retransferred from an object previously touched by
that person to the tape before, during or after the crime (i.e. secondary transfer).

In the network (Figure 3.1), the probabilities of transfer under the three above-
mentioned scenarios have to be estimated by a forensic practitioner\(^7\) based on the
case circumstances\(^8\). As there is a large number of variables that needs to be taken

\(^6\)The word “perpetrator” is often used to refer to the person who performed the activity
(i.e. taped the object). However, when someone taped a person this does not necessarily mean
that a crime was committed. It should thus be noted that interpretation at crime level is not
considered; the word “perpetrator” is strictly used for shortness of notation. Similarly, the word
“crime” refers to the incident that took place.

\(^7\)In order to work with the proposed network, prior probabilities have to be estimated. The
forensic practitioner is sometimes more in a position to do so than the trier of fact (i.e. the
judges and/or jury), for example when estimating the probabilities of transfer ([4]), persistence,
recovery and the efficiency of the DNA analysis. However, when one has to estimate whether
a suspect’s statement is true, this would preferably be done in consultation with the trier of
fact, as this is outside the field of expertise of the forensic practitioner. As the trier of fact
can only provide feedback during the actual hearing of a case, it would be impossible for the
forensic practitioner to take his assessment into consideration when performing calculations in
the network. Therefore, one could instead first assume that the suspect’s story is true, and then
that his story is false. The effect of this on the likelihood ratio at activity level can then be
determined and reported. Another solution would be to calculate and report a list of likelihood
ratios for a number of different probability values. In consultation with the trier of fact, a number
can then be selected from this list.

\(^8\)The amount of DNA in the trace should not be taken into consideration when estimating
the probabilities of transfer, persistence, recovery and the efficiency of the DNA analysis in the
network. [6] discussed the issues which can arise when one wants to take into account the amount
of DNA. A possible solution to these issues was also presented in their article.
3.2. The Bayesian network

into consideration when determining the probabilities for these three scenarios, it was opted not to model all the variables in the network, especially since the probabilities are mostly determined subjectively. It would become challenging to objectively determine the necessary probability tables. Therefore, a transfer node was created for each of the three scenarios, which has states Yes and No. The probabilities that have been subjectively estimated by a forensic practitioner can then be entered into the probability tables of these nodes. Depending on the needs of the case, experiments could be set up to obtain (more) data to support these probabilities (see also [4]).

For the first scenario, the forensic practitioner needs to estimate the probability that DNA of the perpetrator transferred to the sampling location when taping the object. In order to do this, the type of sampling location has to be taken into consideration (e.g. bite mark or tape end). Specific case circumstances also have to be taken into account. For example, in case a person was taped, he might state that the perpetrator wore gloves while taping (which suggests a lower probability of transfer).

For the second scenario, the forensic practitioner needs to estimate the probability that background DNA of the suspect transferred to the sampling location. Background transfer of DNA of the suspect to the tape pieces in general depends on his relation to the tape roll. For example, if the roll under consideration was found at his house and the suspect states that he used that roll regularly, then background transfer is likely. Here, the probability that the suspect’s statement is true has to be taken into account. Also, retransfer of DNA from another object to the sampling location has to be taken into consideration.

As the forensic practitioner wants to estimate the probability of background transfer to a specific sampling area, he needs to scale the total probability of background transfer to the tape accordingly. Here, the size of the sampling area compared to the whole secured tape has to be taken into consideration, as well as the proportion of the tape to which the suspect likely transferred DNA. The location of the sampling area on the pieces itself should not be taken into account, but only the relative size of the sampling area. From Section 3.2.3 and onwards, two different parts will be studied, namely a part in which the composition of the sample is considered (Sections 3.2.4 and 3.2.5), and a part in which the manner of deposition is considered (Section 3.2.6). The position of the sampling area on the original roll is only taken into account in the latter part.

For the third scenario, the forensic practitioner needs to estimate the probability that background DNA of unknown(s) transferred to the sampling location. Here, an unknown person can be any person other than the suspect who had direct or indirect contact with the tape, where the forensic practitioner should not take into account people whose reference DNA profiles are known (such as the person that was taped and forensic examiners) and who can thus later be excluded from a DNA profile. Transfer mostly depends on the number of people who could have had contact with the tape, which can be estimated from case circumstances. Data on background transfer may be collected from case files [12]. Secondary transfer scenarios should be taken into account, as DNA could have been present on the taped object before the crime, or DNA could have been retransferred from the...
3. The interpretation of traces found on adhesive tapes

hands/gloves of the perpetrator to the tape. Just as in the second scenario, the size of the sampling area has to be taken into consideration, as the forensic practitioner wants to determine the probability of transfer to the specific sampling area. The actual position of this area on the original roll of this area should, again, not be used.

In all scenarios, retransfer of DNA from one location on the tape pieces to another should not be taken into consideration (see the discussion, Section 3.3).

3.2.2 Persistence

After DNA has been transferred, it will need to persist will forensic examiners be able to find and recover it from the sampling location. Here, two factors have to be taken into consideration: (1) the degradation of DNA due to environmental conditions, and (2) the retransfer of DNA from the sampling location to another object.

When cellular material is exposed to the environment, DNA present in this material will degrade over time. The rate of decay depends on the specific environmental conditions in which the material resided. DNA will degrade faster in a warm, humid environment and in the presence of UV light than in a cold, dry environment in the absence of UV light ([15]). Whether DNA is still present in a trace on a secured tape piece thus (mostly) depends on the surroundings of the tape. Consideration has to be given to the fact that DNA will not degrade as fast if it is covered by another item, but not to the fact that background DNA could originally have been present at a deeper layer of the roll, and thus covered by other layers of tape on the roll. This is taken into account in the manner of deposition node (see Section 3.2.6).

Re-Transfer of DNA to another object (i.e. the loss of DNA) can occur at any time from when the DNA was deposited up to the moment that the evidence is secured. If an object comes into contact with the tape, DNA can transfer from the tape to that object and, thus, not be present on the tape during forensic examination. An example of this is displacement and retransfer of DNA within forensic packaging [7]. As stated in Section 3.2.1, retransfer from one location on the tape pieces to another should not be taken into consideration (see the discussion, Section 3.3).

In order to use the network (Figure 3.1), the forensic practitioner again has to estimate probabilities for three different scenarios: (1) the persistence of DNA of the perpetrator (that was transferred when taping the object), (2) the persistence of background DNA of the suspect, and (3) the persistence of background DNA of unknown(s). Here, the forensic practitioner only estimates the probability of persistence given that DNA has transferred; when no DNA was transferred, there will be no persistence. Important factors to take into consideration are the time span since transfer for the three scenarios, and the extent to which the item has been handled during that time. If the suspect claims that he had background contact with the roll six months ago and the perpetrator taped the object three days ago, one would expect that if DNA transferred in both scenarios it would persist more likely in the latter, as there would be less chance of degradation.
and/or retransfer.

3.2.3 Composition of the trace and manner of deposition

The three PERSISTENCE nodes are connected to the COMPOSITION OF THE TRACE AND MANNER OF DEPOSITION node (Figure 3.1). This node is also called the SUMMARY node, as it summarizes whose DNA is present at the sampling location and how that DNA transferred to the sampling location. PROPOSITIONS (which contains the propositions under consideration at activity level) is also connected to SUMMARY. This node states who the perpetrator is (the suspect or an unknown person), which is important for the PERSISTENCE PERPETRATOR node.

SUMMARY has 12 mutually exclusive and exhaustive states. These states are displayed in Table 3.1.

<table>
<thead>
<tr>
<th>PROPOSITIONS</th>
<th>Hp</th>
<th></th>
<th>Hd</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERSISTENCE PERPETRATOR</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>PERSISTENCE BACKGROUND SUSPECT</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>PERSISTENCE BACKGROUND UNKNOWN(s)</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>XAB, UB</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XAB</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>XA, UB</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>XA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UAB, XB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UA, XB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UAB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XB, UB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UB</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No DNA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1: Conditional probability table for the SUMMARY node. Here, X= suspect, U = unknown person, A= activity, and B = background. Background DNA from unknown(s) can persist or not, background DNA from the suspect can persist or not, and DNA from the perpetrator (that was transferred when he taped the object) can persist or not. When DNA from the perpetrator has persisted, the particular state in which SUMMARY is depends on the state of PROPOSITIONS. However, if PERSISTENCE PERPETRATOR is in state No, then the state of PROPOSITIONS does not matter. If all three PERSISTENCE nodes are in state No, then there is no DNA present at the sampling location.
3.2.4 Recovery

To be able to perform DNA analysis, forensic examiners first need to recover the DNA that is present at the sampling location. For the collection of DNA, several methods exist, such as tape lifting ([18]) and swabbing. As these methods are not 100% successful, there exists uncertainty whether sampling will succeed. In this article, there is no focus on contamination during recovery (or DNA analysis). For this, the reader is referred to the supplementary material.9

In the network (Figure 3.1), the composition of the sample node is dependent on the recovery and summary nodes. Composition of the sample has four states:

1. DNA of suspect in sample
2. DNA of unknown(s) in sample
3. DNA of suspect and unknown(s) in sample
4. No DNA present in sample

When recovery is in No, then there is no DNA present in the sample (independent of the state of the summary node). The state of summary only matters when DNA was recovered. The network currently only focuses on the “who” component of summary and not at “how” the DNA was originally deposited (i.e. as background, by the perpetrator when he was taping the object, or both). The latter is taken into consideration in manner of deposition (see Section 3.2.6).

When recovery was successful, but no actual DNA was present in the trace, then composition of the sample is in No DNA present in sample. The meaning of this is that the recovery method would have been effective if DNA was been present, but since no DNA is present there will be no DNA in the sample.

Recovery is not meant to be instantiated; if in the end no DNA profile was obtained from the sample, this will allow the forensic practitioner to reason that either no DNA was present at the sampling location or that the recovery method must have been unsuccessful. If, for example, the forensic practitioner expects that in 80% of the cases forensic examiners will be able to collect DNA if it is present at the sampling location, then the prior probability of recovery can be set to 0.8. This probability can be changed on a case by case basis, where the forensic practitioner could take into account the efficiency of the specific collection technique that was used and the effectiveness of (specific) examiners.

3.2.5 DNA analysis

For interpretation at activity level, it is pragmatic to assume a known source. However, with the Bayesian network it is possible to take the likelihood ratio at (sub-)source level into account. When a forensic practitioner is presented with a single source DNA profile (i.e. a profile only containing alleles from one source), he generally addresses the following propositions at sub-source level:

\[ H_p: \text{The suspect is the source of the DNA.} \]
3.2. The Bayesian network

\( H_d: \) An unknown person is the source of the DNA.

For mixed DNA profiles from which the suspect cannot be excluded, the following propositions at sub-source level are addressed:

\( H_p: \) Unknowns \( U_1, \ldots, U_{n-1} \) and suspect \( X \) are the source of the DNA.

\( H_d: \) Unknowns \( U_1, \ldots, U_n \) are the source of the DNA.

If the obtained evidence (i.e. the DNA profile) is denoted by \( E \), then the forensic practitioner would have to determine \( P(E \mid H_p) \) and \( P(E \mid H_d) \), i.e. the probabilities of observing the profile when the suspect was a contributor and when he was not a contributor. These probabilities are usually calculated by using one of the many software tools available for mixture calculations ([2, 8, 9, 13, 14, 17]).

In the network (Figure 3.1), the DNA PROFILE node is dependent on the COMPOSITION OF THE SAMPLE and DNA ANALYSIS SUCCESSFUL? nodes. DNA PROFILE has the following states:

1. Suspect
2. Unknown(s)
3. Suspect and unknown(s)
4. No useable profile

DNA ANALYSIS SUCCESSFUL? has states Yes and No and is not meant to be instantiated. When DNA ANALYSIS SUCCESSFUL? is in state No, then DNA PROFILE is in state No useable profile. Otherwise, DNA PROFILE is in the same state as the COMPOSITION OF THE SAMPLE node. A prior probability for DNA ANALYSIS SUCCESSFUL? has to be estimated by the forensic practitioner in each case, which could, for example, be based on the efficiency of the specific analysis equipment that was used.

Furthermore, the UNCERTAINTY SOURCE ATTRIBUTION node is dependent on the DNA PROFILE node. This node can be used to enter the obtained likelihood ratio at source level. UNCERTAINTY SOURCE ATTRIBUTION has states Match and No match\(^\text{10}\). In tables 3.2 and 3.3, the conditional probability tables for UNCERTAINTY SOURCE ATTRIBUTION when a single source profile and a mixed profile DNA were obtained can be found. Here, \( P(E \mid H_p) \) and \( P(E \mid H_d) \) are as described above.

3.2.6 The position of the sampling location on the original roll

For interpretation at activity level, the position of the sampling location on the original roll can be used. After estimating the circumference of the roll and determining the order in which the tape pieces most likely came from the roll (Sections

\(^\text{10}\)To perform calculations in the network when using software packages such as AgenaRisk [1] and HUGIN [10], UNCERTAINTY SOURCE ATTRIBUTION has to be fixed in state Match. In this manner, the DNA PROFILE node will be in state Suspect (or Suspect and unknown(s), for a mixture) respectively Unknown(s) relative to the probabilities that were inserted for \( P(E \mid H_p) \) and \( P(E \mid H_d) \). As prior probabilities were inserted in the TRANSFER, PERSISTENCE, RECOVERY and UNCERTAINTY SOURCE ATTRIBUTION nodes, the DNA PROFILE node will already have a prior distribution. These probabilities may also have an influence on the specific probabilities with which DNA PROFILE is in its different states.
3. The interpretation of traces found on adhesive tapes

### Table 3.2: Conditional probability table of uncertainty source attribution for a single source DNA profile.

<table>
<thead>
<tr>
<th>DNA PROFILE</th>
<th>Suspect</th>
<th>Unknown(s)</th>
<th>Suspect and unknown(s)</th>
<th>No useable profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>( P(E \mid H_p) )</td>
<td>( P(E \mid H_d) )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No match</td>
<td>1 - ( P(E \mid H_p) )</td>
<td>1 - ( P(E \mid H_d) )</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3.3: Conditional probability table of uncertainty source attribution for a mixed DNA profile.

<table>
<thead>
<tr>
<th>DNA PROFILE</th>
<th>Suspect</th>
<th>Unknown(s)</th>
<th>Suspect and unknown(s)</th>
<th>No useable profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>0</td>
<td>( P(E \mid H_d) )</td>
<td>( P(E \mid H_p) )</td>
<td>0</td>
</tr>
<tr>
<td>No match</td>
<td>1</td>
<td>1 - ( P(E \mid H_d) )</td>
<td>1 - ( P(E \mid H_p) )</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.6 and 3.2.6), it can be determined how likely it is that the sampling location originally would have been on the outside layer of the roll or at a deeper layer (Section 3.2.6).

When someone previously only used the roll for innocent reasons, one would expect that DNA is left behind mostly on the outside of the roll (i.e. the non-adhesive backing of the outer layer of the roll). For example, the suspect can state that his DNA could indeed be present on the tape, but that it was originally left behind when he used the tape in household chores (and that someone else used his roll to tape the object). If a DNA profile matching the DNA profile of the suspect is then obtained from a sampling location which would originally have been a couple of layers deep into the roll, then it is unlikely that his DNA would have ended up there according to his statement. It is more likely that it ended up there as stated by the prosecutor.

**Using physical end matching to determine the order in which the tape pieces came from the roll**

To be able to determine the position of the sampling location on the original roll, forensic practitioners will first have to determine the order in which the tape pieces came from the roll. This is done by a process called physical end matching ([3], [11]). First, the colour and thickness of both the adhesive layer and backing layer of the different tape pieces are compared. Markings and striations can also be examined and compared for the different pieces. After this, the tape ends of the different tape pieces are compared morphologically to determine whether and how the pieces fit each other complementary.

After comparison, the examiner (also called physical end matcher) gives a conclusion on his findings, such as “It is extremely more probable to see these characteristics when the tape pieces originally formed a whole in this specific order.
than when they formed a whole in a different order or did not connect.” In the network, the very small probability of a false negative/false positive\textsuperscript{11} is neglected.

It is assumed that all the found tape pieces can be reassembled in one group and not multiple groups, though extensions to the network can easily be made to deal with multiple groups (see the supplementary material\textsuperscript{12}).

**Determining the position of the sampling location on the original roll**

After determining the order in which the tape pieces came from the roll, physical end matchers try to determine the position of the tape roll from which the pieces came respective to this order. When a tape roll is available (for example, if it was found at the suspect’s house or at the crime scene), physical end matchers can try to match the tape end of that roll to one of the ends of the determined order. If this succeeds they can proceed with a known roll position.

Even in situations where there is no roll available or where it is not possible to connect the obtained pieces to an available roll because pieces are missing, it might still be possible to determine the roll position. When the tape has a fabric layer made of yarn, it is often possible to determine the position of the roll based on the position of the warp yarns (yarns in the direction of the length of the tape) respective to the fill yarns (yarns in the direction of the width of the tape).

As can be seen in Figure 3.2, the distance from the sampling location to the roll end is determined by reassembling the tape pieces in the determined order and by measuring the distance from one end of the order to the sampling area. Depending on whether the roll was positioned on the left end or on the right end of the determined order, different distances from the sampling location to the roll end will be obtained. Therefore, different conclusions may also be drawn.

In the network (Figure 3.1), the total length of the pieces and the distance from the sampling location to the left end of the determined order can be inserted in the nodes **TOTAL LENGTH OF THE PIECES** and **DISTANCE FROM SAMPLING LOCATION TO LEFT END**. These nodes allow for an input on a continuous interval. Here, what is left and what is right is arbitrary: what matters is that one is consistent in using these terms. When a physical match was found between one of the tape ends of the determined order and the roll end, the node **ROLL POSITION** can be set to state **Left** or **Right**, depending on where the roll connects to the pieces. When the roll does not match to one of the ends or when no roll is available, the probability table of **ROLL POSITION** can be altered depending on how certain the forensic practitioner is that the roll was originally connected to either side.

\textsuperscript{11}To cope with the uncertainty of the process of physical end matching, one option would be to weigh this uncertainty when calculating the likelihood ratio at activity level. If, for example, the practitioner is 99.9% certain that the pieces formed a whole in the determined order, then he could weigh the likelihood ratio at activity level corresponding to this order 999 times and the likelihood ratio at activity level corresponding to the situation where he does not have an order one time. When an order cannot be determined, then the likelihood ratio at activity level can be determined by only considering the tape piece on which the trace was found, i.e. disregard the other pieces and calculate the likelihood ratio by using the distance from the trace to the end of the piece on which the sampling location was positioned, and the total length of that piece.

\textsuperscript{12}doi:10.1093/lpr/mgv012
3. The interpretation of traces found on adhesive tapes

![Diagram of adhesive tape and roll](image)

**Figure 3.2:** The position of the roll respective to the order as determined by physical end matchers. The distance of interest is the distance from the sampling location (the highlighted area) to the roll end. These distances are represented by black arrows.

When the position of the roll respective to the tape pieces is uncertain, two distances are calculated, each with their own probability. When **ROLL POSITION** is in **Left**, then **DISTANCE FROM SAMPLING LOCATION TO LEFT END** is equal to **TOTAL LENGTH OF THE PIECES** minus **DISTANCE FROM SAMPLING LOCATION TO ROLL END**. When **ROLL POSITION** is in **Right**, then **DISTANCE FROM SAMPLING LOCATION TO LEFT END** is equal to **DISTANCE FROM SAMPLING LOCATION TO ROLL END**. Hence, **DISTANCE FROM SAMPLING LOCATION TO ROLL END** can easily be determined from the states of the other nodes.

To determine from the distance from the sampling location to the roll end whether the sampling location would have been present on the outside of the original roll (i.e. the backing of the outside layer of the roll) or at a deeper layer of the roll, the circumference of the original roll is needed. This is calculated by multiplying the diameter of the roll (as it was before the activity) by \( \pi \). The diameter of the original tape roll can be deduced from the diameter of the found roll by adding to this diameter the number of layers used to tie up the object times the thickness of each layer. Here, it is assumed that the obtained tape pieces can be physically matched to the roll. When no roll is available, the forensic practitioner could opt to use the largest possible diameter of commercially sold tape rolls (which is 16 cm). This is to give the suspect the benefit of the doubt, as in this case the distance from the sampling location to the roll end will need to be larger for this sampling location to end up at a deeper layer of the roll. However, this might be too conservative in most cases. When no roll was found at the crime scene or at the suspect’s house, it is still possible to deduce the most likely diameter of the roll by visually and chemically comparing the found pieces to rolls of known types and brands. When multiple diameters are possible, an average can be taken, or the LR at activity level can be calculated and reported for each diameter separately.
3.2. The Bayesian network

Deducing from the position of the sampling location on the original roll whether DNA present at this location ended up there as background and/or by the perpetrator when taping the object

In Sections 3.2.1 and 3.2.2, it was stated that the forensic practitioner should estimate the probabilities of transfer and persistence of DNA without taking into consideration the position of the sampling location on the original roll. With the position of the sampling location now known, the prior probabilities of transfer and persistence can be updated.

In the network (Figure 3.1), the MANNER OF DEPOSITION node is dependent on the SUMMARY node. MANNER OF DEPOSITION states how DNA that is present at the sampling location was originally deposited. It has the following states:

1. Background
2. By the perpetrator when taping the object
3. Background and by the perpetrator when taping the object
4. No DNA present

The state of this node is determined by the state of SUMMARY. MANNER OF DEPOSITION is in turn connected to DISTANCE FROM SAMPLING LOCATION TO ROLL END. For the conditional probability table of DISTANCE FROM SAMPLING LOCATION TO ROLL END, it is necessary to know the most likely distances from the sampling location to the roll end given the different states of MANNER OF DEPOSITION. For this, information is needed regarding where on the original roll one would be able to find DNA if it is known that it was deposited as background, by the perpetrator when he was taping the object, or both.

Now, a number of different conditional probability distributions for the DISTANCE FROM SAMPLING LOCATION TO ROLL END node is presented. It has to be stressed that currently no data is available on transfer behaviour; it is unknown where DNA can end up on the roll. Hence, the proposed probability distributions are for illustration purposes only. However, it does seem logical to assume that the probability distributions which are based on data will be of a similar shape as the ones presented here.

Probability distribution regarding the position of the DNA containing sampling location if DNA came upon the roll when the perpetrator was taping the object

When DNA of the perpetrator is present, it intuitively makes sense to assume that this DNA will be distributed uniformly to the roll (see the horizontal red line in Figure 3.3). Although DNA will be transferred more often to tape ends and bite marks, this was already taken into consideration in the TRANSFER nodes (see Section 3.2.1) and this should not be used again here to avoid using this observation twice.
3. The interpretation of traces found on adhesive tapes

The probability distributions for the location of background DNA on the original roll which are presented here (mainly) depend on how this roll was previously used by the suspect and unknown(s). Here, for simplicity it is assumed that a tape roll was secured (for example from the crime scene or from the suspect’s house) and that the tape pieces can be physically matched to this roll. For the moment, it is assumed that a single source DNA profile matching the DNA profile of the suspect is obtained.

When establishing the suspect’s relation to the roll, the statement he has given

**Figure 3.3:** Suggestions for probability distributions regarding the position of DNA on the original roll.

**Examples of probability distributions regarding the position of the DNA containing sampling location when the DNA came upon the roll as background**
3.2. The Bayesian network

to the police can be taken into account. When the suspect states that he previously only touched the roll and has not used it, then forensic practitioners would (very likely) only find his DNA on the backing of the outside layer of the roll. However, when the suspect previously used the roll, then his DNA could end up at other locations. For example, his DNA could be present on the adhesive side of the roll end from when he tore off a piece at some moment before the incident. Also, as he could roll out a length of tape, touch the tape, but then decide to roll the tape back up again instead of tearing it, background DNA can be left behind at deeper layers of the roll (however, the probability that this occurred should be taken into account).

The suggested probability distributions are of the form as displayed in Figure 3.3. For background DNA found on the backing of the tape at a distance smaller than $\pi$ times the diameter (i.e. the circumference of the roll) a uniform probability distribution is assumed: if DNA is present as background, it is assumed that it can end up everywhere on the outside of the roll with the same probability. However, for distances greater than $\pi$ times the diameter (i.e. at a deeper layer of the roll), it is assumed that the probability of finding DNA will decrease drastically as distance increases. One could then, for example, assume that the distribution follows a log-normal distribution, since log-normal distributions have this property. Other distribution functions (such as a step-wise decreasing function) could also be chosen for the same purposes. The crucial point here is that a probability distribution is assumed for which the probability of finding DNA decreases drastically as the distance from the sampling location to the roll end increases. Data needs to be gathered to determine how background DNA is distributed over the roll (see Section 3.2.7 and the discussion, Section 3.3).

If the suspect states that he previously only touched the outside of the roll, this suggests that the majority of the probability mass of the distribution should be underneath the uniform part of the probability distribution. However, if he previously used the roll, one could decide to move part of the mass towards the decreasing (log-normal) part of the distribution. For example, it can be assumed that 99% of the density lies underneath the uniform part and 1% under the log-normal part in the case that the suspect has only previously touched the roll, and that 95% lies underneath the uniform part and 5% under the log-normal part in the case that the suspect has used the roll before. Again, it is important to stress that these numbers are not based on data. For illustration purposes, the same numbers are used throughout the remainder of this article.

For the adhesive side of the tape, the distribution does not follow a uniform distribution for the first layer of tape. The distribution immediately follows a decreasing distribution, as the sampling location is always located at a “deeper layer” (the adhesive side is not readily accessible to someone touching the outside of the roll). Hence, the distribution is independent of the diameter of the roll (Figure 3.3).

When using the network (Figure 3.1), the practitioner first needs to “generate” the distribution fitting his specific case (based on the diameter of the roll, whether the sampling location was positioned on the backing or the adhesive side, and what the suspect’s relation is to the roll). This distribution can then be used as
3. The interpretation of traces found on adhesive tapes

a conditional probability table of **DISTANCE FROM SAMPLING LOCATION TO ROLL END**. This is further described in the supplementary material\(^ {13}\). In the network, nodes have been added so that the forensic practitioner can store distributions for common case circumstances. These nodes are **BACKING OR ADHESIVE SIDE?**, **ROLL DIAMETER** and **DISTRIBUTION**.

After the case specific distribution has been determined, the states of **ROLL POSITION**, **TOTAL LENGTH OF THE PIECES**, and **DISTANCE FROM SAMPLING LOCATION TO LEFT END** can be entered. Software packages like AgenaRisk [1] and HUGIN [10] can then assist in the calculation of the likelihood ratio at activity level.

There are a number of subtleties that were ignored when describing the suggested probability distributions. Firstly, only *transfer* of DNA to the roll was considered and not the persistence of DNA. DNA that was transferred to a deeper layer of the roll is protected from the environment by more superficial layers. The forensic practitioner would thus have to take into account that DNA on the outside layer of the roll will less likely persist than DNA that is deeper into the roll. This can be done by increasing the percentage of the density that lies underneath the decreasing (log-normal) parts of these distributions.

Secondly, it is assumed that the statement that the suspect has given to the police regarding his relation to the roll is true. As can be seen from the above distributions, when the suspect has previously used the roll this will lead to a less steep distribution, and therefore it would be beneficial for the suspect to state that he previously used the roll. Therefore, the prior probability that his statement is true should be taken into account when generating the distribution.

Thirdly, as there is uncertainty with source attribution, there is a probability that an unknown person left behind the background DNA instead of the suspect. When determining the probability distribution, the forensic practitioner would thus also have to take into account how an unknown person could have previously used the roll. However, as the likelihood ratio at source level is often extremely large (larger than one billion) this will only have marginal effect on the distributions described above. One could, therefore, choose to ignore this and only take the suspect’s previous usage into account.

Lastly, it has been assumed that a tape roll was obtained and that the tape pieces can be physically matched to this roll. However, when no roll was found it will not be possible to determine the relation the suspect had to the roll that was used during the crime. When the relation of the suspect to the roll is uncertain, this has to be taken into consideration when generating the probability distribution for the location of background DNA on the original roll. One could, for example, make the distribution less steep to be more in favour of the suspect.

\(^ {13}\) doi:10.1093/lpr/mgv012
3.2. The Bayesian network

Probability distribution regarding the position of the DNA containing sampling location when the DNA came upon the roll as background and by the perpetrator when he was taping the object

When DNA came upon the roll as background and by the perpetrator, the probability distribution for background transfer can be multiplied with the probability distribution for transfer of the perpetrator, as DNA would both had to transfer to a specific location on the tape pieces as background and by the perpetrator when he was taping the object. Thus, if it is assumed that background transfer and transfer by the perpetrator are independent, the combined probability distribution is the product of the individual distributions. Since the distributions are defined on discrete intervals (see the supplementary material\(^{14}\)), one can simply multiply the probabilities for each interval separately. However, as the probability distribution for the perpetrator is uniform, the result will be the same probability distribution as for background alone after normalizing the probability density function.

Probability distribution regarding the position of the sampling location when there is no DNA present

Finally, a distribution should be decided upon regarding where to sample given that there is no actual DNA present on the roll. This is relevant, as there generally is a non-zero probability that the MANNER OF DEPOSITION node is in state No DNA. This probability is equal to the probability that all three PERSISTENCE nodes are in state No.

However, when it is known that no DNA is present, there is no reason to pick a specific sampling location. Therefore, a uniform distribution for the position of the sampling location given that there is no DNA present on the roll is assumed.

3.2.7 Sensitivity analyses

In this section, part of the results of sensitivity analyses that have been performed are presented to illustrate which variables have a substantial influence on the likelihood ratio at activity level. Due to the potential effect that the probability assignments of these nodes have on the outcome, it is necessary to further examine these values, for example by gathering experimental data to support them.

First of all, the results of sensitivity analyses performed on two of the TRANSFER and PERSISTENCE nodes are presented. The numbers behind one variable are varied while the other nodes are fixed in a predetermined set of values. The results of the sensitivity analyses are displayed in Figure 3.4. As can be seen, the likelihood ratio (virtually) fluctuates within two orders of magnitude depending on the probabilities of transfer and persistence that are entered in the network. The likelihood ratio fluctuates similarly for the other transfer and PERSISTENCE nodes. In order to obtain a better estimate of the likelihood ratio at activity level, case specific experiments could be set up (see also [4]).

\(^{14}\)doi:10.1093/lpr/mgv012
Secondly, the influence of the probability distribution for the position of background DNA on the original roll has been determined. As noted in Section 3.2.6, the example distributions are not based on data. The numbers 95% vs. 5% and 99% vs. 1% were chosen for illustration purposes. To demonstrate how the likelihood ratio can vary, it is calculated for different distributions (while keeping the other probability assessments in the network fixed in a predetermined set of values).

As can be seen in Table 3.4, when the distribution is altered, the value of the likelihood ratio fluctuates within one order of magnitude. However, it is when the above factors are changed together that these numbers will change drastically. For example, if the probability of transfer of DNA of the perpetrator is set to 0.6, the probability of persistence of the perpetrator to 0.8, and if the probability
distribution for the location of background DNA on the original roll is taken to be a 95% vs. 5% type distribution, then a likelihood ratio of 215 is obtained (where the other variables are fixed in a predetermined set of values). If the probability of transfer of DNA is changed from 0.6 to 0.99, the probability of persistence of the perpetrator from 0.8 to 0.99, and the distribution from 95% vs. 1% to 99% vs. 1%, then the likelihood ratio at activity level becomes over 25 000.

It is uncertain what the distribution for the position of background DNA actually looks like, thus it may very well be that when someone previously used the roll this distribution could be 99% vs. 1% instead of the 95% vs. 5% which was assumed for illustration purposes in Section 3.2.6. In our opinion, it would be most desirable to gather data for the probability distribution for the position of background DNA on the original roll.

<table>
<thead>
<tr>
<th>% under uniform part vs. % under log-normal part</th>
<th>LR at activity level</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% vs. 5%</td>
<td>215</td>
</tr>
<tr>
<td>99% vs. 1%</td>
<td>560</td>
</tr>
<tr>
<td>99.9% vs. 0.1%</td>
<td>850</td>
</tr>
</tbody>
</table>

Table 3.4: The influence of the steepness of the distribution on the LR at activity level.

3.3 Discussion and conclusion

A Bayesian network which can assist forensic practitioners in the interpretation of human biological traces that can be found on adhesive tapes at activity level was presented. While the focus of this paper lies on human biological traces, similar networks have been developed for other types of transfer traces, such as finger marks. By combining these networks, networks are obtained for multiple traces. These networks are described in the supplementary material15.

The methodology described throughout this article can be used in casework. In a specific case, sensitivity analyses should be performed. The main reason for this is that the probability distributions for the position of background DNA on the original roll used in this paper are not based on data, and the likelihood ratio can fluctuate depending on the specific probability distribution that was chosen (see Section 3.2.7). Intuitively, it makes sense to assume that DNA will be found mostly on the outside layer of the roll if a person only used the roll for innocent purposes, but data has to be gathered to be able to assign more specific probabilities to certain locations. When gathering experimental data on where one would be able to find DNA on the original roll, one would also have to take into consideration retransfer of DNA from one location on the tape pieces to another. As mentioned throughout this article, not much is known about retransfer behaviour of DNA, and experiments should be set up to gather data on this. Intuitively, it makes

15doi:10.1093/lpr/mgv012
sense to assume that retransfer from superficial locations to deeper layers will happen just as easily as vice versa. The forensic practitioner can assume that DNA present at the sampling was also originally deposited there if this is the case. The probability of retransfer from the side of the roll to the surface area of the tape (and subsequently to the sample) can be minimized by avoiding the side of the tape piece when sampling for DNA.

The network should not be used as a black box, but as a reasoning tool. For example, the forensic practitioner could use the network to establish which prior probabilities have the largest influence on his conclusion. It can, therefore, assist him in his interpretation and facilitate the discussion in the courtroom.

Bibliography


Categorical methods for the interpretation of RNA profiles as cell type evidence and their limitations

Abstract

Existing methods for the interpretation of RNA profiles as evidence for the presence of certain cell types aim for making categorical statements. Such statements limit the possibility to report the associated uncertainty. From a statistical point of view, a probabilistic approach is a preferable choice. Here, four popular existing methods for the interpretation of RNA evidence are discussed.

4.1 Introduction

In forensic cases, information on when or what type of DNA material was donated is often needed. RNA evidence can be used to make statements regarding the cell types present in a crime scene stain.

The interpretation of RNA profiles can be difficult due to numerous uncertainties. Most importantly, markers that are used as cell type specific markers infrequently amplify on non-target cell types.

Existing interpretation methods aim to make a categorical identification statement regarding the presence/absence of different cell types in a sample. Here, four popular existing methods for the interpretation of RNA evidence are discussed. We summarize the advantages and limitations of these methods in the conclusion, and propose the use of probabilistic methods.

4. Categorical methods for the interpretation of RNA profiles as cell type evidence and their limitations

4.2 Method overview

4.2.1 Basic method

Juusola and Ballantyne [2] developed a simple method for the identification of four cell types that are commonly encountered in forensic casework analysis. The authors selected two (supposedly) cell type specific markers for each of these cell types. An identification of a cell type is made when both cell type specific markers amplify. This method, although straightforward, has some problems. Firstly, although the tests on these markers within this study did not show any false positives, other studies did. For example, in [5], saliva samples resulted in amplifications on the MUC4 marker. Secondly, this paper does not explain how one should evaluate the evidence when only one of the two cell type specific markers amplify. Finally, the cell types on which these markers were tested do not cover all the possibilities which is a problem when the method is based on the assumption that the selected markers are cell type specific. If it is subsequently shown that some of these markers also amplify for other cell types (skin, urine, ...), then one needs to discard these markers and others need to be identified in order to use the method.

4.2.2 Relative expression method

The method described in [3] uses the relative expression of cell type specific markers with respect to a housekeeping marker in order to identify cell types. A triplex system consisting of two cell type specific markers and a housekeeping marker is developed for the four cell types of interest in this paper (blood, saliva, semen and menstrual blood).

The cycle number, $C_t$, at which the fluorescence signal generated by the amplification of a marker passes a pre-set threshold line is recorded in order to quantify the results of the triplex for interpretation. This $C_t$ value is recorded for both of the cell type specific markers and the housekeeping marker. The idea is that the cell type specific markers should reach this threshold in fewer cycles than the housekeeping marker when the sample consists of this specific cell type.

A $dC_t$ value is obtained by subtracting the average $C_t$ for the cell type specific markers from the $C_t$ score of the housekeeping markers. A positive $dC_t$ value indicates that the gene associated with the cell type specific marker was present at a higher level than the gene associated with the housekeeping marker, whereas a negative $dC_t$ value indicates the opposite. If the $dC_t$ values for both cell type specific markers are positive, then the associated cell type is labelled as present. If both $dC_t$ values are negative, then this cell type is labelled as not detectable. If one of $dC_t$ is positive and the other is negative, then the result is labelled inconclusive.

This approach is problematic for reasons similar to those with earlier method [2]. If subsequent research regarding the amplification of these markers on other cell types shows that the markers also amplify on samples containing these cell types, then new markers need to be identified in order to use the method. Furthermore, the $dC_t$ values, as defined, appear to contain information regarding
the certainty of the presence statement. Two very high $dC_t$ values suggest that the forensic scientist should have more confidence in the interpretation than two low $dC_t$ values. However, the statement obtained using this method ignores this (un)certainty.

### 4.2.3 $n/2$ method

Lindenbergh et al. [4] address the problem where one is unable to make a presence/absence statement when some of the cell type specific markers fail to amplify, or when non-target cell type markers sporadically amplify.

The authors suggest performing multiple PCRs on a sample in order to control the influence of sporadically amplifying non-target cell type markers. If at least half of the (supposedly) cell type specific markers returned a signal above a preset threshold, the corresponding cell type is labelled *observed*. An *observed* cell type is categorized *observed and fits with* if it is co-expressed with another cell type, for example blood concurring with menstrual secretion. The categorization *sporadically observed* is used when the number of amplification peaks above the threshold is less than half of the possible number. The categorization *sporadically observed, not reliable* is assigned to low-level components whereas the categorization *non specific due to high input* corresponds to situations with spurious signals (possibly due to the high input).

Although this method allows the user to place a finer scale on the (un)certainty of the results it still uses thresholds to categorize the results. If there is a maximum of 12 possible peaks, then observing all 12 will result in the same statement as observing 6. On the other hand, the change in the identification statement from 6 peaks down to 5 peaks is enormous. This is another example of the ‘fall-off-the-cliff-effect’. Furthermore, this method does not incorporate the differences in the amplification rates for different markers and cell types. Therefore it is ‘more difficult’ for some cell types to obtain an *observed* score than for other cell types. Lastly, again, if some markers amplify for cell types other than one for which they were selected, then the utility of this method is severely limited.

### 4.2.4 Marker value and scoring threshold method

Roeder and Haas [5] also address the problem that some markers occasionally amplify on non-target cell types. A scoring system which takes the amplification rates for different markers on target and non-target cell types into account is developed. Each marker is assigned a value based on how often it amplifies on target cell types compared to non-target cell types. A high marker value corresponds with a marker that rarely amplifies on non-target cell types.

In this method the marker values belonging to the markers selected for a cell type that showed an amplification peak are summed to obtain a cell type ‘score’ for a sample which contains cells of unknown type. If this score exceeds a predetermined threshold score, then the corresponding cell type is said to be *observed*.

This method would still work even if subsequent research shows that the markers also amplify in the presence of other cell types. The values for these markers
would decrease and predetermined threshold scores may change, but one can still use the same markers. On the other hand, the method also makes a categorical identification (‘fall-off-the-cliff’) statement. A sample score just above a cell type threshold results in a completely different statement than a sample score just below the cell type threshold. Furthermore, in the paper, the marker values are incorrectly derived because the method uses the absolute number of amplifications of a marker to determine the value of a marker instead of the proportion of times that a marker amplified. The method is also dependent on the number of samples that was used in the training database. However, this is something that can be fixed easily, see [1].

### 4.3 Discussion and conclusion

An overview of the advantages and limitations of the discussed method can be found in Table 4.1. All the existing methods for the interpretation of RNA profiles that are discussed here use a method specific threshold to make a categorical statement regarding the presence of a cell type. This type of statements has several important drawbacks, such as the fall-of-the-cliff effect. Furthermore, context information about e.g. the source of the sample cannot be incorporated.

<table>
<thead>
<tr>
<th>method</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic method</td>
<td>[2]</td>
</tr>
<tr>
<td>relative expression</td>
<td>[3]</td>
</tr>
<tr>
<td>$n/2$</td>
<td>[4]</td>
</tr>
<tr>
<td>marker value and threshold score</td>
<td>[5]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Report different levels of certainty</th>
<th>Applicable to RNA profiles with missing expected peaks</th>
<th>Markers need to be cell type specific</th>
<th>Results can be combined with prior information</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>to a certain degree</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

**Table 4.1:** Overview of methods

One solution to address the short-comings of the previous methods is to use a method that returns a probabilistic statement rather than a definitive conclusion. Probabilistic statements inherently address issues of uncertainty and do not necessarily need cell type specific markers. The focus in the field of DNA evidence evaluation is not on identifying the donor(s) to a questioned stain but on reporting the evidential value of the DNA profile as a ratio of conditional probabilities, namely the likelihood ratio. In this approach one is reporting the rarity of observing this match based on multiple hypotheses. We suggest a similar approach for the evaluation of RNA profiles. The amplification variability of markers in
RNA research is something that can be modelled conveniently using a probability model. Also, there is no need to select a new set of markers when some new study shows that a cell type specific marker is not as specific under some other conditions.Suggestions for methods based on probability theory can be found in [1].

Bibliography


A probabilistic approach for the interpretation of RNA profiles as cell type evidence

Abstract

DNA profiles can be used as evidence to distinguish between possible donors of a crime stain. In some cases, both the prosecution and the defense claim that the cell material was left by the suspect but they dispute which cell type was left behind. For example, in sexual offense cases the prosecution could claim that the sample contains semen cells where the defense argues that the sample contains skin cells. In these cases, traditional methods (e.g. a phosphatase test) can be used to examine the cell type contained in the sample. However, there are some drawbacks when using these methods. For instance, many of these techniques need to be carried out separately for each cell type and each of them requires part of the available sample, which reduces the amount that can be used for DNA analysis.

Another option is messenger RNA (mRNA) evidence. mRNA expression levels vary among cell types and can be used to make (probability) statements about the cell type(s) present in a sample. Existing methods for the interpretation of RNA profiles as evidence for the presence of certain cell types aim at making categorical statements. Such statements limit the possibility to report the associated uncertainty. Some of these existing methods will be discussed. Most notably, a method based on a ‘n/2’ scoring rule [Lindenbergh et al, 2012] and a method using marker values and cell type scoring thresholds [Roeder et al, 2013].

From a statistical point of view, a probabilistic approach is the most obvious choice. Two approaches (multinomial logistic regression and naïve Bayes’) are suggested. All methods are compared, using two different datasets and several criteria regarding their ability to assess the evidential value of RNA profiles.

This was chapter originally published as: J. de Zoete, J. Curran and M. Sjerps. A probabilistic approach for the interpretation of RNA profiles as cell type evidence. Forensic Science International: Genetics, (20), pp 30 – 44, 2016.
We conclude that both the naïve Bayes’ method and a method based on multinomial logistic regression, that produces a probabilistic statement as measure of the evidential value, are an important improvement of the existing methods. Besides a better performance, they are flexible and can be adapted to other situations. For example, they could potentially assist in the combination of RNA with DNA evidence.

5.1 Introduction

In forensic cases, it is often insufficient to only consider the question who donated DNA material. To evaluate what actually happened, information on when or what type of material was donated is needed. Other evidence related to the crime stain could be used to assist in answering these questions. For example, the location and/or the amount of cell material that was found might be relevant. Another option is to include evidence regarding the type of cells (body fluid)\(^1\) in the crime stain. This can be done by using traditional methods (microscopy/immunological/chemical/enzymatic) of cell type investigation for forensic purposes [1]. However, there are some drawbacks when using these methods. For example, many of these techniques need to be carried out separately for each cell type and each of these tests requires part of the available sample, which in turn limits the amount that can be used for DNA analysis.

Another option is using messenger RNA (mRNA) evidence. mRNA expression levels vary among cell types and, therefore, analyses for the presence of particular mRNAs can be used to make (probabilistic) statements about the cell type(s) present in the crime scene stain. One benefit of using RNA evidence is that it is possible to co-extract RNA and DNA from the same sample. This is an advantage when the amount of biological material available for analysis is limited. In addition, there exists the ability to simultaneously analyse multiple markers and tissue types within one run which saves time and preserves sample.

The interpretation of RNA profiles, in terms of evidential value, can be difficult. For example, the level of expression (peak height) for different markers can differ substantially due to numerous variables, such as the physical condition of the donor, spurious transcription that occurs whenever RNA polymerase binds to DNA, or the cell type of the tested sample. Furthermore, peaks for distinct markers for the same cell type may differ in heights (or may drop out) due to the different expression levels for the specific mRNAs and to the regulation of mRNA by biological, physiological or environmental factors. Moreover, markers that are used as cell type specific markers (markers that only amplify given that the sample contained a specific cell type) infrequently amplify on non-target cell types. So, making a (probabilistic) statement regarding the cell type of the examined crime stain based on marker expression levels seems very problematic. However, one could use present/absent data of the markers (ignoring the corresponding peak

---

\(^1\)In this study, we will use the term cell type instead of the other commonly used term body fluid. Neither of them covers all considered classes, but we stick with one of them for clarity reasons.
5.2. Literature overview of RNA interpretation methods

5.2.1 Multiplex mRNA profiling for the identification of body fluids

Juusola and Ballantyne developed a multiplex reverse transcription-polymerase chain reaction (RT-PCR) method for the identification of the cell types that are commonly encountered in forensic casework analysis, namely blood, saliva, semen, and vaginal secretions [10]. The authors describe two cell type specific genes for each cell type. These are \( \beta \)-spectrin (SPTB) and porphobilinogen deaminase (PBGD) for blood, statherin (STATH) and histatin 3 (HTN3) for saliva, protamine 1 (PRM1) and protamine 2 (PRM2) for semen, and humanbeta-defensin 1 (HBD-1) and mucin 4 (MUC4) for vaginal secretions. The method that is used to designate a cell type as being present in a sample consists of checking whether the selected markers amplify on an unknown sample. If the markers corresponding to one of the cell types amplify, then this cell type is identified as being (one of) the cell type(s) of the sample.

The interpretation of RNA profiles with this method is a very simple one. An identification of cell types can be made given the amplified markers. However, there is a problem. Although the tests within this study did not show any false positives, other studies did. Both [12] and [17] (see Table 5.15 and 5.16 in the
appendix of this paper) show that some of these markers also amplified on non-target bodily fluids (for example, MUC4 on saliva). A question not addressed by Juusola and Ballantyne (presumably because the study was proof of concept, not an exhaustive approach) is how to interpret the results when only one of the two cell type specific markers amplifies.

5.2.2 $n/2$ scoring method

Lindenbergh et al. [12] address the problem where one is unable to make a presence/absence statement when some of the cell type specific markers fail to amplify, or when non-target cell type markers sporadically amplify. This paper describes a procedure that “…accommodates unbiased analysis and interpretation of RNA profiles.” This procedure reduces the marker information obtained from a RNA profile to a table which summarizes which cell types were observed/observed and fits/sporadically observed and fits/not observed/sporadically observed, not reliable/non-specific due to high input (see Table 5.1).

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Observed</th>
<th>Observed and fits</th>
<th>Sporadically observed, and fits</th>
<th>Not observed</th>
<th>Sporadically observed, not reliable</th>
<th>non-specific due to high input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case #1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>blood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>menstrual secretion</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vaginal mucosa</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mucosa</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>saliva</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>semen</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Example of a table to insert observations from obtained RNA profiles, as suggested in [12]

The authors suggest performing multiple PCRs and selecting, from the obtained RNA profiles, a set of ‘informative’ profiles for further consideration, see [12]. Lindenbergh et al. determine, for each cell type, how often a signal of good peak morphology and above detection threshold is present relative to the number of times a signal could have occurred. For example, if the informative set contains four RNA profiles, and the multiplex consists of three blood markers, then there is a maximum of 12 possible signals for blood. The term ‘observed’ is used when cell type signals are present in at least half of the possible positions. For example,
the term ‘observed’ would be used in our example if at least six (of the possible 12) markers returned a signal above the detection limit. This criterion gives rise to the name of ‘the n/2 scoring method’. An observed cell type is categorized ‘observed and fits with’ if it is co-expressed with another cell type. For example, this may happen when blood co-occurs with menstrual secretions. If cell type signals are present in fewer than half of the possible positions, then the term ‘sporadically observed’ is used. The term ‘sporadically observed, not reliable’ is assigned to low-level components whereas ‘non-specific due to high input’ corresponds to situations with spurious signals (possibly due to the high input). Now, given a set of hypotheses that consider the cell type, a crime stain can be evaluated using the resulting ‘observed/not observed/…’ table (Table 5.1).

One drawback of this method is that it does not incorporate the differences in the amplification rates for different markers and cell types. Blood markers, for example, tend to amplify much more often than semen markers (see Table 5.15 in the Appendix). At least two markers per tissue type are included within the multiplex to account for the fluctuating behaviour of the individual markers. This practice will increase the probability of observing markers for the cell type of the sample, but it does not address the problem that, for some cell types at least, it is more difficult to obtain an ‘observed’ score due to the lower amplification rate of its markers. Also, the different summary statements (observed/observed and fits/…) together with the threshold that determines in which category the evidence from the RNA profiles falls lead to a loss in specificity of the actual data. If there is a maximum of 12 possible peaks, then observing all 12 will result in the same statement as observing 6. That is, there is no change in ‘evidential strength’ with an increase in the number of observed peaks. On the other hand, the change in identification statement from 6 peaks down to 5 peaks is enormous. This is an example of the so-called ‘fall-off-the-cliff-effect’ [16]. Lastly, this method needs cell-specific conditionally independent markers. If some markers amplify for tissue types other than one for which they were selected, or if one marker amplifies only when another marker amplifies regardless of the cell type of the sample, then this severely limits the utility of this method. However, markers which are not cell type specific can still carry information regarding the cell type contained in a sample. For example, the amplification of (housekeeping marker) ACTB is common when the sample only contains blood, saliva or semen, but it rarely amplifies when the sample only contains menstrual blood or skin (see Table 5.15). Such information cannot be taken into account by this method.

5.2.3 Marker value and scoring threshold method

Roeder and Haas [17] address the problem that some markers amplify on non target cell types. In the paper 32 markers are tested for specificity on different cell types. 30 of these are supposed to be indicative for the presence of forensically relevant cell types. Two housekeeping markers are also added to the multiplexes. Each marker in this paper was tested using more than 200 forensically relevant samples of cell types: semen, saliva, cervicovaginal fluid (CVF), blood, menstrual blood (MB), sweat, and skin. The markers are tested using four different multiplexes:
5. A probabilistic approach for the interpretation of RNA profiles as cell type evidence

A multiplex for semen, one for saliva and CVF, one for blood and one for MB. The sweat and skin samples are included to test whether the markers are cell type specific on a bigger prior set of cell types. A scoring system which takes the amplification rates for different markers on target and non target cell types into account is developed using the data obtained from these tests which minimizes the chances of misidentification of a sample due to marker expression in a non-target cell type. The proportion of amplifications on the target cell type for each marker is multiplied by a 100 to obtain a ‘marker value’. For example, the marker value for MSX1 in the MB multiplex is equal to (using the data from Table 5.16),

\[
\frac{19}{19 + 1 + 16 + 8} \cdot 100 = 43.18
\]

Now, for a sample of which the contained cell type is unknown, one adds the marker values of the markers that showed an amplification peak to arrive at a ‘score’ for the sample. The samples of this study were used to determine threshold scores for all considered cell types. These thresholds were set to minimize the chances of obtaining false positive results\(^2\).

This method does take the amplification of marker on non target cell types into account. By using at least 5 markers per cell type, the probability (i.e. the sensitivity) of observing at least one marker also increases when the sample consists of the target cell type. Furthermore, by setting a scoring threshold which implies that multiple markers need to amplify in order to give a positive result for a sample being of a specific cell type, accounts for possible ‘drop-ins’ of a marker. On the other hand, the method aims on making a truth/false statement regarding the cell type of the sample. Setting a scoring threshold implies the same dissatisfactory ‘fall-of-the-cliff’ from the \(n/2\) threshold of the ‘\(n/2\) scoring method’. For example, a sample which shows amplified markers on MSX1, SFRP4, MMP7, MMP10, HBB, ALAS2, and GlycoA gives a menstrual blood score of

\[
2 \cdot (43.18 + 45.45 + 39.53 + 66.67) + 87.69 + 98.04 + 92.50 = 667.89
\]

which is lower than the marker threshold (670). Therefore, the sample does not test positive for menstrual blood. Another sample, which shows amplified markers on MSX1, SFRP4, MMP11, MMP10, HBB, ALAS2, and GlycoA (the samples only differ on 1 marker), gives a menstrual blood score of

\[
2 \cdot (43.18 + 45.45 + 44.12 + 66.67) + 87.69 + 98.04 + 92.50 = 677.07
\]

which is higher than the marker threshold (670). Therefore, this sample does test positive for menstrual blood. Although the sample scores as well as the observed markers are approximately the same, the outcome is completely different.

The previous example showed that using the method as described could result in completely different statements for two samples with very similar RNA profiles.

\(^2\)Note that, minimizing the number of false positives usually comes at the cost of increasing the number of false negatives and vice versa.
This is another example of a ‘fall-off-the-cliff’ decision. In addition, it is also important to note that the method to determine the marker values of the different markers is undesirable. The score is computed as the proportion of amplified samples on the target cell type compared to all the amplifications. Therefore, this score depends on the number of samples that were tested for each cell type. The number of samples that were analysed per cell type differs (between 12 and 50). Therefore, suppose that the proportion of samples that amplifies on a specific marker is correct (for example 19/35 (54%) for the \textbf{MSX1} marker on menstrual blood, which resulted in a marker score of 43.18), then a new experiment under the same conditions but with a different number of samples will result in another marker score (in the \textbf{MSX1} example, if 100 menstrual blood samples were tested, the marker score would be 68.47)\textsuperscript{3}.

If the marker scores were computed while using the proportion (percentage) of amplifications of a marker on a specific cell type, then, marker values that are independent of the number of samples that were tested are obtained. For example, the \textbf{MSX1} marker value becomes

$$\frac{19}{35 + \frac{1}{50} + \frac{16}{49} + \frac{8}{32}} \cdot 100 = 47.64$$

The complete list with marker values based on percentages instead of the observed number of amplification is given in Table 5.2.

<table>
<thead>
<tr>
<th>marker</th>
<th>HBD1</th>
<th>MUC4</th>
<th>Lcris</th>
<th>Ljen</th>
<th>Lcris2</th>
<th>Lgas</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>77.35</td>
<td>83.58</td>
<td>94.25</td>
<td>46.77</td>
<td>94.56</td>
<td>94.63</td>
</tr>
<tr>
<td>original</td>
<td>77.78</td>
<td>81.58</td>
<td>95.45</td>
<td>64.04</td>
<td>95.65</td>
<td>94.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>marker</th>
<th>STATH</th>
<th>MUC7</th>
<th>SMR3B</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>97.06</td>
<td>94.98</td>
<td>88.76</td>
</tr>
<tr>
<td>original</td>
<td>98.04</td>
<td>96.00</td>
<td>92.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>marker</th>
<th>HBB</th>
<th>ALAS2</th>
<th>PRF1</th>
<th>GlycoA</th>
<th>PF4</th>
<th>SPTB</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>92.35</td>
<td>98.85</td>
<td>81.12</td>
<td>92.72</td>
<td>95.83</td>
<td>94.37</td>
</tr>
<tr>
<td>original</td>
<td>87.69</td>
<td>98.04</td>
<td>72.22</td>
<td>92.50</td>
<td>94.34</td>
<td>92.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>marker</th>
<th>MSX1</th>
<th>LEFTY2</th>
<th>SFRP4</th>
<th>MMP7</th>
<th>Hs202072</th>
<th>MMP10</th>
<th>MMP11</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>47.64</td>
<td>87.97</td>
<td>43.24</td>
<td>43.51</td>
<td>80.04</td>
<td>70.47</td>
<td>40.53</td>
</tr>
<tr>
<td>original</td>
<td>43.18</td>
<td>88.89</td>
<td>45.45</td>
<td>39.53</td>
<td>83.33</td>
<td>66.67</td>
<td>44.12</td>
</tr>
</tbody>
</table>

\textbf{Table 5.2:} Marker values per marker when using the percentages (new) and when using the actual number of amplifications (original). The markers PRM1, PRM2, KLK3, SEMG1, TGM4, HTN3, PRB4 and PBGD amplified on target cell types only and, hence, got the same marker value (100) under the new and original system.

Although most of the marker values are approximately the same, there are some markers where there are more substantial differences, such as \textbf{Ljen} and \textbf{PRF1}.

\textsuperscript{3} \frac{19}{35} \cdot 100 + \frac{1}{100} + \frac{16}{50} + \frac{8}{49} \cdot 100 = 68.47
Also, as an example, suppose a sample showed amplified markers on MSX1, SFRP4, MMP7, MMP10, HBB, ALAS2, and GlycoA. The menstrual blood score, with the original marker values, would be

\[ 2 \cdot (43.18 + 45.45 + 39.53 + 66.67) + 87.69 + 98.04 + 92.50 = 667.89 \]

whereas using the new marker values gives a score of

\[ 2 \cdot (47.64 + 43.24 + 43.51 + 70.47) + 92.35 + 98.85 + 92.72 = 693.64. \]

The menstrual blood score threshold from the paper is 670, so, the old marker values will not result in an identification statement for menstrual blood. Using the new marker values gives a sample score that exceeds 670. However, the cell type thresholds will be different for these new marker values since they are determined based on the samples and marker values of the dataset. Even so, the example shows that samples could potentially get substantially different scores using the corrected marker values.

In conclusion, in practice most samples will be classified the same.

5.2.4 Other methods

There are more methods in the literature for the interpretation of RNA profiles than we have described here. However, they tend to exhibit problems similar to those we have highlighted in the previous sections. For example, Juusola and Ballantyne [11] used the relative expression of tissue-specific genes with respect to a housekeeping gene to ‘identify’ tissue types. They developed a triplex system consisting of two cell type specific markers and one housekeeping marker for each of the four cell types of interest in this paper (blood, saliva, semen and menstrual blood). Now, if the amplification peaks of both of the cell type specific markers reach a certain rfu threshold in fewer cycles than the housekeeping gene, then the corresponding cell type is labelled as ‘present’. Similarly, if the housekeeping marker is the first marker to reach this threshold, the corresponding cell type is labelled as ‘not detectable’ in the sample. This method is very similar to the one we discussed in Section 5.2.1 ([10]). Although the method captures part of the problems with ‘drop-in’ amplification peaks by concentrating on the amplification peaks relative to a housekeeping marker, this method still aims to make an identification statement based on a threshold (another “fall-of-the-cliff” example). In addition, the method only works with cell type specific markers and the evidential value of samples in which the cell type specific markers reach this rfu threshold substantially faster than the housekeeping marker is the same as for a sample in which these cell type specific markers reach the rfu threshold ‘just before’ the housekeeping marker.

5.2.5 Summary

In most literature regarding methods to interpret RNA results or research focused on finding markers to use RNA multiplexes, the focus lies at identifying the cell
5.2. Literature overview of RNA interpretation methods

type of the sample or to determine whether a marker is cell type specific. On the one hand, the papers convey the message that it is much harder to interpret RNA profiles (compared to DNA profiles) because of the numerous variables that (can) influence the outcome. On the other hand, a simple present/absent conclusion is made, and the examined markers are evaluated on whether they can assist in ruling out false positives, suggesting that the interpretation does not require probabilistic reasoning. Furthermore, the papers that propose (new) methods to interpret RNA evidence discuss other methods to determine the cell type of a stain based on whether they produce false positive results. For example, in [17], it is mentioned that “a drawback of immunological, chemical and enzymatic assays methods for cell type investigation is that they produce false positive results”. However, all of the methods we discussed here produce false positives/wrong classifications as well. By aiming on making a categorical identification statement, one is not only allowing for possible false positives/negatives but also one ignores the (un)certainty in this statement.

Identification of potential donors to crime scene stains is generally not the main focus when we consider DNA evidence. The ‘match’ between a DNA profile obtained from a crime stain and a DNA profile obtained from a suspect can only be regarded as a ‘positive’. When the evidential value is reported as a ratio of conditional probabilities, i.e. a likelihood ratio, one is only reporting the probability of observing this match based on multiple hypotheses. For example, if the prosecution hypothesis ($H_p$) states that the suspect was the donor of the crime stain whereas the defense hypothesis ($H_d$) states that some other unknown unrelated person was the donor of the crime stain, the reported evidential value of this ‘match’ ($E$) is reported as a likelihood ratio,

$$\frac{\Pr(E|H_p)}{\Pr(E|H_d)}.$$  

This likelihood ratio does not make any categorical statement regarding whether the suspect left the crime stain or not. It only tells us, combined with the prior probabilities for the hypotheses, how probable it is that the suspect was the donor of the crime stain compared to how probable it is that he was not. When more (informative) (unlinked) ‘matching’ markers are available for the DNA profiles, the likelihood ratio increases. So, in contrast to the methods to interpret RNA profiles, it is possible to assign a match specific evidential value where the observation of additional markers will result in a higher evidential value. Also, the focus does not lie on identifying the donor, but assigning probabilities regarding how likely it is to observe some evidence given that the suspect was the donor compared to a situation where an unknown was the donor.

In this paper, we propose two methods which will allow us to do something similar. We believe that, apart from housekeeping markers that are (supposed to) amplify on any tissue type, every marker can potentially provide information regarding the cell type of the sample. Although cell type specific markers can

---

4In some cases, ‘false positive’ does not fully capture the result. For example, when a menstrual blood sample is classified as a sample containing cervicovaginal fluid one can understand the positive result. In these cases, the term ‘wrong classification’ is more appropriate.
be very valuable when making a probabilistic statement about the cell type contained in the sample, other non-cell type specific markers, (like PRF1 from Table 5.16), can be used as well. If there are markers that amplify most often for a subgroup of the relevant forensic cell types, then these can be very informative in discriminating between this subgroup and its complementary part. Furthermore, in these categorical identification statement methods, the information that a cell type specific marker does not amplify is not used as evidence to support that it is one of the other cell types.

The fluctuative behaviour of markers in RNA research is something that can be modelled conveniently using a probability model. In principle, it is possible to capture all the unknown variation that biology presents: MB markers may vary on day in cycle, on age, on ethnicity, per woman. If there is enough data that provide information regarding the amplification rates of markers under these conditions, then it is possible to take all these different situations into account. Although it is unlikely that enough data can be gathered to capture this unknown variation, a method based on a probability model has the capability to incorporate this variation. Another advantage of methods resulting in a probabilistic statement is that there is no need to select a new set of markers when some new study shows that a cell type specific marker is not as specific under some other conditions. In a probability model, one can incorporate the new data without ‘trashing’ the complete model.

When we compare these evaluation methods to the likelihood ratio method that is commonly used to report DNA evidence, we see a number of important advantages of the latter. First, a likelihood ratio is not a categorical statement (yes/no) but a continuous value, that avoids the ‘fall-of-the-cliff-effect’ mentioned above. Second, a likelihood ratio is coherent in the sense that we obtain a higher value when more loci match. Categorical statements do not have this property. Finally, the likelihood ratio combines logically with the other available information that determines the prior probabilities to posterior probabilities. Categorical statements do not depend on the other available information in the case, which is counter-intuitive.

5.3 Material and methods

In this paper, two well known probabilistic methods are introduced and their performance is tested, Naïve Bayes [8] and Multinomial Logistic Regression [2]. Calculations were performed using MATLAB [13] and R [15]. Data were obtained from the authors of [12] and [17].

5.4 Probabilistic methods

Here, two methods that will result in probabilistic statements regarding the present cell type are introduced and discussed. The first method, a Bayesian network with a naïve Bayes’ assumption (Section 5.4.1), generates likelihood ratios which can be converted into posterior probabilities when assuming a prior distribution over
5.4. Probabilistic methods

the cell types. The second method, based on multinomial logistic regression (Section 5.4.2), generates posterior probabilities. This method implicitly uses a prior distribution based upon the relative frequencies of the cell types in the training database.

5.4.1 Bayesian network/naïve Bayes’ approach

If it can be assumed that markers amplify (conditionally) independently of each other given the cell type contained in the sample (naïve Bayes assumption, see [8]), then a (very basic) Bayesian network can be constructed to obtain a posterior probability distribution over the considered cell types (see the network in Figure 5.1). However, before moving forward, the validity of this assumption should be tested.

Figure 5.1: Bayesian network for the evaluation of RNA profiles, $e_1, e_2, \ldots, e_n$ represent a total of $n$ markers.

Conditional independence assumption

In order to test the conditional independence assumption two data sets are used. The data used in [12] (omitting the dilution series, and only considering cell types with sufficient samples), consisted of 148 samples (33 blood, 16 menstrual blood, 23 saliva, 28 semen, 48 skin) on 19 markers. The data used in [17] consisted of 193 samples (25 blood, 49 CVF, 33 menstrual blood, 50 saliva, 36 semen) on 32 markers. A contingency table can be constructed for every pair of markers and for every cell type (see the example in Table 5.3).

<table>
<thead>
<tr>
<th></th>
<th>18S-rRNA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LOR</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>no</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.3: Contingency table for the amplification of 18S-rRNA and LOR on 50 skin samples of the data from [12]

In general, the one-tailed $p$-value for a test of (conditional) independence can be easily computed for a $2 \times 2$ contingency table with Fisher’s Exact test. For the data from Table 5.3, the corresponding $p$-value is,
A probabilistic approach for the interpretation of RNA profiles as cell type evidence

\[ p(\text{18S-rRNA.LOR})(\text{skin}) = 0.16 \]

Hence, using a significance level of \( \alpha = 0.05 \), the \( p \)-value tells that there is no reason to reject the null hypotheses that these markers are conditionally independent when the sample contains saliva. A summary of the number of marker pairs that returned a \( p \)-value smaller than significance level \( \alpha \) can be found in Table 5.4. An overview of the obtained \( p \)-values for all conditional independence tests for semen and skin data from [12] can be found in 5.B.

<table>
<thead>
<tr>
<th>Lindenbergh et al., 2012 data</th>
<th>( p )-values</th>
<th>Total</th>
<th>(&lt; 0.05)</th>
<th>(&lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>blood</td>
<td>171</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>menstrual blood</td>
<td>171</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>saliva</td>
<td>171</td>
<td>12</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>semen</td>
<td>171</td>
<td>10</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>skin</td>
<td>171</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Roeder et al., 2013 data</th>
<th>( p )-values</th>
<th>Total</th>
<th>(&lt; 0.05)</th>
<th>(&lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>blood</td>
<td>171</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CVF</td>
<td>171</td>
<td>27(^5)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>menstrual blood</td>
<td>171</td>
<td>21</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>saliva</td>
<td>171</td>
<td>7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>semen</td>
<td>171</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.4:** Summary of the Fisher exact tests for conditional independence for both databases

These results show some indications of dependency. However, with 171 tests we expect 8.55 \( p \)-values smaller than 1.71 smaller than 1%. Therefore, we conclude that the assumption of conditional independence is supported to some degree (but not fully) by these data.

**Likelihood ratio equation using naïve Bayes assumption**

The conditional independence assumption simplifies the likelihood ratio computation substantially. If the hypotheses pair of interest consists of,

\(^5\)The number of \( p \)-values that are lower than 0.05/0.01 for CVF and menstrual blood samples is remarkably high. However, the CVF samples and menstrual blood samples in this study were analysed in two different laboratories. If we separate the samples into two sets, one for each laboratory, the number of \( p \)-values smaller than 0.05 and 0.01 decreases substantially. For menstrual blood samples, laboratory 1 samples result in 6 \( p \)-values \(< 0.05 \) and 2 \( p \)-values \(< 0.01 \). Laboratory 2 samples result in 2 \( p \)-values \(< 0.05 \) and 0 \( p \)-values \(< 0.01 \). For CVF samples, laboratory 1 samples result in 16 \( p \)-values \(< 0.05 \) and 5 \( p \)-values \(< 0.01 \). Laboratory 2 samples result in 4 \( p \)-values \(< 0.05 \) and 0 \( p \)-values \(< 0.01 \).
5.4. Probabilistic methods

\( H_p \): The crime stain contains cell type \( x_1 \).

\( H_d \): The crime stain contains cell type \( x_2 \).

and the RNA profile is denoted by \( E = (e_1, e_2, \ldots, e_n) \) \((n\text{ markers})\), where \( e_i \) is

\[
e_i = \begin{cases} 
1 & \text{if marker } i \text{ amplified.} \\
0 & \text{if marker } i \text{ did not amplify.}
\end{cases}
\]

then the evidential value (in terms of a likelihood ratio) of the observed RNA profile is given by,

\[
LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} = \frac{\Pr(e_1, e_2, \ldots, e_n|H_p)}{\Pr(e_1, e_2, \ldots, e_n|H_d)} = \frac{\Pr(e_1|H_p) \cdot \Pr(e_2|H_p) \cdots \Pr(e_n|H_p)}{\Pr(e_1|H_d) \cdot \Pr(e_2|H_d) \cdots \Pr(e_n|H_d)}
\]

The conditional probabilities \( \Pr(e_i|H_p), \Pr(e_i|H_d) \) can be estimated from data. For example, if \( H_p \) states that the sample contained menstrual blood and \( H_d \) states that the sample contained saliva, and \( e_i \) is the \textbf{HBD1} marker, then we can use the data from [12] or [17] to estimate the conditional probabilities, see Table 5.5.

<table>
<thead>
<tr>
<th>HBD1</th>
<th>Lindenbergh et al. data</th>
<th>Roeder et al. data</th>
</tr>
</thead>
<tbody>
<tr>
<td>menstrual blood</td>
<td>7/16</td>
<td>13/33</td>
</tr>
<tr>
<td>saliva</td>
<td>4/32</td>
<td>4/50</td>
</tr>
</tbody>
</table>

\textbf{Table 5.5:} Estimated amplification rates from the data of [12] and [17] for \textbf{HBD1}, see Table 5.15 and 5.16

The posterior odds are then given by,

\[
\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = LR \cdot \frac{\Pr(H_p)}{\Pr(H_d)}
\]

\textbf{Dealing with markers which always/never amplify}

The dataset that is used to estimate the conditional probabilities can contain markers that, within the samples used to construct the dataset, always or never amplified for particular cell types. As can be seen from Table 5.16 (the data used in [17]), the \textbf{HBB} marker always amplified when the sample contained blood and never amplified when the marker contained saliva. This would mean that the conditional probability of observing an amplification peak for \textbf{HBB} when the sample contained blood would be estimated as 1. For saliva, this estimated conditional
5. A probabilistic approach for the interpretation of RNA profiles as cell type evidence

probability would be 0. This leaves very little flexibility in the method. If the RNA profile of an unknown sample shows amplification peaks above the threshold level on all markers that one would expect when the sample contains blood except for the HBB marker, then the posterior probability that the sample contains blood given the RNA profile is equal to 0°. Note that the data from Lindenbergh et al. (Table 5.15) shows that they did not observe amplification peaks above their threshold level for some blood samples (and did observe amplification peaks above the threshold level for some saliva samples). To add more flexibility, we advise to add two ‘artificial’ observations to the data. One that returned no amplification peak for any marker for every cell type and one that shows an amplification peak for every marker for every cell type. This is equivalent to assuming a uniform prior and deriving the posterior distribution for the probabilities based on binomial/multinomial sampling ideas. These updated estimated amplification rates for the HBB marker with the data from [17] and [12] can be found in Table 5.6.

<table>
<thead>
<tr>
<th>HBB</th>
<th>Lindenbergh et al. data</th>
<th>Roeder et al. data</th>
</tr>
</thead>
<tbody>
<tr>
<td>blood</td>
<td>31/34</td>
<td>26/27</td>
</tr>
<tr>
<td>menstrual blood</td>
<td>7/18</td>
<td>33/35</td>
</tr>
<tr>
<td>saliva</td>
<td>3/34</td>
<td>1/52</td>
</tr>
<tr>
<td>semen</td>
<td>4/32</td>
<td>1/30</td>
</tr>
<tr>
<td>skin</td>
<td>1/50</td>
<td>1/14</td>
</tr>
<tr>
<td>CVF</td>
<td>-</td>
<td>9/51</td>
</tr>
<tr>
<td>sweat</td>
<td>-</td>
<td>1/14</td>
</tr>
</tbody>
</table>

Table 5.6: Updated estimated amplification rates from the data of [12] and [17] for HBB to make the method more flexible

Prior probabilities in the likelihood ratio equation

In the first example, we worked with a hypotheses pair that stated,

\[ H_p : \text{The crime stain contains cell type } x_1 \]
\[ H_d : \text{The crime stain contains cell type } x_2 \]

which resulted in the following odds form of Bayes’ theorem,

\[
\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{LR}{\left( \frac{\Pr(e_1|H_p)}{\Pr(e_1|H_d)} \cdot \frac{\Pr(e_2|H_p)}{\Pr(e_2|H_d)} \cdots \frac{\Pr(e_n|H_p)}{\Pr(e_n|H_d)} \right)} \left( \frac{\Pr(H_p)}{\Pr(H_d)} \right) \tag{5.1}
\]

With a different pair of hypotheses, namely hypotheses of which at least one consists of ‘sub-hypotheses’, for example,

\footnote{Similarly, if the RNA profile of a sample shows amplification peaks above the threshold level on all markers that one would expect when the sample contains saliva and on HBB, the posterior probability that the sample contains saliva given the RNA profile is equal to 0.}
5.4. Probabilistic methods

$H_p$: The crime stain contains cell type $x_1$

$H_d$: The crime stain does not contain cell type $x_1$

the prior probabilities become part of the likelihood ratio (this phenomenon has been observed in many other situations, for example in [18] and [14]). Using the naïve Bayes assumption, the likelihood ratio simplifies to

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)} = \frac{Pr(e_1, e_2, \ldots, e_n|H_p)}{Pr(e_1, e_2, \ldots, e_n|H_d)} = \frac{Pr(e_1|H_p)}{Pr(e_1|H_d)} \cdot \frac{Pr(e_2|H_p)}{Pr(e_2|H_d)} \cdot \cdots \cdot \frac{Pr(e_n|H_p)}{Pr(e_n|H_d)}$$

A forensic scientist cannot estimate the conditional probabilities $Pr(e_i|H_d)$, without specifying the cell types contained in $H_d$. Since it impossible for a forensic laboratory to have a database that contains all possible cell types, the most informative (but simplified) hypotheses pair that they can use to report a likelihood ratio for the evidence is

$H_p^*: The crime stain contains cell type $x_1$.

$H_d^*: The crime stain does not contain cell type $x_1$, it contains $x_2$, $x_3$, ..., or $x_m$.

where $m$ is the number of other cell types contained in the database. So, $H_d^*$ can be subdivided into

$$H_{d2}^*: The crime stain contains cell type $x_2$.

$$\vdots$$

$$H_{dm}^*: The crime stain contains cell type $x_m$.$$ 

Now, the likelihood ratio can be expanded to

$$LR = \frac{Pr(e_1|H_p^*)}{Pr(e_1|H_d^*)} \cdot \frac{Pr(e_2|H_p^*)}{Pr(e_2|H_d^*)} \cdot \cdots \cdot \frac{Pr(e_n|H_p^*)}{Pr(e_n|H_d^*)}$$

$$= (Pr(H_{d2}^*) + \ldots + Pr(H_{dm}^*)) \cdot \prod_{i=1}^{n} \frac{Pr(e_i|H_p^*)}{Pr(e_i|H_{d2}^*)} \cdot \frac{Pr(e_i|H_{d2}^*)}{Pr(e_i|H_{d3}^*)} \cdot \cdots \cdot \frac{Pr(e_i|H_{dm}^*)}{Pr(e_i|H_{dm}^*)}$$

$$= (Pr(H_{d2}^*) + \ldots + Pr(H_{dm}^*)) \cdot \prod_{i=1}^{n} \frac{Pr(e_i|H_p^*)}{\sum_{j=2}^{m} Pr(H_{dj}^*) \cdot Pr(e_i|H_{dj}^*)}$$

(5.2)
The likelihood that a marker amplifies depends on the cell type contained in a sample. If either hypothesis (in this example the defense hypotheses $H^*_d$) considers multiple cell types, the likelihood of observing the evidence (the RNA profile) is different for all these cell types. These likelihoods are weighted by the prior probabilities that the sample contained these cell types.

In forensic practice, it is common that the forensic practitioner/scientist reports the evidential value of an observation under two hypotheses in the form of the likelihood ratio. The judge or jury may combine this likelihood ratio with prior probabilities for the two hypotheses to obtain the posterior probabilities. In situations, such as this, where there are multiple alternative explanations (hypotheses) under consideration, the $LR$ cannot be separated from the prior probabilities. As such it will be impossible to report the $LR$ to the judge because the role of the expert and the jury has become confounded. One could argue that this is not a problem as the prior probabilities assigned to the competing hypotheses should be determined without knowledge of the likelihood ratio (one should only be aware what ‘background’ information can be used to determine the prior probabilities and what ‘evidence’ cannot, since it is included in the likelihood ratio computation). This would mean that the likelihood ratio would be reported as a formula, which, given the prior probabilities, results in a value. However, when applied in practice, this may result in problems. We will address this issue in the discussion, see Section 5.6.

### 5.4.2 Multinomial logistic regression

A method based on multinomial logistic regression (MLR) can also be used to obtain likelihoods and a posterior probability distribution over the cell types given the results from a RNA profile, see [2]. In logistic regression, it is assumed that,

$$\log \left\{ \frac{\Pr(\text{cell type } i | E)}{\Pr(\text{cell type } K | E)} \right\} = \beta_{0,i} + \beta_i^T E$$

where $K$ is a fixed ‘pivot’ cell type. This can be simplified to obtain equations for the conditional probabilities for the individual cell types.

$$\Pr(\text{cell type 1} | E) = \frac{\exp \left( \beta_{0,1} + \beta_1^T E \right)}{1 + \sum_{j=1}^{K-1} \exp \left( \beta_{0,j} + \beta_j^T E \right)}$$

$$\Pr(\text{cell type 2} | E) = \frac{\exp \left( \beta_{0,2} + \beta_2^T E \right)}{1 + \sum_{j=1}^{K-1} \exp \left( \beta_{0,j} + \beta_j^T E \right)}$$

$$\vdots$$

$$\Pr(\text{cell type } K-1 | E) = \frac{\exp \left( \beta_{0,K-1} + \beta_{K-1}^T E \right)}{1 + \sum_{j=1}^{K-1} \exp \left( \beta_{0,j} + \beta_j^T E \right)}$$

$$\Pr(\text{cell type } K | E) = \frac{1}{1 + \sum_{j=1}^{K-1} \exp \left( \beta_{0,j} + \beta_j^T E \right)}$$
The parameters $\beta_{0,i} \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^p$ are estimated directly using a training database in such a way that they maximize the conditional likelihood of the samples of known cell type. A benefit of the MLR method is that it does not need the conditional independence assumption used in the NB method. Furthermore, there is no need to introduce artificial observations to prevent posterior probabilities equal to zero or one. A drawback of this method is that it returns posterior probabilities based on a prior distribution that is implicitly based on the relative frequencies of the different cell types in the training database (see the discussion in Section 5.6).

The performance of this and the Naïve Bayes method will be tested on several criteria using data from Lindenbergh et al. and Roeder et al. in Section 5.5.

5.5 Method comparison

5.5.1 Data from Lindenbergh et al. [12]

We compare the $n/2$ scoring method with the proposed probabilistic statement methods on the data from [12]. A uniform prior was assumed to be able to calculate posterior probabilities and compare with the $n/2$ scoring method. An overview of the data can be found in Table 5.15. It is not possible to include the marker value and threshold score method on this dataset, since the marker thresholds were set specifically for the dataset from [17].

**Most likely cell type of a known sample**

We compared the methods based on the proportion of ‘known’ samples of which the posterior distribution over the considered cell types puts the most mass on the correct cell type. The performance of the $n/2$ scoring method on these 158 samples is summarized in Table 5.7.

<table>
<thead>
<tr>
<th>True cell type</th>
<th>Blood</th>
<th>CVF</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen</th>
<th>multiple with true</th>
<th>multiple without true</th>
<th>no class.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>23 (72%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (16%)</td>
<td>0 (0%)</td>
<td>4 (13%)</td>
</tr>
<tr>
<td>MB</td>
<td>0 (0%)</td>
<td>6 (38%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (19%)</td>
<td>5 (31%)</td>
<td>1 (6%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>11 (34%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>4 (13%)</td>
<td>8 (25%)</td>
<td>9 (28%)</td>
</tr>
<tr>
<td>Semen</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>25 (83%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (17%)</td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>47 (98%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (2%)</td>
</tr>
</tbody>
</table>

Table 5.7: Classifications for 158 samples, $n/2$ method

This method returned the correct result in 112 (23+6+11+25+47) of the 158 samples (71%). 80% of the samples returned a result in which the correct cell

---

Lindenbergh et al. [12] suggested the use of multiple RNA profiles from the same source with the $n/2$ scoring method. It is not possible to follow this procedure here because the data does not contain multiple RNA profiles from the same source.
type is at least one of the ‘identified’ cell types (126 out of the 158). The remaining 32 (20%) samples resulted in no identification statement (20 samples) or returned an incorrect identification statement (12 samples).

We are interested in comparing the $n/2$ scoring method with the results obtained from using two well-known statistical tools, namely multinomial logistic regression and naïve Bayes’ classification. The use of such tools usually proceeds by building or training the model, and then by using the model to make predictions. Performance of models is, amongst others, judged on the percentage of correctly predicted classifications. In such experiments, the models are trained on a subset of the data which is independent from the data we used to test them. We randomly divided the 158 samples in the data set into 126 (80%) training and 32 (20%) test samples. This procedure is done multiple (100) times to make sure that the results are not heavily influenced by the composition of the training set.

The results for these methods are summarized in Tables 5.8 (NB) and 5.9 (MLR). The NB classification method does not return an absolute classification, but rather a set of posterior probabilities which tell us the probability that an observation comes from each of the different cell types. Similarly, MLR returns a set of odds ratios which, for comparison purposes, can be converted into posterior probabilities. We adopt a ‘majority rules’ scheme here whereby an observation is classified as arising from the cell type with the highest posterior probability. Over these 100 sets, the average classification error for the NB method was 13% (with an associated standard error of 10%). The average classification error for the MLR method was 14% (with an associated standard error of 10%).

<table>
<thead>
<tr>
<th>cell type with the highest posterior probability</th>
<th>Blood</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen</th>
<th>Skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>601 (96%)</td>
<td>0 (0%)</td>
<td>4 (1%)</td>
<td>20 (3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>MB</td>
<td>0 (0%)</td>
<td>285 (83%)</td>
<td>54 (16%)</td>
<td>6 (2%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>1 (0%)</td>
<td>63 (9%)</td>
<td>424 (64%)</td>
<td>139 (21%)</td>
<td>37 (6%)</td>
</tr>
<tr>
<td>Semen</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>555 (89%)</td>
<td>71 (11%)</td>
</tr>
<tr>
<td>Skin</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>16 (2%)</td>
<td>924 (98%)</td>
</tr>
</tbody>
</table>

Table 5.8: Classifications for $100 \times 32$ samples, naïve Bayes method

Both the tables and the average classification error show that the NB method and MLR outperformed the $n/2$ scoring method. All three methods have difficulties with the saliva samples. However, the NB and the MLR method substantially outperform the $n/2$ scoring method. The $n/2$ scoring method returns a saliva identification for 47% of the saliva samples. In comparison, 64% of the saliva samples are correctly classified when using the NB method and 75% when using the MLR method.

---

8Because the training data and the test data are a uniform draw from the same data set, the prior distribution assumed by the MLR method (based on the relative frequencies in the training data) is valid in this case.
5.5. Method comparison

<table>
<thead>
<tr>
<th>True cell type</th>
<th>Blood</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen</th>
<th>Skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>566 (91%)</td>
<td>2 (0%)</td>
<td>20 (3%)</td>
<td>27 (4%)</td>
<td>10 (2%)</td>
</tr>
<tr>
<td>MB</td>
<td>0 (0%)</td>
<td>282 (82%)</td>
<td>40 (12%)</td>
<td>18 (5%)</td>
<td>5 (1%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>37 (6%)</td>
<td>6 (1%)</td>
<td>496 (75%)</td>
<td>104 (16%)</td>
<td>21 (3%)</td>
</tr>
<tr>
<td>Semen</td>
<td>12 (2%)</td>
<td>5 (1%)</td>
<td>86 (14%)</td>
<td>499 (80%)</td>
<td>24 (4%)</td>
</tr>
<tr>
<td>Skin</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>28 (3%)</td>
<td>4 (0%)</td>
<td>908 (97%)</td>
</tr>
</tbody>
</table>

Table 5.9: Classifications for 100 × 32 samples, multinomial logistic regression method

**Distribution of returned posterior probabilities**

The results in the previous section used a ‘majority rules’ classification rule for NB and MLR. This simply means that the class is assigned based on the highest posterior probability. Such an approach is common and acceptable for method comparison purposes, but it does discard some information. If we take a simple example where there are three possible classifications (say A, B and C), then there is an obvious quantitative difference between a result of (0.98, 0.01, 0.01) and (0.34,0.33,0.33). That is, in the first case the probability that the observation belongs to class A is 0.98, and that classification is 98 times more probable than the other two. By contrast, in the second example the probability is 0.34, and it is barely more probable (1.03 times) than the other two possibilities. We examine this behaviour in this section by considering the actual values of the posterior probabilities assigned to the ‘true’ cell type. Figure 5.2 shows the distribution the posterior probabilities that were assigned to the true cell types with boxplots.

These boxplots show that the methods return approximately the same range of posterior probabilities. It appears that the MLR method returns ‘high’ posterior probabilities more often for samples containing saliva compared to the NB method whereas the NB method returns ‘high’ posterior probabilities more often for samples containing menstrual blood, semen or skin compared to the MLR method. The boxplots look very similar for blood samples. Summary statistics for the posterior probabilities for the different cell types with the different methods are given in Table 5.10.

Table 5.10 also shows that the NB method performs poorly when compared to the MLR method on saliva samples. NB performs slightly better than MLR on skin and semen samples, and has similar performance for blood and menstrual blood samples. In legal practice it is common to compare hypotheses or evaluate evidence in ratio form. Therefore, one should be aware that the difference between the minimal posterior probabilities for skin with these methods (10^{-2} and 10^{-5}), should be considered as a relative difference. Although the absolute difference is small, the relative difference in posterior odds is 1000. The posterior probabilities returned by the MLR method appear to be of a range that allows for more discriminating probabilistic statements in odds form (both supporting as opposing the ‘true’ hypothesis).
5. A probabilistic approach for the interpretation of RNA profiles as cell type evidence

Figure 5.2: Boxplot of the posterior probabilities for the ‘true’ cell type for different cell types and methods - Lindenbergh 2012 data [12]. Here ‘⊙’ represents the median, the thick and thin lines are the box and whiskers respectively and a ‘◦’ represents an outlier.

<table>
<thead>
<tr>
<th>NB method</th>
<th>blood</th>
<th>menstrual blood</th>
<th>saliva</th>
<th>semen</th>
<th>skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>$10^{-3}$</td>
<td>$10^{-7}$</td>
<td>$10^{-5}$</td>
<td>$10^{-7}$</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>max</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>mean</td>
<td>0.93</td>
<td>0.82</td>
<td>0.64</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>median</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MLR method</th>
<th>blood</th>
<th>menstrual blood</th>
<th>saliva</th>
<th>semen</th>
<th>skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>$10^{-4}$</td>
<td>$10^{-15}$</td>
<td>$10^{-16}$</td>
<td>$10^{-12}$</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>max</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>mean</td>
<td>0.90</td>
<td>0.81</td>
<td>0.72</td>
<td>0.81</td>
<td>0.96</td>
</tr>
<tr>
<td>median</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5.10: Summary statistics for the posterior probabilities for the different cell types - NB and MLR methods - Lindenbergh data [12]

5.5.2 Data from Roeder et al. [17]

We also compared the marker value and threshold score method with the proposed probabilistic statement methods on the data from [17]. An overview of the data can be found in Table 5.16. It is not possible to include the $n/2$ scoring method on this dataset, since this method is specifically designed to work with cell type specific markers.
Most likely cell type of a known sample

Again, we are interested in the posterior probabilities from the different methods for samples of which we know the true cell type. Therefore, we assumed uniform prior probabilities for the NB method to be able to compare the outcomes. A summary of the identification statements obtained by using the marker value and threshold score method on all 182 samples\(^9\) is given in Table 5.11.

<table>
<thead>
<tr>
<th>True cell type</th>
<th>Blood</th>
<th>CVF</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen multiple with true</th>
<th>Semen multiple without true</th>
<th>Semen no class.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>25 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>CVF</td>
<td>0 (0%)</td>
<td>25 (57%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (1%)</td>
<td>0 (0%)</td>
<td>18 (41%)</td>
</tr>
<tr>
<td>MB</td>
<td>0 (0%)</td>
<td>6 (19%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (10%)</td>
<td>7 (23%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>49 (98%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Semen</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>29 (91%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (9%)</td>
</tr>
</tbody>
</table>

Table 5.11: Classifications for 182 samples, marker value and threshold score method

71% (25 + 25 + 2 + 49 + 29 out of 182) samples are correctly classified using the threshold scoring method (using the ‘old’ marker values and threshold scores). The only cell type for which the method reaches the score threshold is the true cell type. The true cell type is either the only classified cell type, or is part of the group of classified cell types, in 79% (143 of 182) samples.

A similar procedure can be used to evaluate the MLR and NB methods. A summary of results is given in Tables 5.12 and 5.13. The data set in this study is larger (182) so the training set consists of 145 samples and the testing set consists of 37 samples. Repeating this procedure 100 times gives 3700 results. Over these 100 sets, the average classification error for the NB method was 5.76% with an associated standard error of 3.14%. The average classification error for the MLR method was 7.22% with an associated standard error of 3.50%.

<table>
<thead>
<tr>
<th>True cell type</th>
<th>Blood</th>
<th>CVF</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>508 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>CVF</td>
<td>0 (0%)</td>
<td>804 (87%)</td>
<td>97 (11%)</td>
<td>0 (0%)</td>
<td>19 (2%)</td>
</tr>
<tr>
<td>MB</td>
<td>0 (0%)</td>
<td>92 (15%)</td>
<td>505 (85%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1061 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Semen</td>
<td>0 (0%)</td>
<td>5 (1%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>609 (99%)</td>
</tr>
</tbody>
</table>

Table 5.12: Classifications for 100 × 37 samples, naïve Bayes method

---

\(^9\)The dataset consisted of 193 samples. We removed 11 samples that were not tested on all markers.

\(^{10}\)This CVF sample was from a donor that declared time since intercourse was < 1 day. One could argue that this is not a false negative since the sample is most likely a mixture. Nonetheless, we decided to leave the sample in the dataset.
Table 5.13: Classifications for $100 \times 37$ samples, multinomial logistic regression method

Both the NB and the MLR method gave no ‘false positives or negatives’ when concerning saliva samples. This was also the case for NB and blood samples. In addition both of these methods correctly classify the vast majority of MB and CVF samples. In contrast the marker value and threshold score method has problems with these cases (see Table 5.11).

### Distribution of returned posterior probabilities

For the data from Roeder and Haas [17], the distribution of posterior probabilities is similar to that shown in Figure 5.2. Summary statistics for the posterior probabilities returned by both methods are given in Table 5.14.

<table>
<thead>
<tr>
<th>True cell type</th>
<th>Blood</th>
<th>CVF</th>
<th>MB</th>
<th>Saliva</th>
<th>Semen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>500 (98%)</td>
<td>0 (0%)</td>
<td>3 (1%)</td>
<td>0 (0%)</td>
<td>5 (1%)</td>
</tr>
<tr>
<td>CVF</td>
<td>0 (0%)</td>
<td>833 (91%)</td>
<td>67 (7%)</td>
<td>0 (0%)</td>
<td>20 (2%)</td>
</tr>
<tr>
<td>MB</td>
<td>8 (1%)</td>
<td>91 (15%)</td>
<td>493 (83%)</td>
<td>0 (0%)</td>
<td>5 (1%)</td>
</tr>
<tr>
<td>Saliva</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1061 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Semen</td>
<td>0 (0%)</td>
<td>65 (11%)</td>
<td>3 (1%)</td>
<td>0 (0%)</td>
<td>546 (89%)</td>
</tr>
</tbody>
</table>

Table 5.14: Summary statistics for the posterior probabilities for the different cell types - NB and MLR methods - Roeder data

These summarizing statistics support the use of these probabilistic methods for this data set. Almost all of the samples obtain a posterior probability that indicates the true cell type as the most likely cell type contained in the sample. Furthermore, in most cases, this probability is high relative to others. Empirical cumulative distribution functions for the odds between the probability assigned to the true cell type and the highest non true cell type are given in Figure 5.3 (multinomial logistic regression) and 5.4 (naive Bayes). These plots show that both methods generally give very discriminating results towards the true cell type.
Furthermore, the multinomial logistic regression has a larger range of odds, both supporting as opposing the true cell type. Similar plots were obtained for the data from [12]. The results from Roeder and Haas were more discriminating than those from Lindenbergh because their experiments used more markers. The categorical identification statement methods only use the evidence of pre-selected cell type specific markers to come to a conclusion regarding the cell type contained in a sample, whereas the probabilistic methods use the evidence of markers that amplified in combination with markers that did not amplify. Information regarding the markers that did not amplify is as important as information about those that did, as explained earlier. Methods that use this information, such as the two discussed (NB and MLR) will therefore perform better than those that do not.

![Empirical cumulative distribution function](image)

**Figure 5.3**: Empirical cumulative distribution function ($F(x)$) for the posterior odds of the true cell type and the highest non true cell type, [17] data, multinomial logistic regression method
5. A probabilistic approach for the interpretation of RNA profiles as cell type evidence

Figure 5.4: Empirical cumulative distribution function ($F(x)$) for the posterior odds of the true cell type and the highest non true cell type, [17] data, naïve Bayes method

5.6 Conclusion and discussion

The existing literature on the interpretation of RNA profiles focusses on categorical opinions like ‘identifying’ the cell type of the sample or to ‘determine’ whether a marker is cell type specific. However, as most of these papers also mention, RNA profiles depend on numerous variables that can have a substantial influence on the result. With forensic DNA evidence, it is common to report the (un)certainty that is contained in the evidence in the form of a likelihood ratio. Existing methods for the interpretation of RNA profiles do not give the opportunity to so. We have proposed two methods that give the opportunity to report the uncertainty about the evidential value of a RNA profile in the form of a probabilistic statement.

The first method, based on a Bayesian network, is straightforward. By making
the assumption that the amplification of different markers is conditionally independent one can multiply the evidential value of the different markers to obtain a likelihood ratio. The conditional independence assumption that is used in this naïve Bayesian approach was tested using data from [12] and [17]. This provided some, but not full, support for the assumption. When one wishes to avoid this assumption, the Bayesian network can be adapted to include dependencies.

The second method, based on multinomial logistic regression, is very straightforward to implement. It does need a conditional independence assumption. However, it the posterior probabilities obtained by using this method are based on a prior distribution that is implicitly based on the relative frequencies of the different cell types in the training database. When using a method based on multinomial logistic regression, it is common to use a random sample from the population of interest as training database, see [9]. However, in this setting, the training database is usually not based on a random sample of the relevant population. Samples of different cell types are selected based upon their forensic relevance. Hence, it is highly unlikely that the relative frequencies with which the cell types are present in the training database corresponds with the prior distribution over these cell types in casework. One can alter this implicit prior distribution by altering the values of $\beta_0$ subsequently (see [9]), however, this can be problematic and is a topic for further study.

A drawback of any method is that it implicitly assumes that it can only distinguish between the cell types it considers. For example, what happens when one encounters a semen sample if the method can only classify samples as having come from blood or saliva? When a cell type that is not contained in the method (or database used to train the method) would better explain the amplified markers, the likelihood ratio (or the categorical identification statement) might be misleading. Furthermore, in order to determine the value of the likelihood ratio when either of the two considered hypotheses is in a set of other sub-hypotheses (for example, $H_p$: the sample consisted of blood versus $H_d$: the sample did not consist of blood, it can be argued that $H_d$ is a set of sub-hypotheses like $H_d(1)$: the sample consisted of skin cells, $H_d(2)$: the sample consisted of semen cells, etc.) the prior probabilities of these (sub)hypotheses are needed in order to determine the LR, see equation (5.2). Reporting the LR as a formula might result in problems. This is important in practice because it is very unusual that a trier of fact states their choice for the prior probabilities. A practical solution to this may simply be to report a table of likelihoods for each cell type, and additionally, report the LR for the top two of the list, or make pairwise comparison LRs for those cell types deemed most relevant to the case at hand.

Both the MLR and the NB method show satisfying results when applied to real data. The proportion of samples of which the posterior probability is the highest for the ‘true’ cell type is substantially higher than for the $n/2$ scoring and the marker value and scoring threshold methods. Another advantage of these probabilistic methods is the opportunity to train these models on any dataset instantly\textsuperscript{11}. Furthermore, in these probabilistic methods, adding other factors that

\textsuperscript{11}For example, the $n/2$ scoring method is only useful when all markers are cell type specific, and the marker value and threshold score method uses a user defined threshold
influence the amplification rate of markers is very straightforward. For example, several studies have examined the influence of the amount of input material on the amplification rate of markers [4, 3, 5, 6, 7, 19]. If sufficient data is available, the amount of input RNA could be added as another unknown to the model. For the naïve Bayes method, the Bayesian network representing the model is given in Figure 5.5. The MLR method can also be adapted by adding the amount of input RNA as an explanatory variable to the model.

![Figure 5.5: Bayesian network for the evaluation of RNA profiles - input added](image)

The proposed probabilistic methods can also be adapted and trained on mixture data. This is an important feature since the majority forensic casework in which the cell type of the donated material is relevant consists of mixtures. However, before one can conveniently train these models, sufficient data should be gathered on the amplification rate of markers for samples consisting of a mixture of cell types.

We believe that, of the two methods that we propose, the method based on a naïve Bayes assumption is the better one in practice in this situation. Although the methods give similar results, it is easier to adapt the naïve Bayes method to more complex situations and it is not necessary to train a model for every ‘new’ set of hypotheses. Furthermore, the implicit prior distribution assumed by the MLR method, which is based on the relative frequencies in the training database, can lead to problems when applied in casework. Lastly, specialized software/routines (e.g. R ([15]/Matlab [13]) is needed to do MLR. Whilst this barrier is not insurmountable we have found that forensic scientists prefer methods that are easy to understand and explain in court. The NB method however can

---

12 We would like to stress that, when using these methods on another situation or with another dataset one should not carelessly use these methods straight away. For example, it is possible that the markers in this set are completely conditionally dependent.
Apart from these two methods, there are numerous other potential methods for the probabilistic interpretation of RNA profiles. However, in this research, the focus lies on proposing probabilistic methods as an improvement to existing categorical statement methods. Further research could examine which probabilistic method is ‘best’.

Bibliography


5. A probabilistic approach for the interpretation of RNA profiles as cell type evidence


## 5.A Overview of the used marker data

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>18S-rRNA</th>
<th>ACTB</th>
<th>GAPDH</th>
<th>HBB</th>
<th>AMICA1</th>
<th>CD93</th>
<th>KRT4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>32</td>
<td>25</td>
<td>31</td>
<td>29</td>
<td>30</td>
<td>27</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>MB</td>
<td>16</td>
<td>16</td>
<td>5</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Saliva</td>
<td>32</td>
<td>27</td>
<td>17</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>Semen</td>
<td>30</td>
<td>24</td>
<td>18</td>
<td>24</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Skin</td>
<td>48</td>
<td>46</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>SPRR2A</th>
<th>KRT13</th>
<th>STATH</th>
<th>HTN3</th>
<th>PRM1</th>
<th>SEMG1</th>
<th>MMP7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MB</td>
<td>16</td>
<td>14</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Saliva</td>
<td>32</td>
<td>18</td>
<td>17</td>
<td>20</td>
<td>22</td>
<td>1</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Semen</td>
<td>30</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Skin</td>
<td>48</td>
<td>22</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>MMP11</th>
<th>CDSN</th>
<th>LOR</th>
<th>HBD1</th>
<th>MUC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>32</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>MB</td>
<td>16</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Saliva</td>
<td>32</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Semen</td>
<td>30</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Skin</td>
<td>48</td>
<td>0</td>
<td>45</td>
<td>44</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.15: Number of samples amplified on each marker per multiplex using a threshold of 150 rfu, part of the data from [12].
### Table 5.16: Number of samples amplified on each marker per multiplex. Data from [17] 13.

#### (A) Se multiplex

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>PRM1</th>
<th>PRM2</th>
<th>KLK3</th>
<th>SEMG1</th>
<th>TGM4</th>
<th>UCE</th>
<th>TEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>MB</td>
<td>33</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>Saliva</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>CVF</td>
<td>44</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>semen</td>
<td>32</td>
<td>26</td>
<td>19</td>
<td>25</td>
<td>31</td>
<td>28</td>
<td>24</td>
<td>31</td>
</tr>
<tr>
<td>Sweat</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Skin</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> The positive results were obtained from a single MB sample which had TSI time < 1 day

#### (B) SaCVF multiplex

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>HBD1</th>
<th>MUC4</th>
<th>Lcris</th>
<th>Ljen</th>
<th>Lcris2</th>
<th>Lgas</th>
<th>STATH</th>
<th>HTN3</th>
<th>PRB4</th>
<th>MUC7</th>
<th>SMR3B</th>
<th>UCE</th>
<th>TEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>MB</td>
<td>33</td>
<td>13</td>
<td>22</td>
<td>17</td>
<td>27</td>
<td>19</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>Saliva</td>
<td>50</td>
<td>4</td>
<td>13</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td>50</td>
<td>16</td>
<td>48</td>
<td>49</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>CVF</td>
<td>49</td>
<td>15</td>
<td>40</td>
<td>25</td>
<td>30</td>
<td>25</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>semen</td>
<td>32</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>Sweat</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Skin</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>b</sup> The positive results were obtained from a single CVF sample which had TSI time < 1 day

#### (C) Bl multiplex

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>HBB</th>
<th>ALAS2</th>
<th>PRF1</th>
<th>GlycoA</th>
<th>PF4</th>
<th>SPTB</th>
<th>PBGD</th>
<th>UCE</th>
<th>TEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>MB</td>
<td>33</td>
<td>32</td>
<td>25</td>
<td>27</td>
<td>12</td>
<td>25</td>
<td>12</td>
<td>1</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>Saliva</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>35</td>
</tr>
<tr>
<td>CVF</td>
<td>49</td>
<td>8</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>44</td>
<td>43</td>
</tr>
<tr>
<td>semen</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Sweat</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Skin</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### (D) MB multiplex

<table>
<thead>
<tr>
<th>Sample type</th>
<th>No. of samples analysed</th>
<th>MSX1</th>
<th>LEFTY2</th>
<th>SFRP4</th>
<th>MMP7</th>
<th>Hs202072</th>
<th>MMP10</th>
<th>MMP11</th>
<th>UCE</th>
<th>TEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>MB</td>
<td>35</td>
<td>19</td>
<td>16</td>
<td>10</td>
<td>34</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>18</td>
<td>25</td>
</tr>
<tr>
<td>Saliva</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>CVF</td>
<td>49</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>47</td>
<td>48</td>
</tr>
<tr>
<td>semen</td>
<td>32</td>
<td>8</td>
<td>2</td>
<td>12</td>
<td>17</td>
<td>1</td>
<td>5</td>
<td>12</td>
<td>29</td>
<td>31</td>
</tr>
<tr>
<td>Sweat</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Skin</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
5.B p-values overview

Here, the p-values obtained by a Fisher exact test to test for conditional independence between markers given the cell type of the sample are given.

<table>
<thead>
<tr>
<th></th>
<th>housekeeping</th>
<th>blood</th>
<th>mucosa</th>
<th>saliva</th>
<th>semen</th>
<th>menst. secr.</th>
<th>skin</th>
<th>vag. mucosa</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S-rRNA</td>
<td>1.00</td>
<td>0.56</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.65</td>
<td>1.00</td>
</tr>
<tr>
<td>ACTB</td>
<td>1.00</td>
<td>0.61</td>
<td>0.28</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.07</td>
<td>1.00</td>
</tr>
<tr>
<td>GAPDH</td>
<td>0.56</td>
<td>0.61</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.14</td>
<td>1.00</td>
</tr>
<tr>
<td>HBB</td>
<td>0.01</td>
<td>0.28</td>
<td>1.00</td>
<td>1.00</td>
<td>0.26</td>
<td>1.00</td>
<td>0.05</td>
<td>1.00</td>
</tr>
<tr>
<td>AMICA1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CD93</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>KRT4</td>
<td>0.52</td>
<td>0.28</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.09</td>
<td>0.55</td>
</tr>
<tr>
<td>SPRR2A</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.09</td>
<td>1.00</td>
<td>0.39</td>
<td>1.00</td>
</tr>
<tr>
<td>KRT13</td>
<td>0.39</td>
<td>0.07</td>
<td>0.14</td>
<td>0.05</td>
<td>1.00</td>
<td>0.55</td>
<td>0.39</td>
<td>1.00</td>
</tr>
<tr>
<td>STAT3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HTN3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.46</td>
<td>1.00</td>
<td>0.04</td>
<td>1.00</td>
</tr>
<tr>
<td>PRM1</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.04</td>
<td>1.00</td>
<td>0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>SEMG1</td>
<td>0.65</td>
<td>0.02</td>
<td>0.04</td>
<td>0.56</td>
<td>0.56</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>MMP7</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MMP11</td>
<td>0.00</td>
<td>0.11</td>
<td>0.55</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>CDSN</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>LOR</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HBD1</td>
<td>0.38</td>
<td>0.52</td>
<td>1.00</td>
<td>0.18</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MUC4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5.17: p values for fisher exact test on contingency tables regarding the amplification on markers for semen data only, ignoring the almost no output samples (33 samples), Lindenbergh data [12]

\[Based on Table 2 of [17], with kind permission from Springer Science and Business Media\]
<table>
<thead>
<tr>
<th></th>
<th>housekeeping</th>
<th>blood</th>
<th>mucosa</th>
<th>saliva</th>
<th>semen</th>
<th>menst. secr.</th>
<th>skin</th>
<th>vag. mucosa</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S-rRNA</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ACTB</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>GAPDH</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.42</td>
<td>0.32</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HBB</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>AMICA1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CD93</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>KRT4</td>
<td>1.00</td>
<td>1.00</td>
<td>0.42</td>
<td>1.00</td>
<td>1.00</td>
<td>0.05</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SPRR2A</td>
<td>0.49</td>
<td>1.00</td>
<td>0.32</td>
<td>0.08</td>
<td>0.12</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>KRT13</td>
<td>1.00</td>
<td>1.00</td>
<td>0.53</td>
<td>1.00</td>
<td>1.00</td>
<td>0.05</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>STAT3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HTN3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PRM1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SEMG1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MMP7</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MMP11</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>CDSN</td>
<td>0.12</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.23</td>
</tr>
<tr>
<td>LOR</td>
<td>0.16</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HBD1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>MUC4</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 5.18: p values for fisher exact test on contingency tables regarding the amplification on markers for skin data only (48 samples), Lindenbergh data [12]
Cell type determination and association with the DNA donor

Abstract

In forensic casework, evidence regarding the type of cell material contained in a stain can be crucial in determining what happened. For example, a DNA match in a sexual offense can become substantially more incriminating when there is evidence supporting that semen cells are present.

Besides the question which cell types are present in a sample, also the question who donated what (association) is very relevant. This question is surprisingly difficult, even for stains with a single donor. The evidential value of a DNA profile needs to be combined with knowledge regarding the specificity and sensitivity of cell type tests. This, together with prior probabilities for the different donor-cell type combinations, determines the most likely combination.

We present a Bayesian network that can assist in associating donors and cell types. A literature overview on the sensitivity and specificity of three cell type tests (PSA test for seminal fluid, RSID saliva and RSID semen) is helpful in assigning conditional probabilities. The Bayesian network is linked with a software package for interpreting mixed DNA profiles. This allows for a sensitivity analysis that shows to what extent the conclusion depends on the quantity of available research data. This can aid in making decisions regarding further research.

It is shown that the common assumption that an individual (e.g. the victim) is one of the donors in a mixed DNA profile can have unwanted consequences for the association between donors and cell types.

This chapter was originally published as: J. de Zoete, W. Oosterman, B. Kokshoorn, M. Sjerps, Cell type determination and association with the DNA donor. Forensic Science International: Genetics, http://dx.doi.org/10.1016/j.fsigen.2016.08.004, 2016.
### 6.1 Introduction

Biological evidence in forensic casework can often be used to help address two different types of questions. (1) **Who donated the material?** and (2) **What cellular material(s) does the sample contain?** The second question becomes especially relevant when both the prosecution and the defense hypothesis state that the suspect donated the material but dispute the type of material. For example, in a case regarding an alleged sexual offense, a DNA match with a suspect usually is more incriminating when the sample contains semen than skin cells. The association between cell type results and DNA results is not necessarily straightforward. Extra care should be taken in situations which involve mixtures. Gill [15] presents a 2011 case where a man was wrongly charged with rape due to an incorrect assumption that the positive semen test could be associated with a certain donor. Before the actual trial, it was shown that the positive DNA test was due to a contamination with this man’s saliva. Such an unfounded association between a cell type result and a (potential) donor has been labeled an ‘association fallacy’, see [15].

Taylor et al. [53] present a Bayesian network approach that can assist in addressing evaluative questions like “**Is Mr A the source of the blood and the DNA or is he the source of the DNA only?**” The focus lies on combining DNA results with the results of a Hemastix test (a presumptive test for blood) as well as observations of the colour of the stain. Although it is possible to include tests for other cell types, this part is not included in the paper. It is shown how Bayesian networks can assist in making a probabilistic statement when answering questions like **Is it possible that this stain is blood from someone other than the victim over some trace level of DNA from the victim that has been deposited onto the item by a previous chance encounter?**

The Bayesian network presented in Oosterman et al. [35] also assists in combining evidence regarding who donated the material with evidence relevant for the question what kind of material was donated. This network does not concentrate on a specific cell type and can serve as a basis for any cell type-donor situation.

---

**Figure 6.1:** Network for the association of DNA and cell type. The association is effectuated by considering all possible DNA-cell type combinations in the oval configuration nodes, adapted from [35].

In this paper, we will extend the work from Taylor et al. [53] using the Bayesian network structure from Oosterman et al. [35], shown in Figure 6.1. We discuss the kind of data that is needed to evaluate cell type results and present two mock case examples that illustrate the possibilities and pitfalls of such an approach. We
chose to use mock cases typical for real casework instead of real cases in order to avoid the legal complication associated with the use of real cases. The first mock case example focuses on the questions whether semen cells were present in a crime sample combined with who the most likely donor(s) were. A literature overview regarding the specificity and sensitivity of two tests for seminal fluid (RSID semen and PSA test) and a saliva test (RSID saliva) is presented that can be used to assign conditional probabilities for the corresponding nodes in the network. It is shown that the common assumption that an individual is a donor of a mixed DNA profile has unwanted consequence for the association between donors and cell types. The second mock case example focuses on the limitations of this approach. Here, other sources of information regarding the association between donor and cell type are discussed.

Furthermore, the Bayesian network is linked with the R [41] package forensim [17] to allow for more complex mixture calculations. Lastly, distributions reflecting the uncertainty of certain parameter values in the conditional probability tables are used to analyze the sensitivity of the results from the network. Recommendations regarding data collection for assigning the necessary conditional probabilities are given in the discussion.

Thus, in comparison with Taylor et al. [53], we add three things. Firstly we add a literature overview that can assist in assigning probabilities for three different common cell type tests. Secondly, the Bayesian network is linked with forensim a software package for complex mixture calculations. Thirdly, we present a sensitivity analysis for examining the influence of the evidential value of the DNA profile and the specificity of the cell type tests on the posterior probabilities of interest.

There is an ongoing discussion in the forensic literature about dealing with uncertainty when assigning numerical values to probabilities or likelihood ratios [52, 51]. Here, we consider probabilities to have “true” values that can be estimated. A so-called full-Bayesian approach is based on a different point of view and would lead to a different type of data analysis. We leave this as a possible topic for further research.

6.2 Literature overview

In order to assess the evidential value of a positive or negative result of a ‘classical’ test for cell type determination (e.g. Hemastix test, RSID saliva), the specificity and the sensitivity of that test has to be determined. A forensic laboratory can perform experiments in order to estimate these probabilities. Furthermore, in-house (validation) experiments of the manufacturer or studies by other researchers can be used. How such a literature survey can be approached is demonstrated here by an overview of studies on the reactivity of the PSA test. A distinction is made between circumstances that can influence the test outcome, such as the type of sample (stain or intimate swab). In Section 6.3 this overview is used to estimate conditional probabilities. Furthermore, reasonable probability boundaries for use within the sensitivity analysis are determined from this overview. The selected
literature, the methodology used and the number of positive and negative results are summarized in Table 6.10. For the RSID semen and RSID saliva tests, similar overviews can be found in Table 6.11 and 6.12.

6.2.1 The prostate specific antigen test

Prostate specific antigen (PSA) is a protein which is secreted by the prostate gland and is therefore present in high concentrations in seminal fluid. It was initially developed as a test for prostate cancer. In the past decades PSA has been found in various male and female bodily fluids. Low amounts of PSA have been found in blood, saliva, vaginal secretions, breast milk, breast tissue, female urine, in the periurethral glands, the endometrium and in amniotic fluid [24]; while high concentrations have been found in semen, male urine and in the blood of prostate carcinoma patients.

A common PSA immunoassay, the Seratec® Semiquant, can detect PSA down to around 2 ng/mL, but at 0.5 ng/mL a weak positive test result may still be obtained [48]. Despite intrapersonal variation in PSA concentration, no individuals with PSA concentrations equal or lower than the test’s limit of detection have been reported [25, 19]. Samples containing at least a few nano litres of semen are expected to give a positive test result. In the following sections the reactivity of the PSA test with different cell types and sample types will be discussed. An overview of these values with accompanying credible intervals can be found in Table 6.3.

6.2.2 Effect of sample type, time and temperature

The reactivity of the PSA test differs for different sample types. Most seminal samples that are encountered in casework can be divided into one of two categories: stains and intimate swabs. The type of sample can have an influence on the outcome of the test.

175 out of 176 samples returned a positive result in one study where samples of liquid semen were used [22]. However when these samples were kept at 37°C for 48 hours, only 169 out of 176 samples returned a positive result. Proteins in liquid seminal samples that are kept at high temperatures might degrade after time. The results of this study have not been replicated. Additional testing is required for a more generalized statement about PSA degradation in liquid seminal samples.

Test sensitivity with seminal stains is generally reduced compared to liquid samples because of the non-perfect extraction efficiency (<100%) of stains. In Denison et al. [12], semen stains gave a positive result nine out of ten samples and in another study five out of five stains were tested positive [23]. On longer time scales, fabrics seem to be a stable matrix for storing semen (room temperature): positive results have been reported in stains ranging from 2 to 30 years old (n = 10) [19].

A large proportion of the forensic casework with semen concerns intimate swabs. Because semen persists only for a limited amount of time in the vagina, the reliability of the PSA test with post-coital swabs strongly depends on the time that has elapsed between ejaculation and sampling. In a study where females were
exposed to varying amounts of their partner’s semen, the elevated levels of PSA were found to decline rapidly over time [26]. Concentrations were still elevated at 24 hours but after 48 hours base-line levels were reached. Exposure to higher amounts of semen was detectable for a longer time. Since in this study only 1 mL of semen was used and the volume of male ejaculate can range up to about 11 milliliters (mean = 3.2 mL) [42], it is not unlikely that elevated PSA levels could persist in the vagina for longer times. This view is supported by a study where PSA was on average still detectable at 27 hours post-coitus [16]. When post-coital swabs were tested for PSA, the number of positive reactions decreased over time. No negative reactions occurred in the 0-6 hour segment (n = 18) and at the 60 hour mark no more positive reactions were observed [2]. In one study a weak positive reaction from a post-coital swab was still observed 70 hours after intercourse [4]. The exact relation between time and test outcome is difficult to determine, since the probability of a positive test is likely strongly dependent on case circumstances (did the victim wash, sampling technique, was the victim alive etc.). The probability of a positive test with postcoital swabs will have to be decided on a case to case basis, based on case circumstances and elapsed time.

In a similar study with samples simulating oral sexual assault, vomit samples were created by adding semen to gastric juice [30]. PSA was detected in neat samples and samples that were deposited on a ceramic plate after incubation up to 4 hours (n = 54). However, PSA was detected in samples which were spotted on cotton cloth up to only 15 minutes of incubation time. Neat samples can likely be regarded as diluted liquid semen, but with stains a false negative result should be expected.

6.2.3 Cross reactivity of the PSA test

An overview of the cross reactivity of the PSA test is displayed in Table 6.3. Below we will discuss in more detail which cell types/substances have resulted in (occasional) positive results.

**Male urine**

High concentrations of PSA are present in the urine of adult males. With a maximum reported value of around 800 ng/mL [45] there is a large overlap with the range of PSA concentrations found in semen, making the interpretation of samples that might contain urine highly complex. In the PSA test manual [48] it is advised to dilute a positive sample 200 times to differentiate between urine and semen. A semen sample should still give a positive result, urine a negative result. However, since the amount of material is difficult to estimate in forensic samples, such a control cannot provide a definite answer.

In the urine of young children the PSA concentration is low or absent. This is still the case with 10 and 11 year old boys (n = 19) [45]. When boys enter puberty this changes and after the thirteenth year PSA is at adult levels and neat urine samples often give positive test results even when diluted 100 times [43].

Urine that is deposited on fabrics usually gives positive results [12, 45], unless
diluted [45] or if a very small amount (25 mm²) of fabric is used [37]. It seems the reactivity of male urine is strongly dependent on the amount of material that is used with the test. Moreover, different sample preparation protocols garner different results, for example the dilution of a sample is of influence. Neat samples give positive results almost invariably, while stains and diluted samples have varying results.

Test results from samples that could contain male urine can be very difficult to interpret because of the unpredictable test behavior. In our example (Section 6.3), the Bayesian network combines all of these into one category, ‘male urine’. This is a crude approach; one could also inform probabilities for separate trace types (liquid, stain etc.), however the available data is too limited for this. Nonetheless, when data is available, we advise to adapt the network to subdivide this category.

**Vaginal fluid and female urine**

PSA concentrations in vaginal secretions are reported up to a maximum of around 5.0 nl/mL [26]. The PSA concentration in female urine is of a similar level: average reported values range from 0.29 [47] to 3.72 [7] nl/mL. Consequently PSA in vaginal fluid and female urine may be detected by the PSA test based on its limit of detection. This is confirmed by a positive reaction with urine and vaginal fluid with one PSA-high individual during and three days prior to the menstrual cycle [12]. This may be due to the fluctuation of PSA levels with the menstrual cycle. PSA levels in serum were shown to be produced in a cyclical manner with the highest concentrations occurring during active menstruation [56]. Whether this is also the case with urine and vaginal fluid has not been confirmed.

**Blood and saliva**

Reported concentrations of PSA in blood and saliva are generally very low (<1 ng/mL) [14, 28]. Male blood can have higher concentrations up to 4 ng/mL [48], but since samples are diluted during preparation, no instances of false positives have yet been reported. A confounding factor is blood from prostate cancer patients which can contain high PSA levels (<200 ng/mL) [48].

**Breast milk**

Directly after birth very high PSA concentrations in breast milk are possible (up to 2100 ng/mL), but this rapidly declines to lower levels after a few days [25]. Most of the breast milk samples contain around 1 ng/mL PSA but large individual differences were observed. One positive result, [12], was reported in the examined literature out of a total of eighteen samples.

**Feces and sweat**

PSA has never been detected in feces and sweat. 30 (feces) and 33 (sweat) samples respectively all tested negative with the PSA test, [43, 37, 19, 4].

Other substances

In a study of Khaldi et al. [22] using Vedalab® PSA tests 6 out of 102 water samples tested positive, indicating a relatively high likelihood of a false positive when no cells are present. These results show a low reliability of this test, but they are at odds with several (control) experiments on the Vedalab® PSA tests that did not show any (false) positive results. Due to confirmed absence of PSA and neutral pH it is common to include water as a negative control sample when doing casework. Hence, forensic laboratories potentially have a lot of data regarding the cross reactivity of the PSA test with water (the same holds for other tests). Consequently a positive reaction with a water sample has to be explained by a defective test, demonstrating the base-line reliability of a test.

Internal validations of the Seratec® PSA test demonstrated that test result can be influenced by the pH value of sample material. Only low pH (< 5) values of organic acids such as citric acid can cause the false positives; these can usually be recognized as a false result due to a spotted or non-uniformly formed test line [48]. Extreme pH values in general can disrupt antigen-antibody interactions by weakening of hydrogen bonds. The manufacturer recommends adjustment of the pH to a neutral value.

Of various brands of condoms with lubricants and contraceptives, only a condom with the nonoxynol-9 spermicide gave a positive test result in one study [3], other similar studies [4, 37] did not observe a positive in this situation. In one study liquid bleach and Delfen® contraceptive foam gave weak positive results; however the positive results could not be repeated [12]. Some caution with the interpretation of sampled condoms seems warranted since the brand of condom is usually unknown in casework.

No false positives with the following animals were reported: dog, pig, horse, cat, chicken, cow, sheep and various micro organisms [19, 12, 37, 31]. While the PSA Seratec Semiquant seems to be human specific, to our knowledge no experiments with semen of other primates have been performed.

Finally, mixtures of different body fluids could demonstrate new reaction behavior. Semen was tested together with other body fluids in various mixtures of two and three different substances. The detection of seminal fluid does not seem to suffer from interference in these mixtures [12, 4, 43, 37]. However, in one instance when semen was mixed with caustic soda a false negative result was obtained [23]. This could be explained by the extremely low pH of caustic soda (NaOH). It is not entirely clear how a mixture of body fluids affects the reactivity of semen, since neat semen can also produce a false negative result occasionally. The results of these studies support the hypothesis that the ‘most reactive’ body fluid in a mixture determines the reactivity of the whole sample. This is at odds with the hypothesis that through dilution the mixture obtains a new PSA concentration which then determines the probability of a positive test. A third hypothesis would be that the test can react with either fluid in a mixture, causing a higher reactivity in mixtures. At this moment it is not clear which hypothesis is true.
6. Cell type determination and association with the DNA donor

6.3 Mock case example

The following mock case example, which is a typical case that may be encountered in any lab, shows how the Bayesian network can assist in forensic casework. The conditional probabilities for a positive test result using the RSID semen, the PSA test and the RSID saliva test are estimated using the literature overview. Conditional probabilities for a positive Hemastix test are obtained from Taylor et al. [53]. The forensim package [17] is integrated with the BN to allow for mixture calculations and the statistical software package R [41] is used to perform a sensitivity analysis.

Mock case description #1 A woman was orally sexually assaulted in an alley. A suspect was apprehended based on a witness statement. The victim\(^1\) testifies that, after the offender had ejaculated, she spat out the semen. The suspect refuses to give a statement. A sample for investigation of cell type and DNA is obtained from the location pointed out by the victim. The results of a forensic investigation are: a partial mixed DNA profile, a negative Hemastix test for blood, a positive RSID saliva test, a positive RSID semen test and a positive PSA test for seminal fluid.

We assume that the DNA profiles observed in this case are as from Example 3 in [13], see Table 6.1\(^2\). The allele frequencies obtained from the Caucasian population data for the NGM STR loci from Budowle et al. [8] are used as background population data.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Evidence</th>
<th>Victim</th>
<th>Suspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3S1358</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>VWA</td>
<td>14</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>D16S539</td>
<td>9</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>D2S1338</td>
<td>17</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>D8S1179</td>
<td>10</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>D21S11</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>D18S51</td>
<td></td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>D19S433</td>
<td>13</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>TH01</td>
<td>6</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>FGA</td>
<td>20</td>
<td></td>
<td>22</td>
</tr>
</tbody>
</table>

Table 6.1: Alleles in evidence, victim and suspect

It this example, eight cell type categories (and a ‘none/water’ category) are considered, blood, breast milk, feces, saliva, semen, sweat, urine and vaginal secretion. These were selected because we obtained data regarding the behaviour of the PSA, RSID saliva and RSID semen test on these cell types from the literature overview. If data on other substances is available we advise to add these as separate categories.

\(^1\)In sexual offense cases the reliability of the testimony of the victim can be questioned. However, in this example, we will refer to this person as the victim.

\(^2\)Reprinted from [13] with permission from Elsevier
It is assumed that a maximum of two people contributed to the crime stain. This results in a set of 13 donor configurations, see Table 6.2. These correspond with the states of the Individual(s) present in trace node, see Figure 6.1.

<table>
<thead>
<tr>
<th>#</th>
<th>configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
</tr>
<tr>
<td>2</td>
<td>suspect</td>
</tr>
<tr>
<td>3</td>
<td>victim</td>
</tr>
<tr>
<td>4</td>
<td>unknown man</td>
</tr>
<tr>
<td>5</td>
<td>unknown woman</td>
</tr>
<tr>
<td>6</td>
<td>suspect + victim</td>
</tr>
<tr>
<td>7</td>
<td>suspect + unknown man</td>
</tr>
<tr>
<td>8</td>
<td>suspect + unknown woman</td>
</tr>
<tr>
<td>9</td>
<td>victim + unknown man</td>
</tr>
<tr>
<td>10</td>
<td>victim + unknown woman</td>
</tr>
<tr>
<td>11</td>
<td>unknown man + unknown man</td>
</tr>
<tr>
<td>12</td>
<td>unknown man + unknown woman</td>
</tr>
<tr>
<td>13</td>
<td>unknown woman + unknown woman</td>
</tr>
</tbody>
</table>

Table 6.2: Donor configurations

6.3.1 Estimating conditional probabilities for classical cell type determination tests using literature

There are several points to consider when assigning conditional probabilities using a literature study. Firstly, the experimental setup of samples that are used to estimate conditional probabilities should match the protocol of the crime laboratory using the Bayesian network, since these probabilities differ per protocol. Especially the extraction protocol has a large influence on the proportion of the material that is wasted, see [48]. The extraction efficiency of a kit strongly depends on the type of buffer utilized. The sample type (type of fabric, brand of swab etc.) also affects the extraction efficiency. Secondly, the difficulty level of casework samples may differ between laboratories. The set of samples that is used to base the conditional probabilities on should be representative for the specific situation of the examined crime. For this case example, we pool the data from literature as background data. We only use this set for illustrative purposes.

Some cell types always returned a negative result within our dataset (see Table 6.3, for example, if we pool all the studies regarding samples containing blood, zero samples returned a positive out of 27 samples. This suggests that \( \Pr(\text{PSA pos|blood}) = 0 \). Using zeros or ones as conditional probabilities regarding whether a test will return a positive/negative result leaves very little flexibility to the probabilistic model. Furthermore, one can question whether the ‘actual’ probability really is zero in such a situation. For example, three out of four studies found no positive results using the PSA test on samples containing breast milk. The other study did. So, care should be taken with cell types that never or always returned a positive result. Therefore, similar to the procedure in Taylor et al. [53], the conditional probabilities for cell type tests are based on a Dirichlet distribution.
with a uniform prior. For example, \( \Pr(\text{PSA pos} | \text{blood}) \) is calculated as

\[
\Pr(\text{PSA pos} | \text{blood}) = \frac{n_{\text{pos,blood}} + 1}{K + n_{\text{blood}}} = \frac{0 + 1}{2 + 27} = 0.034
\]

Here, \( K \) is the number of states (2, positive and negative) of the PSA test. \( n_{\text{pos,blood}} \) is the number of positive PSA tests on blood samples in our dataset. Similarly,

\[
\Pr(\text{PSA pos} | \text{semen}) = \frac{n_{\text{pos,semen}} + 1}{K + n_{\text{semen}}} = \frac{373 + 1}{2 + 382} = 0.974
\]

An overview of the estimated conditional probabilities and their 95\% (equal-tailed) credible intervals can be found in Table 6.3.

| Cell type       | positive/total samples | \( \Pr(\text{PSA pos} | \text{cell type}) \) | 95\% CI        |
|-----------------|------------------------|---------------------------------|----------------|
| Blood           | 0/27                   | 0.034                           | (0.001, 0.123) |
| Breast milk     | 1/18                   | 0.100                           | (0.013, 0.260) |
| Feces           | 0/30                   | 0.031                           | (0.001, 0.112) |
| Saliva          | 0/48                   | 0.002                           | (0.001, 0.072) |
| Semen           | 373/382                | 0.974                           | (0.956, 0.987) |
| Sweat           | 0/33                   | 0.029                           | (0.001, 0.103) |
| Urine (male)    | 55/96                  | 0.571                           | (0.473, 0.667) |
| Urine (female)  | 1/55                   | 0.035                           | (0.004, 0.096) |
| Vaginal secretion| 2/175                  | 0.017                           | (0.004, 0.040) |
| None/Water      | 6/102                  | 0.067                           | (0.028, 0.122) |

Table 6.3: Overview of the estimated conditional probabilities for the PSA test and their 95\% credible intervals.

Similar tables for the RSID semen and the RSID saliva test can be found in Table 6.13 and 6.14. For the Hemastix test, data from Taylor et al. [53] is used. Since this paper only considered the blood, saliva, semen and none categories data is missing to estimate the necessary probabilities. For illustrative purposes the conditional probability of a positive Hemastix test for the remaining cell types is assumed to be 0.05, see Table 6.15. Since we have insufficient information regarding how these tests react on mixtures of cell types, it is assumed here, for illustrative purposes, that,

\[
\Pr(\text{test pos} | \text{cell type 1, cell type 2}) = \max \Pr(\text{test pos} | \text{cell type 1}), \Pr(\text{test pos} | \text{cell type 2})
\]

In words; the probability that a tests returns a positive result when the sample contains multiple cell types is equal to the maximum of the probabilities that this test gives a positive result for the individual cell types. The validity of this assumption can be tested if enough data is available\(^3\). The conditional probability table for the PSA-test node is given in Table 6.4.\(^4\)

\(^3\)Alternatively, another model can be used to assign conditional probabilities regarding a positive test result when the sample consists of a mixture of cell types

\(^4\)The conditional probability tables for the other cell type tests have the same structure.
6.3. Mock case example

<table>
<thead>
<tr>
<th>cell type(s) present in trace</th>
<th>blood</th>
<th>breast milk</th>
<th>none/water</th>
<th>blood, breast milk</th>
<th>blood, feces</th>
<th>urine (f), vag. sec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>0.034</td>
<td>0.100</td>
<td>0.067</td>
<td>0.100</td>
<td>0.034</td>
<td>0.035</td>
</tr>
<tr>
<td>negative</td>
<td>0.966</td>
<td>0.900</td>
<td>0.933</td>
<td>0.900</td>
<td>0.966</td>
<td>0.965</td>
</tr>
</tbody>
</table>

Table 6.4: Conditional probability table for the PSA-test node

In Taylor et al. [53], conditional probabilities that follow from data obtained in a set of experiments are presented in such a way that it seems that they can be used by other forensic practitioners. In the absence of other data this is a reasonable approach. However, the sensitivity and specificity of tests is dependent on case specific circumstances and can even be lab-dependent. Furthermore, the majority of samples that were used to assign conditional probabilities in Taylor et al. [53] came from forensic casework. This is a representative dataset when the only information regarding a case is that it was submitted for forensic casework. When more background information is available practitioners that use such a probability model should consider to perform specificity and sensitivity tests in their own lab, preferably with a protocol that follows the protocol used for the crime stain of interest. Literature can assist in pointing out unlikely or unexpected positive/negative results. Furthermore, if the in house results follow the published results from another lab/researcher, these results can be used to substantiate the value of the assigned conditional probability.

6.3.2 Integration of LR mix with the BN for mixture calculations

In Taylor et al. [53], DNA evidence is inserted as a likelihood ratio that discriminates between two hypotheses: The POI and an unknown individual are the sources of DNA ($H_1$) and Two unknowns are the sources of DNA ($H_2$). In many situations, the set of hypotheses concerning the donor(s) of a crime stain exceeds two. Furthermore, it is common that the likelihood of observing the DNA profile from a (potentially) mixed stain is determined with the aid of dedicated software packages like [1, 40, 39]. The Bayesian network (Figure 6.1) contains a node DNA interpretation software that allows the user to insert the likelihoods obtained from such a software package as evidence in the network. Likelihoods for observing the DNA profile from the crime stain for the different donor configurations (Table 6.2) are needed. In this example, the forensim package [17] is used for mixture calculations\(^5\). Since this package is implemented in R [41], it is possible to link the mixture software with the Bayesian network for direct calculations. The R [41] package gRain [20] can be used to load a .net Bayesian network file into R. The R code to perform these steps for this (and other similar) example is provided as

\(^5\)Alternatively, other methods can be used to do the mixture calculations. Here, we use forensim since it is freely available in R. When using another method for mixture calculations, the likelihoods from Table 6.5 should be obtained. These can be entered in the conditional probability table.
supplementary material to this paper\textsuperscript{6}.

\textbf{Figure 6.2:} DNA interpretation software component of the network in Figure 6.1

The likelihoods of observing the DNA profile from Table 6.1 given the donor combinations from Table 6.2, as computed by the \textit{forensim} \textsuperscript{[17]} package, assuming,

\begin{align*}
\Pr(\text{contamination}) & = 0.05, \\
\Pr(\text{drop out}) & = 0.10, \\
\theta & = 0,
\end{align*}

are given in Table 6.5. The \textit{forensim} package does not take information on the AMEL locus into account. Hence, it is not possible to distinguish between male and female donors. Nonetheless, specific AMEL information can be included in the model. For example, if the AMEL locus shows a peak at both $X$ and $Y$, the likelihoods for configurations without a male donor could be set to 0.

<table>
<thead>
<tr>
<th>configuration</th>
<th>likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>$3.9 \cdot 10^{-52}$</td>
</tr>
<tr>
<td>suspect</td>
<td>$2.0 \cdot 10^{-19}$</td>
</tr>
<tr>
<td>victim</td>
<td>$1.2 \cdot 10^{-30}$</td>
</tr>
<tr>
<td>unknown man</td>
<td>$4.1 \cdot 10^{-26}$</td>
</tr>
<tr>
<td>unknown woman</td>
<td>$4.1 \cdot 10^{-26}$</td>
</tr>
<tr>
<td>suspect + victim</td>
<td>$5.3 \cdot 10^{-09}$</td>
</tr>
<tr>
<td>suspect + unknown man</td>
<td>$1.0 \cdot 10^{-11}$</td>
</tr>
<tr>
<td>suspect + unknown woman</td>
<td>$1.0 \cdot 10^{-11}$</td>
</tr>
<tr>
<td>victim + unknown man</td>
<td>$2.2 \cdot 10^{-17}$</td>
</tr>
<tr>
<td>victim + unknown woman</td>
<td>$2.2 \cdot 10^{-17}$</td>
</tr>
<tr>
<td>unknown man + unknown man</td>
<td>$9.2 \cdot 10^{-19}$</td>
</tr>
<tr>
<td>unknown man + unknown woman</td>
<td>$9.2 \cdot 10^{-19}$</td>
</tr>
</tbody>
</table>

\textbf{Table 6.5:} Likelihoods for the different donor configurations computed using the \textit{forensim} package, example #1

These likelihoods can be entered into the conditional probability table of the \textit{DNA interpretation software} node. An example is given in Table 6.6 (Note that the unknown man/unknown woman categories are pooled). Most Bayesian network software packages insist that the probabilities sum to 1 in each column. Hence, a ‘dummy’ state (compl.) is needed. The asterisk in Table 6.6 represents the complement of the likelihood, i.e. for the first column $1 - 3.9 \cdot 10^{-52}$.

It is much more likely that one would observe the partial DNA profile from Table 6.1 when the suspect and the victim were the donors than any of the other

\textsuperscript{6}doi:10.1016/j.fsigen.2016.08.004
6.3. Mock case example

<table>
<thead>
<tr>
<th>none</th>
<th>susp</th>
<th>vict</th>
<th>unkn</th>
<th>susp, vict</th>
<th>unkn, vict</th>
<th>unkn, unkn</th>
</tr>
</thead>
<tbody>
<tr>
<td>forensim</td>
<td>3.9E−52 2.0E−19 1.2E−30 4.1E−26 5.3E−09 1.0E−11 2.2E−17 9.2E−19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>compl.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 6.6: Conditional probability table containing the likelihoods for the DNA interpretation software node

categories. Only the combination suspect + unknown cannot be disregarded as likely alternative. The likelihood ratio between these two scenarios can be computed as,

\[
LR_1 = \frac{\Pr(\text{DNA profile}|\text{suspect, victim})}{\Pr(\text{DNA profile}|\text{suspect, unknown})} = \frac{5.3 \cdot 10^{-09}}{1.0 \cdot 10^{-11}} = 530
\]

whereas the alternative victim + unknown has a substantial larger likelihood ratio

\[
LR_2 = \frac{\Pr(\text{DNA profile}|\text{suspect, victim})}{\Pr(\text{DNA profile}|\text{victim, unknown})} = \frac{5.3 \cdot 10^{-09}}{2.2 \cdot 10^{-17}} = 2.4 \cdot 10^8
\]

In the sensitivity analysis section (6.3.4), situations with information on fewer loci will be examined. When performing calculations with the network, the DNA interpretation software node should always be fixed in the state forensim. This way, the different donor configurations get the relative likelihoods that follow from the mixture calculations.

6.3.3 Results

The Configuration contributor #1 and Configuration contributor #2 nodes contain all the possible configurations of person and cell type\(^7\) (see Table 6.16 and 6.17 in the appendix for the states). In Oosterman et al. \cite{35}, the following set of hypotheses is used,

- \(H_1\) The suspect contributed semen.
- \(H_2\) The suspect did not contribute semen but contributed another cellular material.
- \(H_3\) The suspect did not contribute to the trace.

For illustrative purposes, a uniform prior distribution is assumed over the different hypotheses, i.e. \(\Pr(H_1) = \Pr(H_2) = \Pr(H_3) = 1/3\). \(H_2\) and \(H_3\) contain several different states. We assume that the prior probability over these different states is also uniform, see Table 6.16. Furthermore, we assume a uniform prior over all the states in the configuration contributor #2 node, see Table 6.17. Ideally, this prior distribution is based on case related background information.

\(^7\)Under the assumption that each donor contributed one cell type, and limited to the cell types that are considered.
6. Cell type determination and association with the DNA donor

By inserting all the evidence $E$ (a partial mixed DNA profile (Table 6.1), a negative Hemastix test for blood, a positive RSID saliva test, a positive RSID semen and positive PSA test for seminal fluid) posterior probabilities are obtained. By defining different sets of hypotheses, the obtained posterior probabilities can be compared in a meaningful way.

The posterior probabilities for these hypotheses are

$$
\begin{align*}
Pr(H_1|E) &= 0.991 \\
Pr(H_2|E) &= 0.010 \\
Pr(H_3|E) &= 0.000
\end{align*}
$$

The posterior probability that the stain is a mixture of DNA material from the suspect and the victim is very high due to the very strong evidence of the partial DNA profile. The positive tests associated with semen cells suggest the presence of semen that cannot be donated by the female victim. The negative test associated with blood results in a low posterior probability that blood is present. The positive test associated with saliva increases the belief that saliva is present. However, the relatively high conditional probabilities assigned to a positive RSID saliva test result for feces and breast milk (see Table 6.14) result in high posterior probabilities for these cell types. The posterior probability for the node Cell type #1 indicates that it is very likely that donor #1 contributed semen cells. The posterior probability distributions of the cell type # 1 and cell type # 2 nodes can be found in Figure 6.3.

Another example, where the positive cell type tests cannot be associated with a gender is discussed in the discussion, Section 6.4.

6.3.4 Sensitivity analysis

The number of observed loci in the DNA profile

The DNA profiles used in this example (see Table 6.1) are very discriminating (see Table 6.5). By altering the number of loci on which we have information in the crime stain DNA profile, their influence on the posterior probabilities for the hypotheses of interest can be examined. In Table 6.7 the posterior probabilities for the three hypotheses of interest given the evidence as function of the number of observed loci in the crime stain DNA profile are presented.

In order to interpret these results, remember that a uniform prior distribution was assumed over the set of hypotheses. Hence, the prior probabilities for the different hypotheses are,

$$
\begin{align*}
Pr(H_1) &= 0.333 \\
Pr(H_2) &= 0.333 \\
Pr(H_3) &= 0.333
\end{align*}
$$

After inserting the positive (RSID saliva, RSID semen and PSA test) and negative
Figure 6.3: Pareto charts representing the posterior probabilities for the **Cell type #1** (left) and **Cell type #2** (right) nodes

(hemastix test) cell type tests,

\[
\begin{align*}
\Pr(H_1|\text{cell type test results}) &= 0.852 \\
\Pr(H_2|\text{cell type test results}) &= 0.046 \\
\Pr(H_3|\text{cell type test results}) &= 0.102
\end{align*}
\]

As can be seen from Table 6.7, the added value of adding more evidence regarding the donor(s) of the crime stain with respect to the hypotheses of interest decreases. It is remarkable to observe that adding more loci does not only increase the belief
that the suspect is a donor (Pr($H_1$) + Pr($H_2$) increases, Pr($H_3$) decreases) but also allows for the discrimination between what type of material was donated (Pr($H_1$) and Pr($H_2$)). This is (partly) due to the assumption that there is a maximum of two donors for the crime stain and each donor contributed only one cell type. If semen cells were part of the crime stain, then they can only have been left by a male. When the ‘certainty’ increases that the female victim is one of the donors of the crime stain, (possible) present semen cells can only be attributed to a second donor. The probability distribution regarding the sex of the second donor can be obtained from the network (see Figure 6.1). The three rightmost columns of Table 6.7 show the posterior distribution for the Sex #2 node with different number of loci in which information is known.

<table>
<thead>
<tr>
<th>E added locus</th>
<th>Hypotheses</th>
<th>Sex #2 node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr($H_1</td>
<td>E$)</td>
</tr>
<tr>
<td>D3S1358</td>
<td>0.951</td>
<td>0.031</td>
</tr>
<tr>
<td>+ vWA</td>
<td>0.949 ↑</td>
<td>0.048</td>
</tr>
<tr>
<td>+ D16S539</td>
<td>0.981 ↑</td>
<td>0.019</td>
</tr>
<tr>
<td>+ D2S1338</td>
<td>0.944 ↓</td>
<td>0.056</td>
</tr>
<tr>
<td>+ D8S1179</td>
<td>0.982 ↑</td>
<td>0.018</td>
</tr>
<tr>
<td>+ D21S11</td>
<td>0.990 ↑</td>
<td>0.010</td>
</tr>
<tr>
<td>+ D18S51</td>
<td>0.990 ↓</td>
<td>0.010</td>
</tr>
<tr>
<td>+ D19S433</td>
<td>0.991 ↑</td>
<td>0.009</td>
</tr>
<tr>
<td>+ TH01</td>
<td>0.992 ↑</td>
<td>0.009</td>
</tr>
<tr>
<td>+ FGA</td>
<td>0.991 ↓</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table 6.7: Posterior probabilities for $H_1$, $H_2$, $H_3$ and the posterior distribution for the Sex #2 node for different numbers of loci in the DNA evidence.

The three rightmost columns of Table 6.7 shows the influence of the certainty regarding the victim being a donor of the crime stain on the probability that $H_1$ is true. In other words, the more certain that the victim is one of the donors of the crime stain, the more certain that the suspect donated semen cells. Hence, making the (common) assumption that the victim is one of the donors of the crime stain to simplify the mixture calculations can be incriminating regarding the suspect. For example, in this mock case example, $H_1$ represents the prosecutor’s hypothesis where $H_2$ and $H_3$ correspond to situations that favor the suspect.

Specificity and sensitivity of the cell type tests

The conditional probabilities adopted for the classical cell type tests are assigned based on data. The amount of uncertainty of these assigned values influences the uncertainty of the posterior probabilities for the hypotheses of interest, $H_1$, $H_2$ and $H_3$. In Table 6.3, 6.13 and 6.14 the point estimates and 95% credible intervals for the conditional probabilities of interest are presented. Instead of the point estimates, the posterior distributions of obtaining a positive test result can be used to examine the influence of the uncertainty on a parameter of interest. In Figure 6.4 (original data), the cumulative distribution function of Pr($H_1|E$) is plotted based on the posterior distributions for obtaining a positive test result for the PSA and the RSID test.
6.4 Discussion

Pr($H_1|E$) fluctuates between approximately between 0.975 and 0.995, see Figure 6.4 (original data). Obviously, this distribution depends on the assumed distributions for the conditional probabilities of obtaining a positive test result. However, these plots give a better understanding of the sensitivity of the point estimate.

The Bayesian network can assist in identifying parameters that have a substantial influence on the probabilistic conclusions that are returned. For example, due to the very limited number of sweat and none/water samples on which the RSID semen test was tested in the collected data (Table 6.13), combined with the method used to obtain the posterior distribution of a positive test result, the probability estimates can be regarded as unrealistically high in forensic casework. In forensic casework with a RSID semen test, it is common to include a negative control sample. Hence, substantially more data should be available to estimate these probabilities. The influence of adding additional samples to the training database on the resulting probabilities can be examined. For example, if instead of the 7 sweat and 3 none/water samples 100 samples were tested for both categories, all resulting in a negative RSID test. The distributions for Pr($H_1|E$) using the added artificial data is given in Figure 6.4. The point estimates of the posterior probabilities with the added artificial data are:

\[
\begin{align*}
\Pr(H_1|E) &= 0.992 \\
\Pr(H_2|E) &= 0.008 \\
\Pr(H_3|E) &= 0.000
\end{align*}
\]

We see that the additional RSID semen tests result in a more narrow distribution function. Furthermore, the corresponding posterior probability increases (the distribution shifts to the right). Performing additional tests is a useful procedure when the forensic practitioner believes that one of the assigned probabilities does not resemble the ‘true’ value. A Bayesian network implementation can assist in determining the ‘effect’ of performing additional tests on the probability of interest.

6.4 Discussion

6.4.1 An example without positive, gender associated, cell type tests

The evidence in the first case example we discussed consisted of a DNA mixture profile that provided strong evidence that a male suspect and a female victim were the donors. Furthermore, the cell type tests that returned as positive can be associated with male donors. Hence, it was possible to ‘link’ DNA profile results with cell type tests. In a situation where the positive cell type tests cannot be linked to a gender, the resulting posterior probabilities are less discriminative. We consider the following mock case example.

Mock case description #2 A severely injured woman was found unconscious in a park. Witnesses stated that the woman was in a fight with a male individual. Both were seen to be injured. At a certain point the man fled the
scene, leaving the woman on the ground. Crime scene officers followed the route of the perpetrator and found a fresh bloodstain near the exit of the park. Police suspect the bloodstain has been left by the injured perpetrator. The bloodstain was sampled and submitted to the forensic laboratory for biological traces testing and DNA typing. The results of the forensic examination are: a partial, mixed DNA profile matching the victim, a positive Hemastix test for blood, a positive RSID saliva test, a negative RSID semen and a negative PSA test for seminal fluid.

A suspect is apprehended who also matches the partial mixed DNA profile. He states that he has not been involved in the incident. He states that he often hangs around with friends near the entrance of the park and has a habit of spitting on the ground. He therefore suggests that he contributed saliva (or possibly another cell type, not blood) to the sample.

The DNA evidence consists of the profiles from Figure 1 from [18], see Table 6.8\textsuperscript{8}. The likelihoods of observing the DNA profiles from Table 6.8 given the

\textsuperscript{8}Reprinted from [18] with permission from Elsevier
6.4. Discussion

<table>
<thead>
<tr>
<th>Locus</th>
<th>Evidence</th>
<th>Victim</th>
<th>Suspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEL</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D3S1358</td>
<td>15</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>VWA</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>D16S539</td>
<td>11</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>D2S1338</td>
<td>18</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>D8S1179</td>
<td>12</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>D21S11</td>
<td>26</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>D18S51</td>
<td>14</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>D19S433</td>
<td>13</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>TH01</td>
<td>6</td>
<td>9.3</td>
<td>7</td>
</tr>
<tr>
<td>FGA</td>
<td>21</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

Table 6.8: Alleles in evidence, victim and suspect

donor combinations from Table 6.2, as computed by the *forensim* [17] package, assuming,

\[
\begin{align*}
\Pr(\text{contamination}) &= 0.05, \\
\Pr(\text{drop out}) &= 0.10, \\
\theta &= 0,
\end{align*}
\]

are given in Table 6.9. If we assume that the AMEL locus evidence determines that at least one male donor was present, the configuration states without a male donor can be set to 0. The following set of hypotheses is considered,

<table>
<thead>
<tr>
<th>configuration</th>
<th>likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>(1.7 \cdot 10^{-67}) (\rightarrow 0)</td>
</tr>
<tr>
<td>suspect</td>
<td>(1.2 \cdot 10^{-40})</td>
</tr>
<tr>
<td>victim</td>
<td>(1.8 \cdot 10^{-22}) (\rightarrow 0)</td>
</tr>
<tr>
<td>unknown man</td>
<td>(1.2 \cdot 10^{-35})</td>
</tr>
<tr>
<td>unknown woman</td>
<td>(1.2 \cdot 10^{-35}) (\rightarrow 0)</td>
</tr>
<tr>
<td>suspect + victim</td>
<td>(5.1 \cdot 10^{-07})</td>
</tr>
<tr>
<td>suspect + unknown man</td>
<td>(1.8 \cdot 10^{-22})</td>
</tr>
<tr>
<td>suspect + unknown woman</td>
<td>(1.8 \cdot 10^{-22})</td>
</tr>
<tr>
<td>victim + unknown man</td>
<td>(1.6 \cdot 10^{-13})</td>
</tr>
<tr>
<td>victim + unknown woman</td>
<td>(1.6 \cdot 10^{-13}) (\rightarrow 0)</td>
</tr>
<tr>
<td>unknown man + unknown man</td>
<td>(3.1 \cdot 10^{-25})</td>
</tr>
<tr>
<td>unknown man + unknown woman</td>
<td>(3.1 \cdot 10^{-25})</td>
</tr>
<tr>
<td>unknown woman + unknown woman</td>
<td>(3.1 \cdot 10^{-25}) (\rightarrow 0)</td>
</tr>
</tbody>
</table>

Table 6.9: Likelihoods for the different donor configurations computed using the *forensim* package, example #2

\(H_a\) The suspect contributed blood.

\(H_b\) The suspect did not contribute blood but contributed another cellular material.

\(H_c\) The suspect did not contribute to the trace.

For illustrative purposes, a uniform prior distribution is assumed over the different offender/cell type configurations, see Table 6.16 and 6.17. After inserting the
evidence $E$ (cell type results and DNA profiles), the posterior probabilities are,

\[
\begin{align*}
\Pr(H_a|E) & = 0.502 \\
\Pr(H_b|E) & = 0.498 \\
\Pr(H_c|E) & = 0.000
\end{align*}
\]

However, the posterior probabilities that blood is present is,

\[
\Pr(\text{blood present}|E) = 0.836
\]

So, since the positive cell type tests correspond to cell types that cannot be associated with a gender, the resulting posterior probabilities are substantially less discriminative. Other (background) information is needed for associating cell types and donors. For example, in Taylor et al. [53], the quantity of DNA detected for the sample taken and the visual appearance of the stain is taken into account.

6.4.2 Dealing with prior probabilities

We have focused on posterior probabilities in this paper. However, these are based on assumptions about prior probabilities which are outside the domain of expertise of the forensic expert. Therefore, for casework reporting a likelihood ratio is often preferred. This is easy to derive from the network if the hypotheses considered are “simple” alternatives, e.g. blood of suspect X and saliva of the victim versus skin of suspect X and saliva of the victim. However, in practice most of the relevant hypotheses represent a subset of such simple hypotheses. If this is the case, their prior probabilities become part of the likelihood ratio. A suggestion to deal with this is to report a table relating posterior probabilities of the relevant hypotheses as a function of the most relevant prior probabilities. Such a table could clarify that the prior probabilities affect the posterior probability of interest, and illustrate the extent of the effect for the most relevant ones.

6.4.3 Additional information regarding cell type test

In the case example we presented, the results from a literature overview were pooled to assign conditional probabilities to the cell type tests. The background data (Table 6.10, 6.11 and 6.12) show that certain stain related circumstances and/or lab protocols can lead to substantially different results. Hence, one should be careful with pooling data in forensic casework. Conditional probabilities for obtaining a positive cell type test result should ideally be assigned based on data that follow the lab protocol and, if possible, the case specific circumstances. For example, in [36, 10, 34, 5, 9, 6, 54] case specific circumstances are taken into account when examining the sensitivity of the RSID saliva test.

The literature overview presented here can aid the forensic practitioner in identifying potential false positive results. Furthermore, it can assist in substantiating the results obtained from their own lab.
6.5 Conclusion

We presented a Bayesian network that can assist in making the association between donor and cell type. A literature overview is presented that can aid in estimating necessary conditional probabilities. The software package \textit{R} [41] is linked with the Bayesian network which allows for the (direct) use of mixture calculation packages like \textit{forensim} [17]. It has been shown that the common assumption that a known person (for example the victim) is a donor of the mixture potentially influences the belief regarding the cell type that was donated by another donor (for example the suspect). This can have unwanted consequences, for example, one might overestimate the value of the evidence.

By using an underlying distribution for the conditional probabilities for positive cell type test results instead of a point estimate, the sensitivity of the posterior probability of interest can be examined. Such a sensitivity analysis gives a better understanding regarding the uncertainty of the posterior probabilities of interest. Furthermore, the influence of performing additional cell type tests on samples of a certain type on the amount of uncertainty can be examined. In Figure 6.4, an example is presented that shows how the precision of the point estimate is increased after performing additional tests.

If reports regarding cell type tests and DNA results are submitted separately, combining the evidential value becomes the responsibility of a trier of fact. Explanations that provide a likely alternative might be overlooked and underlying dependencies become unclear. By combining the evidence, reports benefit from the expertise of the forensic practitioner and the data present at a forensic laboratory.

Supplementary material

R code to implement the \textit{forensim} package with the Bayesian network and the Bayesian networks used in the case examples are available as supplementary material\(^\text{9}\). This material also includes an overview of the nodes and their conditional probability tables.

Bibliography


\(^{9}\)doi:10.1016/j.fsigen.2016.08.004
6. Cell type determination and association with the DNA donor


6. Cell type determination and association with the DNA donor


6.A Literature overview sensitivity and specificity of cell type tests


6.A Literature overview sensitivity and specificity of cell type tests
<table>
<thead>
<tr>
<th>Cell type</th>
<th>Positives/samples</th>
<th>Ref.</th>
<th>Sample description</th>
<th>Test brand used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>0/2 [43]</td>
<td></td>
<td>Liquid (made with 200 µL on swab); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/6 [37]</td>
<td></td>
<td>Liquid; Extraction in 750 µL TE buffer of 25 mm² fabric; 20 µL aliquot used with</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0/6 [12]</td>
<td></td>
<td>Liquid; Extraction in 300 µL ddH2O; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/4 [23]</td>
<td></td>
<td>Liquid; Extraction in 1000µL buffer of 50 mm²; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/8 [29]</td>
<td></td>
<td>Swab; Extraction in water; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/1 [55]</td>
<td></td>
<td>Test performed conform manufacturer’s protocol</td>
<td>Atlantic; HT; Seratec</td>
</tr>
<tr>
<td>Breast milk</td>
<td>0/4 [43]</td>
<td></td>
<td>Liquid (made with 200 µL on swab); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>1/7 [12]</td>
<td></td>
<td>Liquid; Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td>Saliva</td>
<td>0/4 [23]</td>
<td></td>
<td>Liquid; Extraction in 1000µL buffer of 50 mm²; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/2 [4]</td>
<td></td>
<td>Swab/Stain (one of each); Extraction in 300 µL water; 200 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/3 [37]</td>
<td></td>
<td>Stain; 25 mm² fabric extracted in 750 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td>Abacus</td>
</tr>
<tr>
<td>Feces</td>
<td>0/3 [43]</td>
<td></td>
<td>Liquid (made with 200 µL on swab); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/6 [37]</td>
<td></td>
<td>Liquid; Extraction in 750 µL TE buffer of 25 mm² fabric; 20 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/20 [19]</td>
<td></td>
<td>Liquid; Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/1 [4]</td>
<td></td>
<td>Swab; Extraction in 200 µL water; 200 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td>Urine (m)</td>
<td>175/176 [22]</td>
<td></td>
<td>Liquid (diluted 200x); 200 µL dilution used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>169/176 [22]</td>
<td></td>
<td>Liquid (diluted 200x, exposed at 37°C for 48h); 200 µL dilution used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>9/10 [12]</td>
<td></td>
<td>Stain (diluted 160x); Extraction in 300 µL ddH2O; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td>Semen</td>
<td>10/10 [19]</td>
<td></td>
<td>Stain (2-30 years old); Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>5/5 [23]</td>
<td></td>
<td>Stain; Extraction in 1000µL buffer of 50 mm²; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>3/3 [4]</td>
<td></td>
<td>Stain; Extraction in 300 µL water; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>40/46 [2]</td>
<td></td>
<td>Female cervical vaginal swab 0-24h; Extraction in 500 µL PBS; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>11/26 [2]</td>
<td></td>
<td>Postcoital cervical swab 24-48h; Extraction in 500 µL PBS; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/4 [2]</td>
<td></td>
<td>Postcoital swab &gt;60h; Extraction in 500 µL PBS; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td>Sweat</td>
<td>0/6 [37]</td>
<td></td>
<td>Stain; Extraction in 750 µL TE buffer of 25 mm² fabric; 20 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/20 [19]</td>
<td></td>
<td>Swab; Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/7 [43]</td>
<td></td>
<td>Liquid (made with 200 µL on swab); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Abacus</td>
</tr>
<tr>
<td>Urine (f)</td>
<td>3/3 [55]</td>
<td></td>
<td>Liquid; Test performed conform manufacturer’s protocol</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>10/10 [37]</td>
<td></td>
<td>Liquid; Mixed directly with running buffer (quantity unknown)</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/10 [37]</td>
<td></td>
<td>Liquid; Extraction in 750 µL TE buffer of 25 mm² fabric; 20 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/3 [55]</td>
<td></td>
<td>Liquid (diluted 10x); Test performed conform manufacturer’s protocol</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>2/2 [49]</td>
<td></td>
<td>Liquid (undiluted); Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>7/10 [43]</td>
<td></td>
<td>Liquid (made with 200 µL urine that was 100x diluted on swab, wide range of section ages); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>7/10 [19]</td>
<td></td>
<td>Liquid (diluted 1000x); Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>15/15 [12]</td>
<td></td>
<td>Liquid (undiluted on filter paper, subjects &gt; 21 y/o); Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>15/15 [45]</td>
<td></td>
<td>Liquid (diluted 200x on filter paper, subjects &gt; 21 y/o subjects); Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/8 [23]</td>
<td></td>
<td>Swab; Extraction in 1000µL buffer of 50 mm²; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>2/4 [29]</td>
<td></td>
<td>Swab; Extraction in water; 200 µL aliquot used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/1 [4]</td>
<td></td>
<td>Swab; Extraction in 200 µL water; 200 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td>Urine (l)</td>
<td>0/10 [37]</td>
<td></td>
<td>Liquid; Mixed directly with running buffer (quantity unknown)</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/10 [19]</td>
<td></td>
<td>Liquid (diluted 1000x); Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td>Urine (p)</td>
<td>1/15 [12]</td>
<td></td>
<td>Liquid; Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/20 [12]</td>
<td></td>
<td>Liquid; Extraction in 300 µL ddH2O; Centrifugation; 200 µL supernatant used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/7 [45]</td>
<td></td>
<td>Stain (on filter paper); Test protocol unknown</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/8 [23]</td>
<td></td>
<td>Stain; Extraction in 1000µL buffer of 50 mm²; 200 µL aliquot used</td>
<td>Seratec</td>
</tr>
<tr>
<td></td>
<td>0/4 [29]</td>
<td></td>
<td>Swab; Extraction in water; 200 µL aliquot used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/1 [4]</td>
<td></td>
<td>Swab; Extraction in 300 µL water; 200 µL aliquot used</td>
<td>Abacus</td>
</tr>
<tr>
<td>Vaginal secretion</td>
<td>0/3 [43]</td>
<td></td>
<td>Liquid (made with 200 µL on swab); Extraction in 750 µL SS; Centrifugation; 200 µL supernatant used</td>
<td>Abacus</td>
</tr>
<tr>
<td></td>
<td>0/8 [29]</td>
<td></td>
<td>Swab; Extraction in water; 200 µL aliquot used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/10 [19]</td>
<td></td>
<td>Swab; Extraction in 650 µL HEPES buffer; Centrifugation; 200 µL supernatant used</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/3 [55]</td>
<td></td>
<td>Test performed conform manufacturer’s protocol</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/3 [57]</td>
<td></td>
<td>Test performed conform manufacturer’s protocol</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/3 [57]</td>
<td></td>
<td>Test performed conform manufacturer’s protocol</td>
<td>Seratec; VL; Abacus</td>
</tr>
<tr>
<td></td>
<td>0/1 [5]</td>
<td></td>
<td>Test performed on sterile water</td>
<td>Abacus</td>
</tr>
<tr>
<td>Noise/</td>
<td>1/51 [22]</td>
<td></td>
<td>Test performed on sterile water</td>
<td>Seratec</td>
</tr>
<tr>
<td>Water</td>
<td>5/51 [22]</td>
<td></td>
<td>Test performed on sterile water after exposure to 37°C for 48h</td>
<td>Seratec</td>
</tr>
</tbody>
</table>

Table 6.10: Literature study into the PSA test
### 6.A. Literature overview sensitivity and specificity of cell type tests

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Positives/Ref. samples</th>
<th>Sample description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>0/5 [46]</td>
<td>Liquid; Prototype RSID strip used; Test protocol unknown</td>
</tr>
<tr>
<td></td>
<td>0/45 [34]</td>
<td>Liquid (made with 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/6 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/3 [50]</td>
<td>Stain (made from 4 µL); Extraction in 300 µL RSID buffer; 1-20 µL aliquot used with running buffer</td>
</tr>
<tr>
<td></td>
<td>0/4 [23]</td>
<td>Swab (made with 50 µL); Extraction in 100 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td></td>
<td>0/8 [29]</td>
<td>Swab; Extraction in 350 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td>Breast milk</td>
<td>0/1 [46]</td>
<td>Liquid; Prototype RSID strip used; Test protocol unknown</td>
</tr>
<tr>
<td></td>
<td>0/4 [23]</td>
<td>Stain; Extraction in 1000 µL TBS+-buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td></td>
<td>0/3 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/2 [4]</td>
<td>Stain; Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/45 [34]</td>
<td>Stain (made with 50 µL); Extraction in 1000 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/1 [11]</td>
<td>Swab (made with 100 µL); Extraction in 100 µL RSID buffer; Centrifugation; 100 µL supernatant used</td>
</tr>
<tr>
<td>Feces</td>
<td>0/6 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/3 [50]</td>
<td>Stain (made from 4 µL); Extraction in 300 µL RSID buffer; 1-20 µL aliquot used with running buffer</td>
</tr>
<tr>
<td></td>
<td>0/4 [29]</td>
<td>Swab; Extraction in 350 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td></td>
<td>0/1 [11]</td>
<td>Swab; Extraction in 100 µL RSID buffer; Centrifugation; 100 µL supernatant used</td>
</tr>
<tr>
<td>Saliva</td>
<td>0/3 [46]</td>
<td>Liquid; Prototype RSID strip; Test protocol unknown used</td>
</tr>
<tr>
<td></td>
<td>0/45 [34]</td>
<td>Stain (made with 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/6 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/3 [50]</td>
<td>Stain (made from 4 µL); Extraction in 300 µL RSID buffer; 1-20 µL aliquot used with running buffer</td>
</tr>
<tr>
<td></td>
<td>0/4 [29]</td>
<td>Swab; Extraction in 350 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td></td>
<td>0/1 [11]</td>
<td>Swab; Extraction in 100 µL RSID buffer; Centrifugation; 100 µL supernatant used</td>
</tr>
<tr>
<td>Semen</td>
<td>84/84 [34]</td>
<td>Stain (made from 50 µL on various fabrics); Extraction in 1000 µL RSID buffer from 5 mm fabric punch; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>84/84 [34]</td>
<td>Swab (taken from 50 µL semen on various fabrics); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>5/5 [32]</td>
<td>Stain (33-56 years old); Test performed conform manufacturer’s protocol</td>
</tr>
<tr>
<td></td>
<td>3/3 [4]</td>
<td>Stain (30 years old); Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80µL running buffer</td>
</tr>
<tr>
<td></td>
<td>24/24 [34]</td>
<td>Postcoital swab day 0/1; Extraction in 200 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>8/9 [34]</td>
<td>Postcoital swab day 2; Extraction in 200 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/42 [34]</td>
<td>Postcoital swab day 3-10, 14, 17; Extraction in 200 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>5/5 [4]</td>
<td>Postcoital female undergarments up to 48 h; Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80µL running buffer</td>
</tr>
<tr>
<td></td>
<td>1/1 [4]</td>
<td>Postcoital swab 3h; Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/2 [4]</td>
<td>Postcoital swab &gt;24h; Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80µL running buffer</td>
</tr>
<tr>
<td></td>
<td>40/46 [2]</td>
<td>Postcoital swab 0-24h; Extraction in 500 µL PBS; Centrifugation; 20 µL used with 80 µL running buffer (all negatives after 6h)</td>
</tr>
<tr>
<td></td>
<td>16/26 [2]</td>
<td>Postcoital swab 24-48h; Extraction in 500 µL PBS; Centrifugation; 20 µL used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/4 [2]</td>
<td>Postcoital swab &gt;60h; Extraction in 500 µL PBS; Centrifugation; 20 µL used with 80 µL running buffer</td>
</tr>
<tr>
<td>Sweat</td>
<td>0/6 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/1 [11]</td>
<td>Swab; Extraction in 100 µL RSID buffer; Centrifugation; 100 µL supernatant used</td>
</tr>
<tr>
<td>Urine</td>
<td>0/25 [46]</td>
<td>Liquid; Prototype RSID strip used; Test protocol unknown</td>
</tr>
<tr>
<td></td>
<td>1/1 [50]</td>
<td>Liquid; 20 µL neat urine used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/20 [37]</td>
<td>Liquid sample mixed with TBS+-buffer</td>
</tr>
<tr>
<td></td>
<td>0/10 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/2 [50]</td>
<td>Stain (made from 4 µL); Extraction in 300 µL RSID buffer; 1-20 µL aliquot used with running buffer</td>
</tr>
<tr>
<td></td>
<td>0/45 [34]</td>
<td>Stain (made with 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/2 [4]</td>
<td>Stain; Extraction in 100 µL RSID buffer; 20 µL aliquot used with 80µL running buffer</td>
</tr>
<tr>
<td></td>
<td>1/16 [23]</td>
<td>Stain; Extraction in 1000 µL TBS+-buffer; 100 µL aliquot used; False pos. from female using Femodene contraceptive</td>
</tr>
<tr>
<td></td>
<td>1/2 [11]</td>
<td>Swab (made with 100 µL male urine); Extraction in 100 µL RSID buffer; 100 µL aliquot used; Neg. if diluted 10x</td>
</tr>
<tr>
<td></td>
<td>1/2 [11]</td>
<td>Swab (made with 100 µL female urine); Extraction in 100 µL RSID buffer; 100 µL aliquot used; Neg. if diluted 10x</td>
</tr>
<tr>
<td></td>
<td>0/8 [29]</td>
<td>Swab; Extraction in 350 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td>Vaginal secretion</td>
<td>0/18 [34]</td>
<td>Stain (made with 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/4 [23]</td>
<td>Stain; Extraction in 1000 µL TBS+-buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td></td>
<td>0/3 [37]</td>
<td>Stain; Extraction in 200 µL TE buffer from 25 mm² fabric; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td></td>
<td>0/3 [50]</td>
<td>Stain (made with 4 µL); Extraction in 300 µL RSID buffer; 1-20 µL aliquot with running buffer</td>
</tr>
<tr>
<td></td>
<td>1/2 [11]</td>
<td>Swab (made with 100 µL); Extraction in 100 µL RSID buffer; 100 µL aliquot used; Neg. if diluted 10x</td>
</tr>
<tr>
<td></td>
<td>0/8 [29]</td>
<td>Swab; Extraction in 350 µL RSID buffer; 100 µL aliquot used</td>
</tr>
<tr>
<td>None/water</td>
<td>0/3 [50]</td>
<td>Water sample</td>
</tr>
</tbody>
</table>

**Table 6.11:** Literature study into the RSID semen test
<table>
<thead>
<tr>
<th>Cell type</th>
<th>Positives/Ref.</th>
<th>Sample description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blood</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0/6 [36]</td>
<td>Liquid; Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/1 [10]</td>
<td>Swab (made with 50 µL); Extraction in 300 µL RSID buffer; 12 µL aliquot used with 108 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/1 [34]</td>
<td>Swab (made with 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>1/7 [33]</td>
<td>Liquid; Various amounts added to diverse buffers; False positive with 20 µL undiluted blood in 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/5 [38]</td>
<td>Liquid; 50 µL “assayed according to manufacturer’s instruction”.</td>
<td></td>
</tr>
<tr>
<td><strong>Breast milk</strong></td>
<td>1/1 [38]</td>
<td>Liquid; 50 µL “assayed according to manufacturer’s instruction”.</td>
</tr>
<tr>
<td>2/4 [34]</td>
<td>Swab (made with 50 µL); Extraction in 1000 µL RSID buffer, 20-1 µL aliquot used with 80-99 µL running buffer (total 100 µL)</td>
<td></td>
</tr>
<tr>
<td>1/3 [36]</td>
<td>Swab; Extraction in 200 µL TE buffer, 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td><strong>Feces</strong></td>
<td>6/18 [34]</td>
<td>Stain; Extraction in 1000 µL RSID buffer; 5, 20 and 100 µL aliquots used with running buffer; All 100 µL aliquots positive.</td>
</tr>
<tr>
<td>3/3 [36]</td>
<td>Swab (from stain); Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>3/4 [10]</td>
<td>Swab (from stain); Extraction in 300 µL RSID buffer; 12 µL aliquot used with 108 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>3/3 [33]</td>
<td>Suspension: 0.2 - 20 µL feces suspension with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>1/2 [27]</td>
<td>Suspension (made from 0.3/30 µL feces in 500 µL RSID buffer); 100 µL aliquot used; 30 µL sample positive</td>
<td></td>
</tr>
<tr>
<td><strong>Saliva</strong></td>
<td>17/17 [27]</td>
<td>Liquid (made from 0.1 µL in 500 µL RSID buffer); 100 µL aliquot used</td>
</tr>
<tr>
<td>17/17 [27]</td>
<td>Stain; Extraction of cutout or sample in 500 µL RSID buffer; 100 µL aliquot used</td>
<td></td>
</tr>
<tr>
<td>15/17 [27]</td>
<td>Liquid (made from 0.1 µL aged saliva (7 days at 25°C) on various textiles and other backgrounds); Extraction of cutout or sample in 500 µL RSID buffer; 100 µL aliquot used; Concrete and wood samples negative</td>
<td></td>
</tr>
<tr>
<td><strong>Semen</strong></td>
<td>7/7 [33]</td>
<td>Liquid (made from varying amounts of saliva in various different extraction buffers)</td>
</tr>
<tr>
<td>13/14 [44]</td>
<td>Swab (made from 0.1 µL saliva on various substrates); Extraction of cutout or swab in 500 µL RSID buffer; 100 µL aliquot used; Sample from cactus negative</td>
<td></td>
</tr>
<tr>
<td>9/14 [44]</td>
<td>Stain (made from 0.1 µL saliva on various substrates, exposed to room temperature for 7 days); Extraction in 500 µL RSID buffer; 100 µL aliquot used; Wood, rust, sand and concrete samples negative</td>
<td></td>
</tr>
<tr>
<td><strong>Sweat</strong></td>
<td>0/7 [36]</td>
<td>Liquid; Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td>1/4 [10]</td>
<td>Swab (made from 50 µL); Extraction in 300 µL RSID buffer; 12 µL aliquot used with 108 µL running buffer</td>
<td></td>
</tr>
<tr>
<td><strong>Urine</strong></td>
<td>6/7 [36]</td>
<td>Liquid; Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
</tr>
<tr>
<td>1/4 [10]</td>
<td>Swab (made from 50 µL); Extraction in 300 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/9 [10]</td>
<td>Swab (made from 50 µL); Extraction in 300 µL RSID buffer; 12 µL aliquot used with 108 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>7/12 [38]</td>
<td>Liquid (50 µL) “assayed according to manufacturer’s instruction”.</td>
<td></td>
</tr>
<tr>
<td>1/3 [33]</td>
<td>Liquid (Varying amounts added directly to running buffer); False positive with 20 µL undiluted urine with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/1 [34]</td>
<td>Swab (made from 50 µL); Extraction in 1000 µL RSID buffer; 25 µL aliquot used with 75 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>1/4 [27]</td>
<td>Liquid (made from 0.3 - 30 µL urine in 500 µL RSID buffer); 100 µL aliquot used; One male sample of 30 µL positive</td>
<td></td>
</tr>
<tr>
<td>0/7 [36]</td>
<td>Swab; Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td><strong>Vaginal secretion</strong></td>
<td>1/3 [33]</td>
<td>Liquid (Varying amounts added directly to running buffer); False positive with 20 µL undiluted fluid with 80 µL running buffer</td>
</tr>
<tr>
<td>0/8 [36]</td>
<td>Swab; Extraction in 200 µL TE buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/1 [38]</td>
<td>Liquid (50 µL) “assayed according to manufacturer’s instruction”</td>
<td></td>
</tr>
<tr>
<td>0/20 [34]</td>
<td>Swab (no semen); Extraction in 300 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td><strong>None/water</strong></td>
<td>1/19 [27]</td>
<td>Liquid (Various textiles and other backgrounds added to 500 µL RSID buffer); 100 µL aliquot used; asphalt sample positive</td>
</tr>
<tr>
<td>0/4 [33]</td>
<td>Liquid (water on cotton); Extraction in 200 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>0/5 [33]</td>
<td>Liquid (Swab from coffee, tea and cola); Extraction in 300 µL RSID buffer; 20 µL aliquot used with 80 µL running buffer</td>
<td></td>
</tr>
<tr>
<td>1/14 [44]</td>
<td>None (various substrates); Extraction of cutout/swab in 500 µL RSID buffer; 100 µL aliquot used; Sample on cactus was positive</td>
<td></td>
</tr>
<tr>
<td>Cell type</td>
<td>positive/total samples</td>
<td>Pr(RSID semen test pos</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Blood</td>
<td>0/72</td>
<td>0.014</td>
</tr>
<tr>
<td>Breast milk</td>
<td>0/56</td>
<td>0.017</td>
</tr>
<tr>
<td>Feces</td>
<td>0/18</td>
<td>0.050</td>
</tr>
<tr>
<td>Saliva</td>
<td>0/70</td>
<td>0.014</td>
</tr>
<tr>
<td>Semen</td>
<td>179/179</td>
<td>0.994</td>
</tr>
<tr>
<td>Sweat</td>
<td>0/7</td>
<td>0.112</td>
</tr>
<tr>
<td>Urine</td>
<td>4/131</td>
<td>0.038</td>
</tr>
<tr>
<td>Vaginal secretion</td>
<td>1/37</td>
<td>0.051</td>
</tr>
<tr>
<td>None/water</td>
<td>0/3</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Table 6.13: Overview of the estimated conditional probabilities for the RSID semen test and their 95% credible intervals.

| Cell type      | positive/total samples | Pr(RSID saliva test pos|cell type) | 95% CI          |
|---------------|------------------------|------------------------|-------------|
| Blood         | 1/20                   | 0.091                 | (0.012, 0.238) |
| Breast milk   | 4/8                    | 0.500                 | (0.212, 0.788) |
| Feces         | 16/30                  | 0.531                 | (0.360, 0.698) |
| Saliva        | 78/86                  | 0.900                 | (0.827, 0.952) |
| Semen         | 2/16                   | 0.167                 | (0.038, 0.365) |
| Sweat         | 1/21                   | 0.087                 | (0.011, 0.228) |
| Urine         | 16/47                  | 0.347                 | (0.222, 0.484) |
| Vaginal secretion | 1/35              | 0.054                 | (0.007, 0.145) |
| None/water    | 2/42                   | 0.068                 | (0.015, 0.158) |

Table 6.14: Overview of the estimated conditional probabilities for the RSID saliva test and their 95% credible intervals.

| Cell type      | positive/total samples | Pr(Hemastix test pos|cell type) | 95% CI          |
|---------------|------------------------|---------------------|-------------|
| Blood         | 115/122                | 0.935               | (0.886, 0.971) |
| Breast milk   | *                      | 0.050               | *           |
| Feces         | *                      | 0.050               | *           |
| Saliva        | 0/12                   | 0.071               | (0.002, 0.247) |
| Semen         | 0/11                   | 0.077               | (0.002, 0.264) |
| Sweat         | *                      | 0.050               | *           |
| Urine         | *                      | 0.050               | *           |
| Vaginal secretion | *                     | 0.050               | *           |
| None/Water    | 11/38                  | 0.300               | (0.170, 0.449) |

Table 6.15: Overview of the estimated conditional probabilities for the Hemastix test and their 95% credible intervals.
6.B. States of the configuration nodes

<table>
<thead>
<tr>
<th>#</th>
<th>configuration</th>
<th>prior probability, case example #1</th>
<th>prior probability, case example #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>suspect, semen</td>
<td>1/3</td>
<td>1/20</td>
</tr>
<tr>
<td>2</td>
<td>suspect, blood</td>
<td>1/15</td>
<td>1/20</td>
</tr>
<tr>
<td>3</td>
<td>suspect, feces</td>
<td>1/15</td>
<td>1/20</td>
</tr>
<tr>
<td>4</td>
<td>suspect, saliva</td>
<td>1/15</td>
<td>1/20</td>
</tr>
<tr>
<td>5</td>
<td>suspect, sweat</td>
<td>1/15</td>
<td>1/20</td>
</tr>
<tr>
<td>6</td>
<td>suspect, urine (male)</td>
<td>1/15</td>
<td>1/20</td>
</tr>
<tr>
<td>7</td>
<td>none</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>8</td>
<td>unknown man, blood</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>9</td>
<td>unknown man, feces</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>10</td>
<td>unknown man, saliva</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>11</td>
<td>unknown man, semen</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>12</td>
<td>unknown man, sweat</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>13</td>
<td>unknown man, urine (male)</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>14</td>
<td>unknown woman, blood</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>15</td>
<td>unknown woman, breast milk</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>16</td>
<td>unknown woman, feces</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>17</td>
<td>unknown woman, saliva</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>18</td>
<td>unknown woman, sweat</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>19</td>
<td>unknown woman, urine (female)</td>
<td>1/42</td>
<td>1/20</td>
</tr>
<tr>
<td>20</td>
<td>unknown woman, vaginal secretion</td>
<td>1/42</td>
<td>1/20</td>
</tr>
</tbody>
</table>

Table 6.16: States of the configuration contributor #1 node for case example #1 and #2

<table>
<thead>
<tr>
<th>#</th>
<th>configuration</th>
<th>prior probability (case example #1 and #2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>none</td>
<td>1/21</td>
</tr>
<tr>
<td>2</td>
<td>victim, blood</td>
<td>1/21</td>
</tr>
<tr>
<td>3</td>
<td>victim, breast milk feces</td>
<td>1/21</td>
</tr>
<tr>
<td>4</td>
<td>victim, feces</td>
<td>1/21</td>
</tr>
<tr>
<td>5</td>
<td>victim, saliva</td>
<td>1/21</td>
</tr>
<tr>
<td>6</td>
<td>victim, sweat</td>
<td>1/21</td>
</tr>
<tr>
<td>7</td>
<td>victim, urine (female)</td>
<td>1/21</td>
</tr>
<tr>
<td>8</td>
<td>victim, vaginal secretion</td>
<td>1/21</td>
</tr>
<tr>
<td>9</td>
<td>unknown man, blood</td>
<td>1/21</td>
</tr>
<tr>
<td>10</td>
<td>unknown man, feces</td>
<td>1/21</td>
</tr>
<tr>
<td>11</td>
<td>unknown man, saliva</td>
<td>1/21</td>
</tr>
<tr>
<td>12</td>
<td>unknown man, semen</td>
<td>1/21</td>
</tr>
<tr>
<td>13</td>
<td>unknown man, sweat</td>
<td>1/21</td>
</tr>
<tr>
<td>14</td>
<td>unknown man, urine (male)</td>
<td>1/21</td>
</tr>
<tr>
<td>15</td>
<td>unknown woman, blood</td>
<td>1/21</td>
</tr>
<tr>
<td>16</td>
<td>unknown woman, breast milk</td>
<td>1/21</td>
</tr>
<tr>
<td>17</td>
<td>unknown woman, feces</td>
<td>1/21</td>
</tr>
<tr>
<td>18</td>
<td>unknown woman, saliva</td>
<td>1/21</td>
</tr>
<tr>
<td>19</td>
<td>unknown woman, sweat</td>
<td>1/21</td>
</tr>
<tr>
<td>20</td>
<td>unknown woman, urine (female)</td>
<td>1/21</td>
</tr>
<tr>
<td>21</td>
<td>unknown woman, vaginal secretion</td>
<td>1/21</td>
</tr>
</tbody>
</table>

Table 6.17: States of the configuration contributor #2 node
Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

Abstract

In deze bijdrage wordt het gebruik van schakelbewijs in strafzaken vanuit twee verschillende en elkaar soms tegensprekende gezichtspunten beschreven. Naast een juridische beschouwing zal bezien worden hoe schakelbewijs vanuit een kanstheoretische invalshoek benaderd en geduid kan worden. Met deze bijdrage hopen wij inzichtelijk te maken dat verschillende disciplines fundamenteel anders tegen dergelijk bewijs kunnen aankijken. Vanuit dit inzicht hopen wij dat zowel strafjuristen als wiskundigen van elkaars benaderingen leren en interpretatiefouten worden voorkomen. Tevens willen wij een discussie op gang brengen over de verschilpunten tussen de juridische en de kanstheoretische visies op schakelbewijs.

7.1 Inleiding

De opbouw van het artikel is als volgt. In paragraaf 2 wordt de juridische context voor schakelbewijs uiteengezet, in paragraaf 3 de kanstheoretische. In paragraaf 4 worden de verschillende schakelbewijsconstructies in de juridische praktijk besproken. Vervolgens wordt de manier waarop schakelbewijs binnen een kanstheoretisch wiskundig model gebruikt kan worden behandeld. In paragraaf 5 wordt besproken aan welke eisen de modus operandi moet voldoen om te kunnen schakelen.

---

1 Wij gebruiken de term ‘kanstheoretische invalshoek’ i.p.v. vergelijkbare andere termen als ‘statistische-’ of ‘stochastische invalshoek’.
2 Ons doel is derhalve praktische handreikingen te bieden voor een specifiek soort bewijs (schakelbewijs); zie voor een algemene vergelijking van theoretische benaderingen van het bewijs in strafzaken Dubelaar 2014, p. 38-48 [8] en Prakken 2014 [14].
7. Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

In paragraaf 6 worden verschillende aspecten van de bewijsmotivering belicht. Achtereenvolgens worden de vragen beantwoord of van elk feit vastgesteld moet worden of het om een delict gaat (6.1), of schakelbewijs verankerd moet worden (6.2) en of een schakelbewijsconstructie lineair dient te verlopen (6.3). Ten slotte inventariseren wij in paragraaf 7 de punten waarop en de redenen waarom juristen en kanstheoretici verschillend tegen schakelbewijs aankijken, waarmee wij hopen een aanzet te geven tot verdere discussie.

7.2 Juridische context; dubbele bevestiging als beginsel van strafrechtelijk bewijs

Om te kunnen oordelen over het gebruik van schakelbewijs in strafzaken moet aller eerst duidelijk zijn hoe juristen naar het bewijs in strafzaken kijken, en waarom zij dat doen. In zoverre aan het gebruik van schakelbewijs door juristen overwegingen ten grondslag liggen die voortvloeien uit de wijze waarop het strafproces is ingericht, c.q. het bewijsrecht is vormgegeven, kunnen zij immers weinig leren van kanstheoretici. Waar kanstheoretici en juristen het over verschillende zaken hebben, kunnen geen misverstanden rijzen. Maar dan moeten zij zich wel realiseren dat zij het over verschillende zaken hebben. ³ Uitgangspunt van strafrechtelijk bewijsrecht is dat het oordeel omtrent de schuld van de verdachte aan een hem ten laste gelegd feit in beginsel - uitzonderingen als art. 344 lid 2 Sv daargelaten - niet mag berusten op informatie uit een enkele bron. Positief geformuleerd dient de tenlastelegging - direct of indirect - door meer bronnen bevestigd te worden. Van directe bevestiging is sprake als het rechterlijke oordeel dat de verdachte schuldig is aan het ten laste gelegde feit steunt op bewijsmiddelen uit meer dan één bron. Het beginsel van dubbele bevestiging wordt dan materieel opgevat, doordat het betrokken wordt op de bewezenverklaring.⁴ Van indirecte bevestiging is sprake als de deugdelijkheid van de informatie waaraan de rechter zijn overtuiging ontleent getoetst kon worden aan informatie uit een andere bron. Het beginsel wordt dan formeel opgevat, doordat het ziet op de bewijsmiddelen als zodanig. Deze onderscheiden juridische perspectieven worden hieronder nader toegelicht.

7.2.1 Het formele perspectief: hoe komt de rechterlijke overtuiging tot stand?

Verdachte zijn drie verkrachtingen van prostituees ten laste gelegd, die hij in de periode augustus-september 1998 te Amsterdam, Arnhem en Groningen begaan zou hebben. Volgens de aangeefsters beloofde hij met drugs voor seks te betalen, maar kwam hij zijn belofte niet na. Verdachte erkent met elk van de vrouwen seks gehad te hebben maar ontkent verkrachting. Herhaalde pogingen de aangeefsters ter terechtzitting in hoger beroep te doen verschijnen lopen op niets uit; van de verslaafde vrouwen is geen woon- of verblijfplaats bekend.

³ Vgl. Nijboer 2009, p. 53 e.v. [12].
⁴ Zo opgevat is het beginsel van dubbele bevestiging de ratio van zgn. bewijsminimumregels; zie Nijboer 2011, p. 76 e.v. [13].
7.2. Juridische context; dubbele bevestiging als beginsel van strafrechtelijk bewijs

Het Hof ziet in de omstandigheid dat verdachte ten aanzien van de aangeefsters zijn ondervragingsrecht niet heeft kunnen uitoefenen geen beletsel voor veroordeling, die het - behalve op de verklaringen van verdachte en de aangeefsters - baseert op een deskundigenrapport en processen-verbaal waaruit blijkt dat elk van de ten laste gelegde verkrachtingen dezelfde bizarre details vertoonde. Deze bewijsconstructie doorstaat de toets aan het Europees Verdrag tot bescherming van de rechten van de mens (EVRM): naar het oordeel van het Europees Hof voor de Rechten van de Mens (EHRM) kan niet worden gezegd dat verdachtes veroordeling ‘only or to a decisive extent’ op de verklaringen van de aangeefsters gebaseerd was. Daartoe overweegt het dat uit de bewijsmiddelen - in het bijzonder verdachtes verklaringen - blijkt dat de zaak drie vrijwel identieke incidenten, met een vergelijkbare modus operandi betrof die zich in een betrekkelijk korte tijdspanne, in drie verschillende plaatsen hadden voorgedaan.5

In de praktijk van het EVRM overheerst het formele perspectief.6 Art. 6 lid 1 EVRM kent verdachten in strafzaken het recht op een eerlijk proces toe, en volgens lid 3 is de mogelijkheid getuigen te (doen) ondervragen op de openbare terechtzitting - ten overstaan van de rechter die over zijn schuld moet oordelen - daar een aspect van. Het gaat dus om de fairness van de procedure, niet om de toelaatbaarheid c.q. waardering van bewijs als zodanig.7 Weliswaar strekt dit vereiste zich uit tot de beantwoording van de bewijsvraag, maar toch ziet het op de procedure als geheel.8

Met de vaststelling dat een getuige niet ter terechtzitting ondervraagd is, is derhalve niet gegeven dat van een fair trial geen sprake was.9 Die conclusie kan niet getrokken worden als de verdediging onvoldoende initiatief genomen heeft om de verschijning van de getuige te bewerkstelligen,10 niet-verschijning van de getuige gecompenseerd is door middel van counterbalancing measures11 of - en daar gaat onze aandacht naar uit - de desbetreffende verklaring voldoende bevestiging vindt in andere bewijsmiddelen.12

Er zijn echter grenzen aan de compensatie van de beperking van het ondervragingsrecht door middel van steunbewijs: met een fair trial is onverenigbaar dat...

5Casusbeschrijving ontleend aan EHRM 5 april 2005, 39209/02 (Scheper/The Netherlands)
6Wij spreken bewest van ‘overheersen’, om niet de suggestie te wekken dat het materiële perspectief geen enkele rol speelt. Weliswaar zegt het EHRM zich niet in het bewijsrecht van staten-partijen te willen begeven, dat doet er niet aan af dat zijn rechtspraak zich - gegeven de ratio van tegensprekelijkheid als aspect van een fair trial: de deugdelijkheid van het schuldoordeel te bevorderen - ook doet voelen bij motivering van de bewezenverklaring.
7EHRM 19 december 1990, 11444/85 (Delta/France), para. 35.
9Alink & van Zeben 2007, p. 8 [2]; Dubelaar 2014, p. 122 e.v. [8].
10EHRM 10 november 2005, 54789/00 (Bocos Cuesta/The Netherlands), para. 65-66
12Alink & Van Zeben 2007, p. 34 [2] en Dubelaar 2014, p. 127. [8] geven o.i. geen zuivere voorstelling van zaken waar zij stellen dat de noodzaak om compenserende maatregelen te treffen afhankelijk is van aard en omvang van ander bewijsmateriaal; de beschikbaarheid van steunbewijs komt eerst aan de orde als de gelegenheid tot ondervragen, c.q. compensatie d.m.v. counterbalancing measures ontbrak
een veroordeling *solely, mainly of to a decisive degree* berust op de verklaring van een getuige ten aanzien van wie de verdachte zijn ondervragingsrecht niet heeft kunnen uitoefenen. Want hoewel steunbewijs de rechter in staat stelt te beoordelen of de verklaring van een getuige geloof verdient, die verklaring is niet tot stand gekomen in een procedure op tegenspraak. Het gebruik van zo’n verklaring is *unfair* als het de uitkomst van de zaak beslissend zou beïnvloeden; of dat het geval is, hangt ervanaf of steunbewijs voorhanden is, c.q. hoe sterk het steunbewijs is. Daarom verlangt het EHRM bevestiging van de verklaring dat verdachte bij het ten laste gelegde strafbare feit betrokken was.

Het EHRM lijkt bovendien te verlangen dat verdachtes betrokkenheid bij het ten laste gelegde feit bevestigd wordt; onvoldoende is dat de verklaring van een niet ondervraagde getuige als zodanig bevestiging vindt, zelfs als het steunbewijs op betwiste onderdelen van die verklaring betrekking heeft. Aldus verstaan brengt de eis van steunbewijs mee dat de bewezenverklaring van elk ten laste gelegd feit op meer bewijsmiddelen moet berusten - t.w. de verklaring van een niet ondervraagde getuige die wat verdachtes betrokkenheid bij het feit betreft door ander bewijs bevestigd wordt. Het EHRM heeft schakelconstructies - waarin de gezocht betrokkenheid wordt afgeleid uit verdachtes betrokkenheid bij andere feiten - echter niet uitgesloten. In de hiervoor beschreven zaak-*Scheper* oordeelde het Hof dat verdachtes veroordeling ter zake van drie verkrachtingen niet ‘only or to a decisive extent’ op de verklaringen van de - niet ondervraagde - aangeefsters berustte, waartoe het overwoog dat:

‘A number of leads, with which these witnesses had provided the police, had been followed up and had resulted in supporting evidence. Bearing in mind that [de strafzaak tegen verdachte] concerned three nearly identical incidents with a similar *modus operandi* by the perpetrator and which occurred within a relatively short time span in three different towns, and having regard to all the material used in evidence against the applicant, including his own statements made before the police and the trial courts, the Court holds that the applicant’s conviction cannot be said to have been based only or to a decisive extent on the statements given by Ms A., Ms B. and Ms C. to the police.’

De omstandigheden dat de aangeefsters aanwijzingen gegeven hadden waarmee aanvullend bewijs vergaard was, kan niet de doorslag hebben gegeven; dat bewijs zag zelfs niet op de door verdachte betwiste onderdelen van hun verklaringen.
Van beslissende betekenis was dat de ten laste gelegde verkrachtingen gedurende een korte periode, maar in verschillende plaatsen en met dezelfde bizarre details hadden plaatsgevonden; de verklaring van ieder van de aangeefsters kon aan de verklaringen van de anderen worden getoetst. Daaraan behoeft het ontbreken van neutraal steunbewijs, c.q. de omstandigheid dat toetsing van de betwiste verklaringen een cirkelredenering impliceert niet in de weg te staan. Slechts in het geval dat de rechter tot ‘grossly unfair or arbitrary conclusions’ komt, is het recht op een eerlijk proces geschonden.

7.2.2 Het materiële perspectief: op welke gronden berust het rechterlijke oordeel?

Verdachte staat terecht ter zake van oplichting van vijf vrouwen over een periode van twee jaar. De modus operandi van de dader was bij alle feiten dezelfde. Hij vond zijn slachtoffers via een advertentie op tv-zender The Box en legde vervolgens contact via sms-berichten en telefoongesprekken, om ten slotte de desbetreffende vrouw in levenden lijve te ontmoeten. Hij deed zich voor als een rijke man en gebruikte daarbij wisselende - valse - namen; soms vertelde hij dat hij dure sportauto’s bezat. Hij nam zijn slachtoffer vaak mee naar het casino en liet haar dan geld pinnen waarmee hij vervolgens kon gokken. Verscheidene vrouwen heeft hij enorme bedragen beloofd in ruil voor een avondje uit of seks. Verdachte heeft een gokverslaving.

Ten aanzien van drie vrouwen heeft de verdachte bekend deze te hebben opgelicht. Het bewijs bestaat hier steeds uit de aangifte van het slachtoffer en de bekennende verklaring van verdachte. Ten aanzien van de overige twee vrouwen heeft verdachte geen verklaring afgelegd. De rechtbank gebruikt voor deze twee feiten een schakelbewijsconstructie. Zonder een dergelijke constructie zou vrijspraak hebben gevolgd: er lagen immers alleen de aangiften van de twee vrouwen.\(^{18}\)

De rechterlijke overtuiging dient ingevolge art. 338 Sv gebaseerd te zijn op de inhoud van wettige bewijsmiddelen. De rechter zal, naast de rechtmatige verkrijging, steeds de betrouwbaarheid van de verschillende bewijsmiddelen dienen te beoordelen. Vervolgens zal hij de bewijswaarde van de bewijsmiddelen op enigerlei wijze dienen te combineren. In het onderstaande is het ons te doen op dit combineren van bewijs(middelen).\(^{19}\) De vraag hoe verschillende bewijsmiddelen op een epistemologisch verantwoorde wijze gecombineerd kunnen worden, wordt zelden expliciet gesteld: de wijze waarop de rechter zich overtuigd acht in de kern een subjectief, strikt persoonlijk proces van oordeelsvorming. Weliswaar wordt een te grote subjectieve overtuiging tegengegaan door de plicht om de bewezenverklaring van een degelijke motivering te voorzien, dit laat toch onverlet dat het ‘sprongetje’ van het voorhanden bewijsmateriaal naar de bewezenverklaring zich notoir lastig onder woorden laat brengen.\(^{20}\)


\(^{19}\)De situatie waarin de rechter zich op basis van één bewijsmiddel - t.w. het proces-verbaal van een opsporingsambtenaar: art. 344 lid 2 Sv - overtuigd mag achten van de schuld van de verdachte blijft hier onbesproken.

Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

Wat betekent het eigenlijk als de rechter bewijs combineert? In elk geval zal, wil een bewezenverklaring kunnen volgen, elk onderdeel van het probandum - de tenlastelegging - door ten minste één bewijsmiddel moeten worden ondersteund. Het op deze wijze combineren van bewijsmiddelen is tamelijk eenvoudig en inzichtelijk: de verschillende bewijsmiddelen bieden steun aan de verschillende onderdelen van het probandum en indien al deze onderdelen bewezen kunnen worden kan de rechter tot een bewezenverklaring komen. So far, so good. Het wordt echter gecompliceerder, en daarmee interessanter, indien de rechter verschillende bewijsmiddelen die zien op hetzelfde onderdeel van het probandum in ogenschouw neemt. Als, bijvoorbeeld, twee getuigen onafhankelijk van elkaar verklaren over de toedracht van een vechtpartij in een uitgaansgebied, zal de rechter zowel de bewijswaarde van elke verklaring afzonderlijk als van de verklaringen gezamenlijk moeten bepalen. Dit bepalen van de bewijswaarde van de afzonderlijke verklaringen en het combineren daarvan wordt gecompliceerder indien aan de juistheid van de waarnemingen van een getuige getwijfeld kan worden, bijvoorbeeld vanwege het gebruik van alcohol of drugs dat een nadelige invloed heeft op het waarnemingsvermogen. In dit verband heeft Nijboer gesteld dat de bewijsregeling in het Wetboek van Strafvordering gekenmerkt wordt door het algemene beginsel van de dubbele bevestiging:

‘Zo bezien behelst de bewijsregeling ex artt. 338-344a WvSv een algemeen beginsel dat door ons wel wordt aangeduid als de eis van de dubbele bevestiging (van de bewezenverklaarde tenlastelegging): er is (…) steeds een pluraliteit aan informanten vereist om tot een bewezenverklaring te komen.’

Kenmerkend voor de bewijsregeling in art. 338-344a Sv is dat de rechter veel vrijheid wordt gelaten om het bewijs naar eigen inzicht te waarderen en te beoordelen of het voorhanden zijnde bewijs voldoende is om tot een bewezenverklaring te kunnen komen. Slechts in een drietal gevallen heeft de wetgever expliciet gesteld dat de rechter niet tot een bewezenverklaring mag komen als hij niet over meer dan een minimale hoeveelheid bewijs beschikt: de bewijsminimumregels vervat in art. 341 lid 4, art. 342 lid 2 en art. 344a Sv. De enkele verklaring van de verdachte, de enkele verklaring van een getuige, anonieme getuigenverklaringen en de verklaring van een kroongetuige zijn als zodanig onvoldoende bewijs om tot een bewezenverklaring te komen. Dergelijke regels zijn een beperking van de vrije bewijsleer, waarin de rechter maximale vrijheid wordt gelaten om het bewijs te waarderen.

Hoewel het Nederlandse bewijsrecht gewoonlijk negatiefwettelijk genoemd wordt, laat het zich in zijn toepassing beter omschrijven als een voorwaardelijk vrij bewijsstelsel. Immers, vrijwel elk bewijs (d.w.z. materiaal met enige bewijswaarde) kan gecategoriseerd worden onder een van de in art. 339 Sv genoemde bewijsmiddelen en de rechter wordt in beginsel vrijgelaten in zijn waardering van het beschikbare bewijsmateriaal. Slechts in die gevallen waarin een bewijsminimumregel van toepassing is, wordt de rechter beknot in zijn waarderingsvrijheid.

Nijboer 2011, p. 77 (cursivering in origineel). [13].
als aan de voorwaarden van de bewijsminimumregels is voldaan, stelt de wet geen beperkingen aan de bewijswaardering. Praktisch beperkt het Nederlandse bewijsrecht de rechter evenmin in zijn oordeelsvorming als een vrij bewijsstelsel. Dit laat overigens onverlet dat via de voorschriften met betrekking tot de bewijsmotivering wel degelijk controle op de bewijswaardering mogelijk is, zij het indirect.

In de recente jurisprudentie over de toepassing van de unus-testis-regel blijkt dat de Hoge Raad dit voorschrift niet uitsluitend als een kwantitatieve waarborg beschouwt voor een deugdelijke bewezenverklaring. Met de in deze arresten geformuleerde maatstaf van ‘voldoende steun’ wordt de bewijsminimumregel uit art. 342 lid 2 Sv ook inhoudelijk ingevoerd. Deze maatstaf is o.i. ook van belang bij de beoordeling van het inhoudelijke verband dat in een schakelbewijsconstructie tussen de schakels dient te bestaan.

Gezien het feit dat de Hoge Raad als cassatierechter de bewezenverklaring slechts marginaal - over de band van de bewijsmotivering - kan toetsen, heeft de jurisprudentie op dit gebied een sterk casuïstisch karakter. Borgers heeft er echter op gewezen dat een omgekeerd evenredig verband lijkt te bestaan tussen de betrouwbaarheid van een getuigenverklaring en de vereiste ondersteuning daarvan. Omdat schakelbewijsconstructies met de nodige behoedzaamheid dienen te worden gehanteerd (er is immers per definitie sprake van bewijsnood), dient de rechter deugdelijk te motiveren hoe de bewezenverklaring van het ene feit redengevend kan zijn voor het bewijs van het andere feit. Met andere woorden: op welke wijze bieden de schakels voldoende steun aan het bewijs voor een ander feit?

7.3 Kanstheoretische context; modellen en Bayesiaanse netwerken

In veel of misschien wel in de meeste zaken speelt onzekerheid over bepaalde feiten en omstandigheden een belangrijke rol. Als we een situatie willen beschrijven met een wiskundig model, dan ligt het dus voor de hand om dat te doen met een model waarin onzekerheden prominent figureren: een kanstheoretisch model. Dit wiskundige model dient aan de ene kant wiskundig hanteerbaar te zijn, en aan de andere kant de werkelijkheid adequaat te beschrijven. Het is evident dat deze randvoorwaarden elkaar in de weg kunnen zitten. Men is doorgaans vooral geïnteresseerd in de vraag of een verdachte de dader is, en het is verleidelijk om te spreken over de kans hierop, meestal gegeven een aantal bewijsmiddelen en een groot aantal andere onzekere factoren. Het begrip ‘kans’ is echter notoir gecompliceerd en vaag, en er zijn vele manieren om het te interpreteren. Sommige onderzoekers stellen dat menubèrbaupt niet kan spreken over de kans dat een verdachte een bepaald misdrijf heeft begaan, omdat de verdachte het simpelweg wel of niet gedaan heeft. Van deze interpretatie van het kansbegrip heeft het recht inderdaad weinig te verwachten. Er zijn voor het recht echter zinvollere interpretaties, waarin een kans wordt opgevat als een kwantitatieve uitdrukking van iemands overtuwing dat iemand bijvoorbeeld het misdrijf heeft begaan waarvan hij verdacht wordt. Kansen

zijn dan persoonsafhankelijk en daarom wordt deze interpretatie ook wel subjectieve kansrekening genoemd. Dit is inderdaad een geheel andere zienswijze en het verschil tussen de zienswijzen kan wellicht met een voorbeeld worden verduidelijkt. Als persoon A een dobbelsteen onder een omgekeerde beker gooit en stiekem kijkt naar de uitkomst, dan is het aantal ogen voor deze persoon bekend. Persoon B, die deze informatie niet heeft, kan waarschijnlijk niet anders stellen dan dat de kans op elk mogelijks aantal ogen 1/6 bedraagt, ook al staat de uitkomst in feite al vast. Een kansinschatting is dus mede gebaseerd op de voor de betreffende persoon beschikbare informatie en is daarmee niet langer een absolute grootte. Onder bepaalde aannames (waarop wij hier niet ingaan) gaat rekenen met behulp van subjectieve kansinterpretatie op dezelfde manier als rekenen met klassieke absolute kansen. Dit betekent dat wij met behulp van elementaire kansrekening inzicht kunnen krijgen in de validiteit van bepaalde redeneringen in situaties waarin sprake is van onzekerheid.

Vanuit een kanstheoretisch perspectief is schakelen niets meer dan het gezamenlijk modelleren van diverse delicten met inachtneming van de samenhang daar-tussen. In een model dat diverse delicten omvat kan het dan zijn dat informatie over het ene delict de kansinschatting betreffende de dader van een ander delict beïnvloedt. Er zijn verschillende manieren om hier inhoud aan te geven, en we beschrijven eerst het Bayesiaanse (netwerk)model dat ons in staat stelt om onderlinge samenhang op een wiskundig correcte wijze te beschrijven. Om kansinschattingen te maken gebruikt men doorgaans een kansmodel dat bekend staat als het ‘Bayesiaanse model’ of het ‘Likelihood Ratio model’. De ingrediënten van dit model bestaan uit twee of meer elkaar uitsluitende hypothesen, een of meer bewijsmiddelen en eventueel achtergrondinformatie. In het Engels worden deze ingrediënten aangeduid met respectievelijk hypotheses, evidence, en background information, vaak afgekort als $H$, $E$ respectievelijk $I$. Voor een roofoverval zouden de hypothesen bijvoorbeeld kunnen zijn:

$$H_1:$$ de verdachte is de dader
$$H_2:$$ de verdachte is niet de dader

Het bewijsmiddel $E$ zou kunnen bestaan uit de verklaring van een getuige. De achtergrondinformatie $I$ zou onder andere kunnen bestaan uit informatie over de plaats waar de getuige zich bevond tijdens de overval en de hoeveelheid licht ter plekke. Het Bayesiaanse kansmodel vertelt hoe de kansen op de juistheid van de hypothesen $H_1$ en $H_2$ (d.w.z. de persoonlijke overtuiging daaromtrent) veranderen als wij de verklaring van de ooggetuige toevoegen of in aanmerking nemen. De achtergrondinformatie wordt hierbij ook betrokken. Deze verandering wordt inzichtelijk gemaakt door de verhouding van de kansen op de hypothesen te beschouwen. Deze verhouding is:

$$\frac{\text{kans op hypothese } 1}{\text{kans op hypothese } 2}$$

Als de kans dat de verdachte de dader is bijvoorbeeld 3/4 is, dan is de kans dat hij de dader niet is 1/4. De kans dat de verdachte de dader is, is dan drie keer
7.3. Kanstheoretische context; modellen en Bayesiaanse netwerken

groter dan de kans dat hij dat niet is: de kansverhouding is 3. Hierbij maken we een onderscheid tussen de kansverhouding vóór en nádat wij het bewijsmiddel \( E \) beschouwen: de a-priorikansverhouding en de a-posteriorikansverhouding. De zogenaamde regel van Bayes, een basisregel uit de kansrekening, geeft de relatie daartussen als volgt weer:

\[
\text{a-priorikansverhouding} \cdot \text{Likelihood ratio} = \text{a-posteriorikansverhouding}
\]

De \textit{Likelihood Ratio} (afgekort als LR) die in deze formule staat is zelf ook een verhouding van twee kansen, namelijk:

\[
\begin{align*}
\text{kans op } E \text{ als hypothese 1 waar is} &= \text{kans op } E \text{ als hypothese 2 waar is}
\end{align*}
\]

De forensische wetenschap levert een bijdrage aan deze kansinschatting via deelhypothesen als ‘het DNA-spoor is afkomstig van de verdachte’ versus ‘het DNA-spoor is afkomstig van een onbekende niet-verwante persoon’, of ‘de kogel is verschoten met dit vuurwapen’ versus ‘de kogel is verschoten met een ander vuurwapen met dezelfde kenmerken’.

Als beide kansen op het observeren van dit bewijsmateriaal (\( E \)) even groot zijn, is de LR gelijk aan 1. Aan de regel van Bayes zien wij dat de kansverhouding van de hypothesen in dat geval niet verandert als we \( E \) toevoegen. \( E \) is in dit geval neutraal bewijs. Naarmate \( E \) beter past bij hypothese 1 dan bij hypothese 2 wordt de LR groter, en dus ook de kansverhouding van de hypothesen als wij \( E \) in aanmerking nemen. De LR meet dus in feite de mate waarin het bewijs hypothese 1 ondersteunt ten opzichte van hypothese 2. De LR wordt daarom ook wel gezien als een maat voor de bewijskracht in het licht van de gegeven hypothesen, waarbij er meer steun is voor hypothese 1 dan voor hypothese 2 als de LR groter is dan 1 en meer steun voor hypothese 2 dan voor hypothese 1 als de LR kleiner is dan 1.

De a-priorikansverhouding is een inschatting van de juistheid van de gebruikte hypothesen voordat het bewijsmiddel in aanmerking genomen wordt. Zelfs bij een hoge LR kan hypothese 1 nog steeds onwaarschijnlijk zijn, t.w. als de a-priorikans op deze hypothese erg laag was. In ons voorbeeld laat de regel van Bayes zien dat de verhouding van de kansen op hypothese 1 en hypothese 2 verandert als de getuigenverklaring in aanmerking genomen wordt, door deze te vermenigvuldigen met de desbetreffende LR:

\[
\begin{align*}
\text{kans dat de ooggetuige de verklaring aflegt als de verdachte de dader is} &= \text{kans dat de ooggetuige de verklaring aflegt als de verdachte niet de dader is}
\end{align*}
\]

Als de ooggetuige de dader goed heeft kunnen zien en een opvallend zeldzaam kenmerk beschrijft dat de verdachte ook blijkbaar te hebben, bijvoorbeeld heel erg scheel kijken, dan is de kans groot dat hij dit verklaart als de verdachte de dader is, en klein als de verdachte niet de dader is. De LR is in dat geval (veel) groter dan 1. Hoe groot precies vertelt het Bayesiaanse model ons meestal niet, maar het model vertelt ons wel welke kansen een rol spelen en hoe wij alle kansen op een correcte manier moeten combineren. Soms kunnen kansen worden gekwantificeerd, bijvoorbeeld bij een DNA-match, maar zoals gezegd gaat het meestal om een persoonlijke inschatting op basis van kennis en ervaring.
Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

Een kenmerk van schakelbewijs is dat verschillende ten laste gelegde feiten worden beschouwd, hetgeen de waarschijnlijkheidsredenering compliceert. Wiskundigen werken in dergelijke situaties vaak met een zogenaamd Bayesiaans netwerkmodel. Een Bayesiaans netwerk is een grafische weergave van hypotheses en bewijsmiddelen waarin de relaties daartussen zijn aangegeven. Een dergelijk netwerk maakt de onderliggende kansrekening transparanter. Wij zullen hier geen uitvoerige uiteenzetting omtrent Bayesiaanse netwerken in het algemeen geven maar volstaan met enkele voorbeelden, waaronder een netwerk dat te gebruiken is bij schakelbewijs. Een Bayesiaans netwerk van het hierboven beschreven Bayesiaanse model in zijn meest basale vorm is gegeven in figuur 7.1.

De bolletjes in het netwerk worden *knopen* genoemd. Zij representeren hypotheses waarvan wij willen weten hoe groot de kans erop is en bewijsmiddelen die de kansen op deze hypotheses beïnvloeden. Elke knoop kan in verschillende toestanden verkeren. Voor de hypothesesknoop ‘is de verdachte de dader?’ zijn dit de toestanden ‘ja’ en ‘nee’. De omstandigheid dat een bewijsmiddel van invloed is op de kans op een hypothese wordt weergegeven met een pijl tussen de knopen. Achter elke knoop schuilt een kansverdeling. De kansverdeling die achter de knoop ‘getuigenverklaring’ schuilt bevat de kansen dat de ooggetuige verklaart dat de dader bepaalde kenmerken heeft (bijvoorbeeld erg scheel kijkt) *gegeven* of de verdachte al of niet de dader is. Dit zijn precies de kansen die in de LR staan. Knopen waar geen pijl naartoe leidt krijgen een a-priorikansverdeling. Op basis van de beschikbare achtergrondinformatie wordt daaromtrent een beredeneerde aanname gedaan.

De regel van Bayes vertelt ons nu wat de kans is op de verschillende hypotheses gegeven de inhoud van een bepaalde ooggetuigenverklaring. Het Bayesiaanse netwerk kan deze berekening uitvoeren. Samengevat bestaat het Bayesiaanse netwerkmodel uit (1) keuze en definitie van de verschillende knopen, (2) de afhanke-

---

23Recent bijvoorbeeld [16].
lijkheidsstructuur tussen de knopen, (3) kansmodellen die de kansverdelingen achter de pijlen bepalen en (4) achtergrondinformatie die de (a-priori)kansverdelingen bepalen.

In dit eenvoudige voorbeeld is de onderliggende formule voor de kansrekening eenvoudig genoeg om de berekening met de hand te doen. In complexere gevallen, bijvoorbeeld als sprake is van schakelbewijs, maakt het gebruik van Bayesiaanse netwerken het echter mogelijk om het overzicht te behouden van de factoren die beschouwd zijn en de samenhang daartussen en de kansrekening zonder fouten uit te voeren.

In de voorbeeldzaak uit figuur 7.1 is sprake van slechts één hypethese-knoop. Een voorbeeld van een Bayesiaans netwerk voor een schakelbewijsconstructie van twee ten laste gelegde feiten is gegeven in figuur 7.2 (hierin is aangenomen dat elk feit door één dader begaan is en de onderscheiden feiten door dezelfde dader begaan kunnen zijn). In dit netwerk is er een a-priorikans op de hypothese dat de verdachte dader van feit 1 is, en op de hypothese dat beide feiten door dezelfde dader begaan zijn. Een beschrijving van de dader door een getuige van het eerste feit die overeen komt met een beschrijving van de dader door een getuige van het tweede feit vergroot de kans op de hypothese dat de dader van feit 1 ook feit 2 begaan heeft.

Figuur 7.2: Bayesiaans netwerk voor een schakelbewijsvoorbeeld

Opnieuw laten de pijlen zien welke elementen van invloed zijn op de verschil-
7. Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

lende hypotheses. Wanneer wij de bewijsbeelden (de getuigenverklaringen) invoeren berekent het netwerk wat de kans is dat de daders dezelfde zijn. Natuurlijk moet hierbij niet uit het oog worden verloren dat er a-priorikansverhoudingen bepaald moeten worden om de berekening te kunnen maken. De afhankelijkheidsrelaties tussen bewijsmiddelen en hypotheses in dit netwerk zijn gecompliceerder dan in figuur 7.1, maar wij merken wel op dat de onderliggende kansrekening eigenlijk dezelfde is: de regel van Bayes, herhaaldelijk toegepast. Er bestaat uitstekende software om in een Bayesiaans netwerk berekeningen te kunnen doen. Het netwerk is een visueel hulpmiddel om de situatie goed te beschrijven, en het maakt ook direct duidelijk welke kansen veranderen als wij bepaalde bewijsmiddelen veranderen. De bijbehorende software maakt dat het netwerk ook kan functioneren als rekenmachine voor de kansberekeningen.

Het werken met Bayesiaanse netwerken is de facto een manier om te laten zien welke elementen worden beschouwd en hoe zij precies met elkaar samenhangen. Wiskundig gezien is het niet nodig om het op deze manier aan te pakken. Hiervoor merkten wij op dat schakelen vanuit een wiskundige context niets anders is dan het gezamenlijk modelleren van verschillende delicten, en vervolgens binnen dat model kansen op gebeurtenissen uit rekenen. Dit kan met een netwerk, maar ook op de klassieke wiskundige manier met formules. Een voordeel van netwerken is dat zij in ingewikkelde situaties de beschouwde hypotheses en bewijsmiddelen en hun afhankelijkheidsstructuur overzichtelijker weergeven dan formules. Tevens kunnen door middel van een netwerk mogelijke complicaties zichtbaar gemaakt worden.

7.4 Schakelbewijs

In deze paragraaf worden de verschillen tussen het gebruik van schakelbewijs in de juridische praktijk en de mogelijkheden van schakelbewijs volgend uit het Bayesiaanse netwerkm model besproken. In het bijzonder wordt ingegaan op de vragen wat onder schakelbewijs verstaan wordt en aan welke eisen het moet voldoen.

7.4.1 Juridische praktijk

Elders is reeds een overzicht gegeven van de verschillende bewijsconstructies die in de Nederlandse rechtspraak als schakelbewijs worden benoemd. Hieronder zullen zij slechts gememoreerd worden door middel van voorbeelden ontleend aan de jurisprudentie. De eerste variant betreft feiten waartussen wordt geschakeld nadat zij als zodanig bewezen zijn verklaard. Feiten met een gelijksoortige modus operandi worden in deze variant weliswaar geschakeld, maar dit schakelen vindt ten overvloede plaats: de bewezenverklaring van de verschillende feiten volgt reeds uit de desbetreffende bewijsmiddelen. Zo overwoog het Gerechtshof Amsterdam:

‘Het hof stelt voorop dat het gebruik van schakelbewijs in beginsel niet ongeoorloofd is, maar dat dit slechts dan toelaatbaar is bij bewezenverklaring van soortgelijke feiten indien en voor zover het bewijs voor

\[\text{zie [17].}\]
het ene feit redengevend is voor het bewijs van het andere ten laste gelegde feit, waarbij opmerking verdient dat de andere bewezenverklaringen die worden “geschakeld” daarnaast zelfstandig moeten worden gefundeerd.\textsuperscript{25}

Bij een dergelijke bewijsconstructie spreken wij niet van schakelbewijs omdat er geen noodzaak bestaat om de feiten onderling te schakelen: als zij zelfstandig gefundeerd kunnen worden, kan de bewezenverklaring het stellen zonder een schakelbewijsconstructie. De tweede variant betreft de situatie waarin één bewijsmiddel bewijskracht heeft voor meer tenlastegelegde feiten. In een dergelijke situatie spreken wij evenmin van schakelbewijs, maar van het meermaals gebruiken van hetzelfde bewijsmiddel. Een voorbeeld hiervan is te vinden in een arrest van het Gerechtshof Arnhem-Leeuwarden betreffende een zedenzaak. Onder het kopje ‘Schakelbewijs’ overwoog het hof:

‘Met de advocaat-generaal is het hof van oordeel dat in deze zaak ook ruimte is voor een zogenaamde schakelbewijsredenering, die ten aanzien [sic] van alle vier de feiten gebruikt kan worden bij de bewijsconstructie. Het gaat om soortgelijke feiten, waarbij delen van het bewijs in de ene zaak ook redengevend zijn voor het bewijs in (een of meer van) de andere zaken. Uit de bewijsmiddelen blijkt van een specifieke manier van handelen van verdachte, die overeenstemt met de gang van zaken in elk van de afzonderlijke feiten. Die specifieke kenmerken zijn de volgende:

- bij alle vier feiten legt verdachte contact via de sociale media, waarbij hij zich telkens jonger voordoet dan hij in werkelijkheid is;
- met de meisjes wordt afgesproken bij de [supermarkt] aan het [adres], waarna hij ze meeneemt naar zijn kamer, volgens hem om naar een film te kijken of op de computer te spelen;
- alle vier de meisjes geven aan dat verdachte op zijn kamer onbeschermd seks met hen had;
- verdachte heeft bij alle vier de meisjes alleen de benedenkleding uitgetrokken.

Het hof acht deze kenmerken dermate specifiek en eensluidend bij alle vier de feiten, dat deze manier van handelen van verdachte mede redengevend is voor de bewezenverklaring van deze vier feiten.’\textsuperscript{26}

De derde variant betreft de situatie waarin daadwerkelijk geschakeld wordt. De Rechtbank Rotterdam verwoordde dat aldus:

‘Volgens de rechtbank is van “echt” schakelbewijs sprake als voor het bewijs, leidend tot de conclusie dat de verdachte een tweede, hem ten

\textsuperscript{25}Hof Amsterdam 8 september 2010, ECLI:NL:GHAMS:2010:BN6408.
\textsuperscript{26}Hof Arnhem-Leeuwarden 26 april 2013, ECLI:NL:GHARL:2013:6093.
laste gelegd feit heeft begaan, wordt gebruikt de net daarvoor bereikte conclusie dat de verdachte een ander strafbaar feit heeft begaan.  

Wij reserveren de term ‘schakelbewijs’ voor gevallen als deze: de bewezenverklaring van één ten laste gelegd feit wordt redengevend geacht voor het bewijs van een ander. Alleen in dergelijke gevallen kan worden gezegd dat de bewezenverklaring van het tweede feit afhankelijk is van de bewezenverklaring van het eerste (of de bewezenverklaringen over en weer afhankelijk zijn, waarover hieronder meer), en blijkt de meerwaarde van Bayesiaanse netwerken.

7.4.2 Kanstheoretisch

Een schakelbewijsconstructie maakt het mogelijk om de overtuiging dat de verdachte zich schuldig gemaakt heeft aan één ten laste gelegd feit te gebruiken bij beantwoording van de vraag of hij een ander feit begaan heeft. Als voor twee feiten een schakelbewijsconstructie gebruikt wordt spelen de volgende drie vragen een cruciale rol:

1. Is de verdachte de dader van feit 1?
2. Is de verdachte de dader van feit 2?
3. Is de dader van feit 1 dezelfde persoon als de dader van feit 2?

De antwoorden op deze vragen zijn afhankelijk van elkaar; een antwoord op twee van de vragen is vaak voldoende om het derde antwoord te bepalen. Als de verdachte de dader van feit 1 is en de dader van feit 1 is dezelfde persoon als de dader van feit 2, dan is de verdachte ook de dader in feit 2. Ook als onzekerheid bestaat over de antwoorden op deze drie vragen (zoals in de praktijk het geval is) zijn zij afhankelijk van elkaar. Als waarschijnlijk is dat de verdachte de dader van feit 1 is en dat de dader van feit 1 dezelfde persoon is als de dader van feit 2, dan is waarschijnlijk dat de verdachte de dader van feit 2 is.

Er is vanuit het oogpunt van de kanstheorie weinig verschil tussen de drie genoemde varianten van schakelbewijs in de juridische praktijk. In alle varianten wordt de bewijswaarde van bewijsstukken direct (zoals in variant 2) of indirect via een juridisch oordeel over een bepaald feit (varianten 1 en 3) gebruikt voor andere feiten. Het verschil tussen de eerste en de derde variant is gelegen in het antwoord op de vraag of de rechter al tot een veroordeling kan komen zonder gebruik te maken van schakelbewijs. Het gebruik van schakelbewijs in variant 1 is vooral bedoeld om de algehele bewijsconstructie te versterken, niet om onderscheiden feiten te kunnen bewijzen. Daarin is het verschil met variant 3 gelegen. In variant 2, waarin één bewijsmiddel bewijskracht heeft voor meer ten laste gelegde feiten, is de link tussen de (twee) feiten van belang. Deze bewijsmiddelen zijn van belang.

---


28 Als op de eerste twee vragen het antwoord ‘nee’ is, of als op vraag 1 (of 2) en vraag 3 het antwoord ‘nee’ is volgt het antwoord op de laatste vraag niet.
7.5. Verantwoording van het rechterlijke oordeel: de modus operandi als conditio sine qua non

voor de vraag of de verdachte de dader van een bepaald feit is alsmede voor de vraag of de dader van feit 1 ook de dader is van feit 2 is. Een voorbeeld hiervan is het geval waarbij ooggetuigen van verschillende feiten beschrijvingen geven die overeenkomen met elkaar én met het uiterlijk van een verdachte. Bewijsmiddelen die voor het bewijs van verschillende feiten relevant zijn vergroten de kans dat dezelfde persoon aan alle feiten schuldig is.

Wij zien dat voor de kanstheoretische invalshoek slechts van belang is in hoeverre de verschillende kansen van elkaar afhangen. De kanstheoretische invalshoek voor schakelen tussen verschillende feiten kan dan ook gedefinieerd worden als het berekenen van een zogenaamde gezamenlijke kansverdeling van alle relevante factoren. Een gezamenlijke verdeling is niets anders dan exact vastleggen hoe informatie over een bepaalde hypothese (bewijsmiddel) de kans op een andere hypothese (bewijsmiddel) doet veranderen, of, algemener, hoe informatie over een verzameling hypothesen de kansen op andere hypothesen beïnvloedt.

7.4.3 Balans

In de juridische praktijk zijn er drie bewijsconstructies waarbij van schakelbewijs gesproken wordt. Wij reserveren de term ‘schakelbewijs’ echter voor gevallen waarin de bewezenverklaring van één ten laste gelegd feit meewerkt tot het bewijs van een ander. Deze definitie impliceert twee belangrijke voorwaarden:

• om gebruik te maken van een schakelbewijsconstructie is noodzakelijk dat er voor minimaal één feit voldoende bewijs is om, afzonderlijk van de andere feiten, tot een bewezenverklaring te komen (een zogenaamd ‘anker’ of ‘locomotief’ voor de schakelbewijsconstructie);

• schakelen gebeurt serieel. Een nieuwe schakel (feit) in de keten wordt pas behandeld wanneer over het voorgaande feit een oordeel is geveld.

Vanuit het oogpunt van de kansrekening zijn deze voorwaarden niet noodzakelijk, en is het op deze manier zelfs mogelijk om bewijs over of onder te waarderen. Een tweede verschil tussen de juridische praktijk en de kansrekening is dat juridisch gezien het gebruik van schakelbewijs slechts dan toegestaan is indien de van de verschillende feiten op essentiële punten overeenkomt. Vanuit het oogpunt van de kansrekening is dit niet nodig, en is samenhang tussen de bewijsmiddelen of hypothesen voldoende om te (kunnen) schakelen (dat wil dus zeggen, om de gezamenlijke kansverdeling zinvol te kunnen beschouwen). Op deze verschillen wordt in de volgende paragrafen nader ingegaan.

7.5 Verantwoording van het rechterlijke oordeel: de modus operandi als conditio sine qua non

In deze paragraaf wordt besproken wat de rol van de modus operandi voor het schakelen van feiten is. Vanuit de juridische praktijk is dit een eerste vereiste om te kunnen schakelen. Vanuit de kansrekening is dit niet het geval.
7. Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

7.5.1 Juridische praktijk

Uit de bestudeerde jurisprudentie blijkt dat rechters zich in het algemeen terdege bewust zijn van de complexiteit van het gebruik van schakelbewijs. Het gebruik van schakelbewijs wordt gemotiveerd, waarbij men zich er veelal rekenschap van geeft bekend te zijn met de mogelijke gevaren van schakelbewijs. Een mooi voorbeeld van een bewijsmotivering waarbij de rechtbank specifiek heeft stilgestaan bij de vereisten voor het gebruik van schakelbewijs is te vinden in een vonnis van de Rechtbank Rotterdam. De rechtbank overweegt:

‘Schakelbewijs heeft een grotere waarde naarmate (1) er meer feiten zonder gebruik te maken van schakelbewijs kunnen worden bewezen, (2) de methode om de feiten te plegen uitzonderlijker, unieker, specifieker, minder voorkomend is, (3) er meer overeenkomsten zijn in de methode of methodes die is of zijn gebruikt bij de verschillende feiten en (4) de voor het schakelbewijs te gebruiken verklaringen onafhanke lijker van elkaar zijn afgelegd.’

Met name het onder (2) genoemde criterium is o.i. van belang. De rechtbank verwoordt hier in feite dat de *modus operandi* van de verschillende feiten, kort gezegd, overeenkomstig dient te zijn. Daartoe dient hij na te gaan of er voldoende overeenkomsten zijn, c.q. te motiveren waarin deze gelegen zijn.

Die motivering kan niet gevonden worden in de blote gelijkenis tussen verschillende feitencomplexen, want dan dreigt een drogreden: de vraag is niet of de feiten op elkaar lijken, maar of zij van andere onderscheiden kunnen worden. De rechter dient allereerst te bepalen of er een gelijkssoortige *modus operandi* bestaat ten aanzien van de verschillende feiten (de *intra-individuele* variatie). Vervolgens, en dat is o.i. de kern van het gebruik van schakelbewijs, dient hij te bepalen hoe uitzonderlijk een dergelijke *modus operandi* is (de *interindividuele* variatie). Met andere woorden: springt de werkwijze van de verdachte in het oog, of is slechts sprake van een werkwijze die nauw samenhangt met het type delict? Met dit laatste wordt bedoeld dat de werkwijze van de verdachte niet typisch of veelvoorkomend mag zijn: zakkenrollers, bijvoorbeeld, plegen vaak in groepjes te opereren waarbij het slachtoffer door één groepsled wordt afgeleid om zo een ander lid de gelegenheid te geven toe te slaan. Een schakelbewijsconstructie waarbij een dergelijke *modus operandi* wordt vastgesteld is weinig solide: een dergelijke *modus operandi* springt

---

7.5. Verantwoording van het rechterlijke oordeel: de modus operandi als \textit{conditio sine qua non}

immers niet in het oog.\textsuperscript{33}

7.5.2 Kanstheoretisch

Uit kanstheoretisch oogpunt is een overeenkomstige modus operandi geen vereiste om te kunnen schakelen. De bewijswaarden van verschillende bewijsmiddelen voor een feit zijn van belang voor de bewijswaarden betreffende een ander feit als er aanwijzingen zijn dat het om dezelfde dader gaat. Bijvoorbeeld omdat de \textit{modus operandi} overeenkomt, maar ook als bijvoorbeeld het slachtoffer in verschillende zaken dezelfde (soort) persoon is of er overeenkomstige waarnemingen zijn (DNA-profielen, schoenafdrukken, getuigenverklaringen) betreffende verschillende feiten, enz. In al deze gevallen is het gebruik van een schakelbewijsconstructie relevant. Een voorbeeld van een zaak met verschillende \textit{modi operandi} maar met een aanwijzing dat de desbetreffende feiten mogelijk door dezelfde persoon zijn gepleegd is het geval waarin dezelfde persoon eerst met een mes wordt aangevallen en er vervolgens met een vuurwapen een aanslag op hem gepleegd wordt. Behalve de voorwaarde van een overeenkomstige \textit{modus operandi} spreken juristen verder alleen van schakelbewijs bij feiten die in tijd gescheiden zijn. Wanneer verschillende strafbare feiten niet in tijd gescheiden zijn - dus deel uitmaken van hetzelfde feitencomplex - valt er juridisch gezien niets te schakelen. Implicit wordt er dan vaak aangenomen dat de dader van de verschillende strafbare feiten dezelfde persoon is, wat vanuit kanstheoretisch oogpunt op schakelen neerkomt. Een voorbeeld.

\textit{Met beveiligingscamera's is een roofoverval door één dader vastgelegd waarbij de eigenaar van de zaak gedood wordt. Kort na de overval worden bij een verdachte thuis een pistool en spullen die ontvreemd zijn tijdens de roofoverval gevonden. Alhoewel het hier om twee te onderscheiden feiten (roofoverval en doodslag) gaat, is het dankzij de camerabeelden evident dat de persoon die de roofoverval pleegde ook de eigenaar doodde. Derhalve zijn de gestolen spullen dus niet alleen relevant voor de vraag wie de roofoverval pleegde maar ook voor de vraag wie de eigenaar doodde.}

Zelfs zonder het directe bewijs dat de persoon die de eigenaar doodde ook de roofoverval pleegde (de camerabeelden) kan het bij een kort tijdsverloop tussen overval en inbeslagneming waarschijnlijk zijn dat de persoon die de spullen stal ook de eigenaar doodde. Als er bijvoorbeeld alleen camerabeelden van de voordeur zijn waarop te zien is hoe één gemaskerde persoon het huis binnendringt, en geen beelden van de achterdeur, dan bestaat er geen direct bewijs maar is het wel waarschijnlijk dat het om één dader gaat. Als de overval en de doodslag in de context van dezelfde strafzaak behandeld worden is er vanuit kanstheoretisch oogpunt sprake van een schakelbewijsconstructie tussen twee feiten met een verschillende \textit{modus operandi} (de \textit{modus operandi} van overvallen verschilt immers van die van doden). Er moet uiteraard wel rekening gehouden worden met de


\textit{‘De rechtbank is van oordeel dat de enkele omstandigheid dat zowel bij feit 7 als bij de onder 1 tot en met 6 tenlastegelegde feiten de panden zijn betreden via een gat in het dak, niet voldoende is om via een schakelbewijsconstructie tot de bewezenverklaring van al deze feiten te komen. Immers, dit is een \textit{modus operandi}, die in zijn algemeenheid bij meer inbraken wordt toegepast.’}

ondzekerheid betreffende de vraag of de persoon die de roofoverval pleegde ook de eigenaar doodde. Voor de vraag wie de eigenaar doodde zijn de gestolen spullen die bij de verdachte gevonden zijn niettemin relevant. De bewijswaarde hiervan is afhankelijk van de waarschijnlijkheid dat de twee strafbare feiten dezelfde dader hebben. Omdat de feiten waar het om gaat niet in tijd gescheiden zijn en dit dus als één feitencomplex beschouwd wordt, is er in juridische zin geen sprake van een schakelbewijsconstructie. Kanstheoretisch gezien komt de redenering echter wel op hetzelfde neer. Bewijsmiddelen die van belang zijn voor het ene feit worden van belang voor het andere feit omdat er sterke aanwijzingen zijn dat het om dezelfde dader gaat. De gestolen spullen zijn relevant voor de overval (feit 1). De mate waarin zij relevant zijn voor het doden van de eigenaar (feit 2) hangt af van de kracht van de aanwijzingen dat het om dezelfde dader gaat. De onderliggende redenering is wiskundig gezien daarom exact hetzelfde als wanneer feit 1 en 2 wel in tijd gescheiden zouden zijn: een schakelbewijsredenering.

Een derde situatie waarin wiskundig gezien sprake is van schakelbewijs maar juristen dit woord niet zullen gebruiken betreft het geval waarin een schakelbewijsconstructie relevant is om te bewijzen dat de verdachte niet de dader is. Bijvoorbeeld, als er een roofoverval wordt gepleegd in Rotterdam en er op hetzelfde moment een moord plaatsvindt in Amsterdam, dan zijn gestolen spullen uit Rotterdam gevonden bij een verdachte relevant voor de vraag of hij de dader van het Amsterdamse feit is. De schakelbewijsconstructie werkt dus de andere kant op. Het aan treffen van de gestolen spullen verkleint namelijk de kans dat de verdachte schuldig is aan de Amsterdamse moord.

7.5.3 Balans

In de juridische praktijk mag alleen in eigenlijke zin geschakeld worden als de modus operandi van de verschillende feiten op essentiële punten overeenkomt. De interindividuele variatie dient groot en de intra-individuele variatie dient klein te zijn. Slechts in deze gevallen mag de bewezenverklaring van een feit redengevend zijn voor het bewijs van een ander gelijksoortig feit. Verder spreekt men in de praktijk alleen van schakelbewijs als het om feiten gaat die in tijd gescheiden zijn en het gaat om belastend bewijs. Vanuit de kansrekening kan een schakelbewijsconstructie veel breder beschouwd worden als iedere constructie waarin bewijs voor een feit relevant is voor de vraag wie de dader is van andere feiten omdat het de waarschijnlijkheid van de hypothese de dader van feit 1 is de dader van feit 2 bevoedt. Een modus operandi die op essentiële punten overeenkomt vergroot de waarschijnlijkheid van deze hypothese. Maar hetzelfde geldt voor situaties waarin bijvoorbeeld het slachtoffer van twee feiten dezelfde persoon is of bewijsmiddelen die in het onderzoek betreffende verschillende feiten gevonden zijn met elkaar overeenkomen (DNA-profieLEN, vingerafdrukKEN, ooggetuigenbeschrijvingEN, enz.). Ook als de feiten niet in tijd gescheiden zijn of als er bewijs bestaat dat twee feiten niet door dezelfde persoon gepleegd kunnen zijn (d.w.z. bewijsmateriaal dat de waarschijnlijkheid van de hypothese de dader van feit 1 is de dader van feit 2 negatief bevoedt), kan een schakelbewijsconstructie zinvol zijn. Kortom, een schakelconstructie is ook zinvol voor andere vormen van samenhang dan overeen-
komstige en kenmerkende *modi operandi*.

### 7.6 Motivering van de bewezenverklaring

Bij deze vraag gaat het om de juridische voorwaarden waaronder de bewezenverklaring van een strafbaar feit gemotiveerd kan worden door middel van een schakelbewijsconstructie. In het voorafgaande zijn deze voorwaarden al aan de orde geweest. De oorsprong van schakelbewijs is gelegen in het beginsel van dubbele bevestiging: teneinde het rechterlijke oordeel controleerbaar te maken, moet het oordeel omtrent de schuld van de verdachte aan het ten laste gelegde direct of indirect bevestigd worden. Daarom beperken juristen de term ‘schakelbewijs’ doorgaans tot gevallen waarin als uitvloeisel van dit beginsel waarschijnlijk een vrijspraak gevolgd zou zijn.

#### 7.6.1 Moet voor elk feit afzonderlijk worden vastgesteld dat een strafbaar feit begaan is?

**Juridische praktijk**

Het meest in het oog springende aspect van deze vraag betreft de toedracht van de gebeurtenissen waarop de bewijsconstructie betrekking heeft. Die toedracht stond centraal in het arrest van het Hof Den Haag in de zaak-*Lucia de Berk*.[34] Hoewel het hof slechts voor twee van de ten laste gelegde levensdelicten bewijs van verdachtes handelen gevonden had, veroordeelde het verdachte ter zake van tien levensdelicten. De verdediging had bepleit dat schakelbewijs in deze zaak eerst in aanmerking zou komen als er voor elk ‘geschakeld’ feit ten minste een aanwijzing gevonden zou worden dat een strafbaar feit begaan was, daaruit zou blijken welk delict haar kon worden verweten en het schakelbewijs verdachtes handelen zou betreffen.[35] Strikt genomen kan dan nog sprake zijn van ‘echt’ schakelbewijs, maar de door de verdediging geformuleerde eisen zijn wel streng: schakelbewijs voorziet slechts in gevallen waarin een bewijsminimumregel - en niet het ontbreken van de rechterlijke overtuiging - tot vrijspraak zou dwingen.[36] Het zou geen uitkomst bieden voor gevallen als dit, waarin de toedracht van een gebeurtenis c.q. verdachtes betrokkenheid daarbij ongewis is.

Het hof sprak echter niet vrij van de acht feiten waarvoor direct bewijs ontbrak. Daartoe overwoog het hof onder meer dat (a) verdachte de twee bedoelde delicten op een specifieke wijze - door toediening van een levensgevaarlijke dosis van een niet voorgeschreven geneesmiddel, in een ziekenhuis - had begaan, (b) de tenlastelegging betrekking had op soortgelijke incidenten (onverwachte, medisch

---

onverklaarbare incidenten gedurende diensten van verdachte), (c) zich in één ziekenhuis gedurende een periode van een jaar zeven incidenten hadden voorgedaan, (d) zes incidenten op een gewone verpleegafdeling hadden plaatsgevonden en (e) een herkenbaar en gelijkssoortig patroon kon worden vastgesteld als verdachtes handelen en haar verklaringen in aanmerking werden genomen.37

In cassatie overwoog de Hoge Raad dat het hof aldus in zijn oordeel dat verdachte schuldig was aan de acht feiten waarvoor geen direct bewijs voorhanden was, had betrokken dat de feitelijke gang van zaken ten aanzien van deze feiten - context en omstandigheden van de feiten, resp. handelen en verklaringen van verdachte - op wezenlijke punten belangrijke overeenkomsten vertoonde met de gang van zaken ten aanzien van de twee feiten waarvoor wel direct bewijs voorhanden was. Dat mocht het hof volgens de Hoge Raad doen; het hoefde niet aan de strenge eisen van de verdediging te voldoen.38 In het bijzonder is het gebruik van schakelbewijs niet beperkt tot de gevallen waarop de verdediging het oog had, waarin de bewezenverklaring overwegend berust op ander bewijs.39

Dat in de zaak-Lucia de Berk de aandacht vooral uitging naar de toedracht van de ten laste gelegde incidenten is wel verklaarbaar. Als natuurlijke oorzaken uitgesloten konden worden zou het bewijs van verdachtes schuld, c.q. van haar veroorzaken van die incidenten immers goeddeels geleverd zijn. Toch moeten deze kwesties onderscheiden worden. Het Hof Den Haag onderscheidde althans ten dele. Het overwoog dat met betrekking tot acht feiten weliswaar bewijsmiddelen ontbraken waaruit kon worden afgeleid welk handelen van verdachte tot het desbetreffende incident had geleid, maar dat dit voor het bewijs van verdachtes opzet, c.q. haar voorbedachte raad ook niet vereist was. In elk geval kon bewezen verklaard worden dat het incident telkens was veroorzaakt door verrichtingen ten aanzien van de desbetreffende patiënt, wat ondenkbaar is zonder opzet, c.q. voorbedachte raad.40 Dat kon er in cassatie mee door: uit een identieke handelwijze kan voorbedachte raad worden afgeleid.41

Kanstheoretisch

Over de zaak-Lucia de Berk is ook vanuit kanstheoretisch perspectief veel gepubliceerd, vooral over de ondeugdelijkheid van het originele statistische rapport.42 Op dat aspect gaan wij hier niet in, maar wij maken wel enkele algemene observaties. In zaken waarin nog niet vaststaat of wel een strafbaar feit is begaan (zoals in de zaak-Lucia de Berk) is het nog steeds mogelijk om te schakelen. Het antwoord op de vraag of sprake is van een delict en, zo ja, of de verdachte verantwoordelijk is voor dit delict zijn nauw met elkaar verbonden. Als binnen korte tijd, in hetzelfde ziekenhuis, verscheidene baby’s overlijden, rijst de vraag of sprake is van (een) onbekende factor(en) die deze gebeurtenissen verklaren. Neem het volgende hypothetische voorbeeld.

38 HR 14 maart 2006, ECLI:NL:HR:2006:AU5496, r.o. 6.3.2.
41 HR 14 maart 2006, ECLI:NL:HR:2006:AU5496, r.o. 6.5.
42 Zie o.m. [10, 11, 7].
7.6. Motivering van de bewezenverklaring

Binnen een nacht branden in de binnenstad honderd auto’s uit. Uit historische gegevens is bekend dat gemiddeld één auto per nacht door kortsluiting en dergelijke uitbran delt en dat uitschieters maar zelden voorkomen. De observatie dat er op een nacht honderd auto’s uitbranden doet vermoeden dat een onbekende onnatuurlijke factor hiervoor de oorzaak is. Dit kan brandstichting zijn, maar ook bijvoorbeeld een nieuwe populaire auto waarbij kortsluiting vaker voorkomt. ‘Onschuldige’ verklaringen worden echter ook na uitvoerig onderzoek niet gevonden. Een persoon wordt opgepakt. Uit locatie-informatie van zijn mobiele telefoon en spullen gevonden bij hem thuis wordt aannemelijk dat hij verantwoordelijk is voor een aantal branden, maar niet precies welke.

Zonder zekerheid te hebben welke branden de verdachte veroorzaakt heeft is het wellicht wel ‘voldoende’ aannemelijk te maken dat de verdachte een rol heeft in een aanzienlijk deel van de branden. Hoe groter het aantal branden, des te groter de kans dat bij elke specifieke brand sprake is van brandstichting. Wij kunnen dan in de situatie komen dat er voor elke specifieke brand een zeer grote kans is dat de verdachte deze gesticht heeft, terwijl wij tegelijkertijd ook verwachten dat één brand een natuurlijke oorzaak heeft, maar we niet kunnen zeggen welke brand. Dit lijkt een paradox, soortgelijk aan de ‘gatecrasher paradox’, waarin van een populatie mensen duidelijk is dat er een aantal schuldig zijn, maar niet kan worden vastgesteld wie.43

Balans

Om te schakelen is het in de juridische praktijk evenmin als in de kansrekening noodzakelijk om voor elk afzonderlijk feit vast te stellen dat sprake is van een delict. Het observeren van een opmerkelijk aantal overeenkomende incidenten vergroot meestal de waarschijnlijkheid dat er een onnatuurlijke factor aan ten grondslag ligt. De aannemelijkheid dat dezelfde persoon daarvoor verantwoordelijk is hangt af van het aantal feiten maar ook van de beschikbaarheid van alternatieve verklaringen. Als het om feiten gaat waarvoor ook een ‘onschuldige’ verklaring bestaat (hier is de a-priorikans op deze verklaringen relevant) is het, zonder aanvullend bewijs, niet mogelijk om de feiten te onderscheiden.

7.6.2 Moet schakelbewijs verankerd worden?

Met de afwijzing van kwantitatieve beperkingen aan schakelbewijs - t.w. tot gevallen waarin de bewezenverklaring overwegend op ander bewijs berust - is niet gezegd dat geen kwalitatieve beperking in acht genomen moet worden. In het bijzonder is daarmee niet uitgesloten dat schakelbewijs slechts in aanmerking komt in het geval dat voor ten minste één ten laste gelegd feit voldoende bewijs voorhanden is en andere feiten bewezen verklad worden vanwege hun gelijkenis met dat feit. In de zaak-Lucia de Berk kwam die vraag eerst in herziening bij het Hof Arnhem, na verwijzing door de Hoge Raad, aan de orde.

43[1, 6, 9]
7. Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

Juridische praktijk

De Hoge Raad verklaarde de aanvraag tot herziening gegrond op de grond dat de bewijsvoering ter zake van het onder 1 bewezen verklaarde feit - een van de feiten waarvoor naar ’s hofs oordeel voldoende direct bewijs voorhanden was - wezenlijk ondergraven was: een natuurlijke oorzaak van overlijden kon niet langer worden uitgesloten. Bovendien was twijfel gerezen aan de representativiteit van een tot het bewijs gebezigd bloedmonster en aan het tijdstip waarop een fatale dosis digoxine toegediend zou zijn: gedurende de desbetreffende periode werd het slachtoffer onderzocht.44 Naar het oordeel van het Hof Arnhem was het noodzakelijke anker daarmee weggeslagen:45

‘Het hof acht zich ontslagen van diepergaande beschouwingen, nu het bij geen enkel ten laste gelegde feit tot een zelfdragende bewezenverklaring komt. Zonder schakels kan nu eenmaal geen ketting worden gemaakt.’

Kanstheoretisch

Vanuit het oogpunt van de kansrekening is een anker simpelweg niet noodzakelijk. De overtuiging dat de verdachte schuldig is aan een feit dat hem ten laste gelegd wordt hangt af van het bewijs dat deze hypothese ondersteunt. Als meer feiten ten laste gelegd zijn en geschakeld wordt benvloeden bewijsmiddelen betreffende onderscheiden feiten de kans op de hypothese dat de verdachte de dader van alle geschakelde feiten is. Als een feit waarvoor zonder schakelbewijsconstructie vrijspraak zou volgen verankerd wordt, dan vergroot dit de overtuiging dat de verdachte de dader is van het geschakelde feit en vice versa. Vanuit het oogpunt van de kansrekening ontstaat er in de praktijk ook een ongerijmdheid tussen twee theoretisch gezien dezelfde situaties wanneer een ankerfeit noodzakelijk zou zijn. Dit wordt gellustrueerd door het volgende voorbeeld.

**Twee feiten, met overeenkomstige modi operandi worden geschakeld. Er is een verdachte. Achtergrondinformatie over het aantal mogelijke daders maakt dat de a-priorikans dat de verdachte de dader van een specifiek feit is op 1 op 1000 wordt gesteld.**46

Situatie A: er is een ankerfeit  Wat het eerste feit betreft bestaat het bewijs uit een DNA profiel verkregen uit een spoor waarvan wordt aangenomen dat het is achtergelaten door de dader en dat overeenkomt met het DNA-profiel van de verdachte. De random-matchkans van dit DNA profiel is 1 op 100 000. Het bewijs voor het tweede feit is een vingerspoor waarvan wordt aangenomen dat het is achtergelaten door de dader en dat overeenkomt met de vingerafdruk van de verdachte. De random-matchkans van dit bewijs is 1 op 1000. Met het Bayesiaanse netwerk uit figuur 7.3 kan de a-posteriorikans dat de verdachte de dader van een specifiek feit is op 1 op 1000 worden berekend.

---

44 HR 7 oktober 2008, ECLI:NL:HR:2008:BD4153, r.o. 5.1.
45 Hof Arnhem 14 april 2010, ECLI:NL:GHARN:2010:BM0876, r.o. 3.1.3.
46 Het inschatten van deze a-priorikans is een probleem op zich, waar wij hier niet nader op ingaan.
7.6. Motivering van de bewezenverklaring

feit is bepaald worden (hetzelfde resultaat wordt bereikt door gebruik te maken van de regel van Bayes). De a-posteriorikans dat de verdachte de dader van het eerste feit is bedraagt 99,01%. De a-posteriorikans dat de verdachte de dader van het tweede feit is bedraagt ongeveer 50% (50,025%). Als we een ankerfeit definiëren als een feit ten aanzien waarvan de a-posteriorikans dat de verdachte de dader is minimaal 99% is, kunnen deze feiten dus geschakeld worden. De a-priorikans dat de beide misdaden door dezelfde persoon zijn gepleegd wordt op basis van de modus operandi geschat op 90%. Het Bayesiaanse netwerk dat gebruikt kan worden om de a-posteriorikans dat de verdachte de dader van specifieke feiten is te bepalen is gegeven in figuur 4. De a-posteriorikans, na schakelen, dat de verdachte de dader van het eerste feit is bedraagt 99,999%. Wat het tweede feit betreft is deze kans gelijk aan 99,988%. Dit leidt mogelijk tot een veroordeling voor beide delicten.

Situatie B: er is geen ankerfeit  Als de random-matchkansen van de bewijzen anders zijn, bijvoorbeeld 1 op 10 000 voor zowel het DNA-profiel als het vingergespoor, en het model uit figuur 7.3 wordt gebruikt om de a-posteriorikans op dat de verdachte de dader is te berekenen zonder te schakelen, vinden wij een a-posteriorikans van 90,92% voor beide feiten afzonderlijk. Aangezien dit minder is dan 99% kan feit 1 evenmin als feit 2 als anker dienen in een schakelbewijsconstructie. De verdachte wordt daarom vrijgesproken van beide delicten. Na schakelen, volgens het Bayesiaanse netwerk uit figuur 7.4, vinden wij echter a-posteriorikansen van 99,998% dat de verdachte de dader van specifieke feiten is. Deze getallen zijn van eenzelfde orde als in de vorige situatie. Ook al ontbreekt het in deze tweede situatie aan een ankerfeit, de a-posteriorikansen dat de verdachte de dader is als de feiten gezamenlijk beschouwd worden zijn nagenoeg hetzelfde.

![Figuur 7.3: Bayesiaans netwerk voor een enkel feit](image-url)

Het is dus mogelijk dat het belastende bewijs voor twee feiten zonder anker net zo sterk is als het bewijs voor twee vergelijkbare feiten met een anker. Dat maakt het ongerijmd dat er zonder anker altijd een vrijpraak zou moeten volgen, terwijl
Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

Is de dader uit zaak 1 ook de dader in zaak 2?

Is de verdachte de dader in zaak 1?

DNA match

Is de verdachte de dader in zaak 2?

Vingerspoor match

Figuur 7.4: Bayesiaans netwerk voor de geschakelde feiten

47 Het verschil in tussen dit Bayesiaanse netwerk en het Bayesiaanse netwerk uit figuur 7.2 zit hem in de ongelijksoortigheid van de bewijzen betreffende de onderscheiden feiten. Twee signalementen zijn afhankelijk van elkaar wanneer het om dezelfde onbekende dader gaat. Voor een DNA-profiel en een vingerspoor geldt dit niet.
7.6. Motivering van de bewezenverklaring

Balans

In de juridische praktijk lijkt vereist te worden dat ten minste een van de feiten bewezen verklaard wordt zonder te schakelen. Vanuit de kanstheorie is dit niet het geval. Het is mogelijk dat twee feiten zonder anker gezamenlijk beschouwd net zo sterk zijn als twee vergelijkbare feiten met een anker. Als een anker noodzakelijk is zal dit dus niet tot eenzelfde conclusie leiden, hetgeen ongerijmd is. Sterker nog, omdat voor de vraag of de verdachte de dader van een bepaald feit is bewijs voor andere feiten relevant wordt, kunnen de gronden voor bewezenverklaring wegvallen door te schakelen. De analogie van het anker, of nog beter, de locomotief die de gehele trein moet trekken gaat dan ook niet op. Deze locomotief kan misschien op zichzelf wel rijden, maar als er na schakelen een locomotief gekoppeld wordt die de andere kant op rijdt is het mogelijk dat het geheel blijft staan, of zelfs de andere kant op rijdt.

7.6.3 Telkens één schakel: een waarborg tegen overwaardering van bewijs?

Hetgeen in par. 7.6.2 is opgemerkt met betrekking tot de verankering van schakelbewijs impliceert dat schakelen lineair verloopt: er wordt telkens één schakel aangelegd. De bewezenverklaring van een enkel feit wordt gemotiveerd door verwijzing naar de kenmerkende gelijkenissen tussen dat feit en hetgeen bewezen verklaard is. En dan behoeft het geen betoog dat elke bewezenverklaring het gemakkelijker maakt een volgend feit te bewijzen; ook rechters zijn gevoelig voor confirmation bias.48

Juridische praktijk

Niet geheel ten onrechte, want als de gevonden gelijkenissen voldoende onderscheidend vermogen hebben, verhoogt de bewezenverklaring de a-priorikansverhouding ten aanzien van het te bewijzen feit: aan de desbetreffende bewijzen komt groter gewicht toe. En bewezenverklaring daarvan verhoogt op haar beurt de a-priorikansverhouding ten aanzien van het volgende te bewijzen feit.49 Er is dus wel sprake van ‘confirmation’, niet van ‘bias’: de rechter moet zich rekenschap geven van de wijze waarop, c.q. de mate waarin bewezenverklaring van het ene feit zijn waardering van bewijzen voor het andere feit bevoedt.

Kanstheoretisch

Vanuit het oogpunt van de kansrekening is het serieel schakelen problematisch. Immers, de bewijsstukken betreffende afzonderlijke feiten zijn relevant voor alle andere feiten in de keten. Als het schakelen lineair verloopt is er maar één richting waarop bewijswaarde in de bewijsconstructie overgedragen kan worden; van de reeds ‘bewezen’ feiten naar de ‘te bewijzen’ feiten. De beslissing over specifieke

48Goede rechters zijn zich niet alleen van hun ‘bias’ bewust, maar ook van de wijze waarop zij deze kunnen beperken: Hof Arnhem 14 april 2010, ECLI:NL:GHARN:2010:BM0876, r.o. 3.1.3.
49zie [17], p. 133.
feiten wordt dan afhankelijk van de volgorde waarin ze behandeld worden, zie het volgende tegenvoorbeeld.

Een schakelconstructie wordt gebruikt om drie feiten te schakelen. De link tussen de feiten bestaat uit een overeenkomstige, specifieke modus operandi (d.w.z. het is waarschijnlijk dat dezelfde persoon verantwoordelijk is voor alle drie de feiten). Wat twee feiten betreft is er sterk bewijs dat tegen de verdachte pleit. Wat het derde feit betreft is er overtuigend bewijs dat de verdachte niet de dader is. Als de feiten lineair behandeld worden maakt het uit in welke volgorde dit gebeurt. Als de feiten in serie behandeld worden, dan is de a-priorikans dat de verdachte de dader van het tweede feit is afhankelijk van de uitspraak betreffende het eerste feit. Als de verdachte duidelijk niet de dader is van het eerste feit dat behandeld wordt, betekent dit dat de a-priorikans dat de verdachte de dader van het eerste feit bij beneden bijgesteld moet worden, er is immers (sterk) bewijs dat er een andere persoon dan de verdachte verantwoordelijk is voor een feit met een overeenkomstige modus operandi. Als de verdachte duidelijk de dader van het eerste feit is, dan is de a-priorikans veel groter. De kans dat de verdachte ter zake van het tweede feit wordt veroordeeld hangt af van de volgorde waarin de feiten behandeld worden.

Omdat de volgorde waarin de feiten behandeld worden arbitrair is, is dit een onacceptabel gevolg. Dit kan voorkomen worden door feiten binnen een schakelbewijsconstructie als een geheel te beschouwen. Als er overeenkomsten tussen feiten bestaan, benvloeden de bewijzen betreffende alle feiten de vraag of de verdachte de dader van een specifiek feit is. De onderliggende bewijsmiddelen bepalen het antwoord op deze vraag en zijn dus direct relevant voor het beantwoorden hiervan ten aanzien van de andere feiten. Door slechts naar één feit te kijken loopt men het gevaar bewijs over te waarderen doordat afhankelijkheden worden genegeerd, bijvoorbeeld als bij het onderzoek naar onderscheiden feiten hetzelfde onvolledige DNA-profiel wordt gevonden. Ook kan het gebeuren dat de implicaties van de omstandigheden dat bewijzen betreffende verschillende feiten niet overeenstemmen (bijvoorbeeld ooggetuigen met niet overeenkomende signalementen) over het hoofd worden gezien. Met andere woorden: de seriële benadering kan leiden tot zowel over- als onderwaardering van het bewijs. Als geschakelde feiten gezamenlijk behandeld worden en de conclusies parallel getrokken worden is hier geen sprake van. Dit betekent uiteraard niet dat de beslissing wat alle feiten betreft dezelfde moet zijn, het is best mogelijk dat de verdachte aan sommige feiten wel en aan ander feiten niet schuldig bevonden wordt.

Een schrikbeeld voor het gebruik van schakelbewijsconstructies is dat voor elk nieuw feit dat aan de keten toegevoegd wordt er minder bewijs nodig is om tot een bewezenverklaring te komen. Hiertegen kan echter worden ingebracht dat een groot deel van het bewijs in de schakels tussen de feiten zit. Als feiten geschakeld worden, is het bewijs voor specifieke feiten relevant voor de vraag wie de dader van alle feiten is. De mate waarin bewijs betreffende een specifiek feit relevant is voor de andere feiten is afhankelijk van de sterkte van de schakel tussen de feiten (hoe zeker is het dat de feiten door dezelfde persoon gepleegd zijn?). Als vaststaat dat beide feiten door dezelfde persoon gepleegd zijn, is de waarde van bewijs betreffende een specifiek feit net zo relevant voor het beantwoorden van
7.6. Motivering van de bewezenverklaring

de dadervraag betreffende andere feiten. In een hypothetische casus waarin de sterkte van de schakels tussen alle geschakelde feiten even groot is, bijvoorbeeld omdat het bewijs dat de verschillende feiten door dezelfde persoon gepleegd zijn uitsluitend bestaat uit het feit dat in alle gevallen precies dezelfde *modus operandi* gebruikt werd, draagt elk feitenspecifiek bewijs evenveel bij aan beantwoording van de schuldvraag betreffende andere feiten. Dus als een ‘nieuw’ feit aan de keten van feiten wordt toegevoegd, zal de aannemelijkheid van de hypothese dat de verdachte de dader is zeer sterk bevoorvloed zijn door de eerdere gevonden bewijsstukken. Als er voor dit (nieuwe) feit wederom (onafhankelijk) bewijs is dat steun biedt aan de hypothese dat de verdachte de dader is, dan zal dit de a-priorikans op de hypothese dat de verdachte de dader van een nieuw feit is weer doen stijgen. Dit impliceert dat voor elk nieuw feit met dezelfde *modus operandi* steeds minder bewijs nodig is om tot een bewezenverklaring te komen.

**Balans**

In de juridische praktijk verloopt schakelen serieel. Na een uitspraak betreffende het eerste feit wordt het tweede feit onderzocht. Wat dat feit betreft kan de omstandigheid dat het eerste feit bewezen verklaard is gebruikt worden als bewijs. Nadat de bewijsconstructie verankerd is, kunnen opeenvolgende schakels worden aangelegd. Vanuit het oogpunt van de kansrekening is het echter niet noodzakelijk om het schakelen serieel te laten verlopen. Sterker nog, door feiten een voor een te beschouwen is het mogelijk dat de uitspraak over onderscheiden feiten afhankelijk wordt van de volgorde waarin zij behandeld worden. Het kan er bovendien toe leiden dat bewijs wordt over- of ondergewaardeerd.

7.6.4 Conclusie

In dit artikel hebben wij onderzocht hoe juristen en kanstheoretici tegen schakelbewijs in strafzaken aankijken: hoe bepalen hun onderscheiden perspectieven op het combineren van bewijs of - en, zo ja: onder welke voorwaarden - de omstandigheid dat de verdachte schuldig wordt bevonden aan een strafbaar feit kan meewerken tot beantwoording van de vraag of hij aan een ander feit schuldig is. Daarbij bleek in de eerste plaats dat de juridische praktijk verankering van de bewijsconstructie lijkt te verlangen, wat meebrengt dat schakelen vervolgens serieel verloopt. Uit kanstheoretisch oogpunt zijn deze voorwaarden niet noodzakelijk, en is het op deze manier zelfs mogelijk om bewijs over of onder te waarderen.

Voorts verlangen juristen dat de *modus operandi* van de verschillende feiten op essentiële punten overeenkomt. Slechts in deze gevallen wordt de bewezenverklaring van een feit redengevend geacht voor het bewijs een ander gelijksoortig feit. Het gaat dus om belastend bewijs, betreffende feiten die in de tijd gescheiden zijn. Uit kanstheoretisch oogpunt wordt daarentegen geschakeld als bewijs de waarschijnlijkheid van de hypothese de dader van feit 1 is de dader van feit 2 bevoorvloedt. Een *modus operandi* die op essentiële punten overeenkomt vergroot de waarschijnlijkheid van deze hypothese, maar dat geldt voor meer omstandigheden. Ook als de feiten niet in tijd gescheiden zijn en zelfs als er bewijs bestaat dat twee
feiten niet door dezelfde persoon gepleegd kunnen zijn, kan een schakelbewijscon-
structie zinvol zijn.

Wat de motivering van de bewezenverklaring betreft ten slotte is het in de 
juridische praktijk evenmin als in de kanstheorie noodzakelijk om voor elk afzon-
derlijk feit vast te stellen dat sprake is van een delict. Het observeren van een 
opmerkelijk aantal overeenkomende incidenten vergroot de waarschijnlijkheid dat 
er een onnatuurlijke factor aan ten grondslag ligt. Wel lijkt de juridische praktijk 
te verlangen dat ten minste één strafbaar feit bewezen verklaard wordt zonder te 
shakelen. In de kanstheorie daarentegen kan het bewijs voor twee feiten zonder 
anker gezamenlijk beschouwd net zo sterk zijn als het bewijs voor twee vergelijk-
bare feiten met een anker. Dat maakt de voorwaarde van verankering ongerijmd, 
te meer nu de gronden voor bewezenverklaring kunnen wegvallen doordat geschak-
keld wordt: dat maakt andere bewijzen relevant. In de kanstheorie wordt derhalve 
niet serieel geschakeld, omdat dit de uitspraak over onderscheiden feiten afhanke-
lijk maakt van de volgorde waarin zij behandeld worden. Het kan ertoe leiden dat 
bewijs wordt over- of ondergewaardeerd. Wij willen benadrukken dat het ons er 
in dit artikel niet om te doen was uit te maken of juristen dan wel kanstheoretici 
het gelijk aan hun zijde hebben. Gegeven dat wij de validiteit van het juridische, 
resp. het kanstheoretische perspectief, c.q. van de daaraan ten grondslag liggende 
uitgangspunten niet ter discussie gesteld hebben, konden wij dat niet uitmaken. 
Wel hebben wij getracht duidelijk te maken in welke opzichten die uitgangspun-
ten verschillen, en welke daarvan de praktische consequenties (kunnen) zijn. Wij 
hopen dat wij met onze analyse een aanzet geven tot verdere discussie omtrent de 
praktijk van schakelbewijs.

Bibliografie


Modelling crime linkage with Bayesian networks

Abstract

When two or more crimes show specific similarities, such as a very distinct modus operandi, the probability that they were committed by the same offender becomes of interest. This probability depends on the degree of similarity and distinctive-ness. We show how Bayesian networks can be used to model different evidential structures that can occur when linking crimes, and how they assist in understanding the complex underlying dependencies. That is, how evidence that is obtained in one case can be used in another and vice versa. The flip side of this is that the intuitive decision to “unlink” a case in which exculpatory evidence is obtained leads to serious overestimation of the strength of the remaining cases.

8.1 Introduction

Suppose that two similar burglaries occur in a small village within a small time span. In the second one, a suspect is identified. The question whether this person is also responsible for the first crime arises. Clearly this depends on possible incriminating or exculpatory evidence in this first case, but also on the degree of similarity between the two burglaries. Several interesting questions arise in such common situations. For instance, can one “re-use” evidence incriminating the suspect in the second case as evidence in the first case? How does the evidence “transfer” between the two cases? How does the degree of similarity between the two cases affect this transfer? What happens when the evidence in the two cases partially overlaps, or shows dependencies? How can we make inferences for more than two cases?

In practice, it is generally assumed by the police, prosecution and legal fact finders that when there are two or more crimes with specific similarities between them there is an increase in the belief that the same offender(group) is responsible for all the crimes. The probability that there is only one offender(group) depends on the degree of similarity between the crimes. Even for a small number of crimes, the probabilistic reasoning rapidly becomes too difficult. In such situations it is recognized that a Bayesian Network (BN) model can help model the necessary probabilistic dependencies and perform the correct probabilistic inferences to evaluate the strength of the evidence [15]. We can use BNs to examine how evidence found in one case influences the probability of hypotheses about who is the offender in another case.

In this paper, we will show how BNs can help in understanding the complex underlying dependencies in crime linkage. It turns out that these complex dependencies not only help us understand the impact of crime similarities, but also produce results with important practical consequences. For example, if it is discovered that in one of the similar crimes the suspect is not involved, then simply discarding that crime from the investigation could lead to overestimation of the strength of the remaining cases due to the dependency structure of the crime linkage problem. Hence, the common procedure in law enforcement to select from a series of similar crimes only those cases where there is evidence pointing to the suspect and disregard the other cases and evidence can be misleading. Our analysis thus extends the analysis of Evett et al.[6]

The notion of ‘crime linkage’ may be perceived and dealt with differently at different levels in the judicial process. During investigation (i.e., not at trial) considerations are typically very broad and connections among crimes may be made on other criteria than probability. In this paper, we focus on understanding the underlying logic regarding crime linkage. The examples we present serve as “thought experiments”. Such experiments are commonly used in mathematics to focus on the logic of the argumentation. In a thought experiment, a simple situation is considered that may not be very realistic but which contains the essence of the problem, showing the most important arguments. In reality all sorts of detail will complicate the problem but the essence will remain the same. Thus, although the model does not incorporate all the difficulties involved when dealing with crime linkage in practice, it can highlight flaws in the reasoning and create a better understanding of the main line of reasoning.

The paper is structured as follows: In Section 8.2 we present a selection of the relevant literature and state of the art on crime linkage. In Section 8.3 we will model different situations in crime linkage using BNs, starting with the simplest example of two linked cases. We introduce and extend, step-by-step, to a network with three cases in Section 8.4 where the evidence is directly dependent on each other. In Section 8.5 we discuss our conclusions and give some ideas for future research.
8.2 Literature and state of the art on crime linkage

Crime linkage is a broad topic that has been extensively reported (for example in [12, 18, 3]). Here we focus only on two aspects of the literature that are relevant for our analysis, namely: (1) how to identify linked cases, and (2) how to model crime linkage. We discuss a (non extensive) selection of some key papers on these topics.

8.2.1 Literature on how to identify linked cases

For identifying linked cases, it is necessary to assess how similar two crimes are, how strong the link between the cases is and how sure we are that the offender in one case is also the offender in another case.

The authors of [4, 8, 9] investigate the behavioural aspects of sexual crime offenders in solved cases. These studies concentrate on the consistency of the behaviour of serial sexual assault offenders. The authors conclude that certain aspects of the behaviour can be regarded as a signature of the offender. These aspects can be used to identify possibly linked crimes.

The notion of such a ‘signature’ is discussed by Petherick in the chapter Offender Signature and Case Linkage [11]. It is noted that a signature in criminal profiling is a concept and not a ‘true’ signature. A may signature suggest that it is unique, whereas in criminal profiling it can only serve as an indication of whether or not two or more crimes are connected to each other.

Bennell and Canter [1] are interested in the probability (or indication) that two commercial burglaries are linked, given the modus operandi of these crimes. They use a database of solved commercial burglaries. Some of the burglaries studied had the same offender, which made it possible to identify behavioural features that reliably distinguish between linked and unlinked crime pairs. The authors present a model in which the distance between burglary locations and/or the method of entry can be used to determine the probability that the crimes are linked.

Tonkin et al. did a similar study [17]. They concentrate on the distance between crime locations and the time between two crimes to distinguish linked and unlinked crimes. They conclude that the distance between crime locations found and/or the temporal proximity is able to achieve statistically significant levels of discrimination between linked and unlinked crimes.

The discussed papers show that, in practice, it is possible to select certain features of crimes (like the distance or temporal proximity) to assign a probability to the event ‘the crimes are committed by the same person’. Taroni [13] discusses how such crime-related information may be used for the automatic detection of linked crimes.

8.2.2 Literature on modelling crime linkage

The papers discussed here focus on how to model possibly related crimes.

Taroni et al. [16] introduce Bayesian networks that focus on hypothesis pairs that distinguish situations where two items of evidence obtained from different
crime scenes do or do not have a common source. They show how Bayesian networks can help in assigning a probability to the event there is one offender responsible for both crimes. We concentrate on a different topic, namely the offender configuration (who is the offender in which case) and on how evidence implies guilt in one case influences the probability that a suspect is guilty in another case. Taroni et al. also present a Bayesian network for linking crimes with a utility and a decision node, which can help determine the direction for further investigation. Their study concentrates on how evidence from different cases influences the belief that there is a single offender responsible for both cases.

In Evett et al. [6] the hypothesis of interest does concern the offender configuration. Two case examples of similar burglaries are considered. In the first case the evidence consists of a DNA profile with a very discriminative random match probability and in the second case the only evidence is the report of an eye witness. The influence of the evidence in the first case on the question of guilt in the second case is investigated. They vary the strength of the evidence that suggests that there is one offender responsible for both cases to see how this influences the event that a suspect is guilty in the individual cases. The most important observation from their work is that when there is evidence that there is one offender responsible for both cases, the evidence in the individual cases becomes relevant to the other cases as well. This can either increase or decrease the probability that the suspect is the offender in a particular case. Evett et al. classify evidence into two categories that concern: (1) a specific crime only and (2) evidence that relates to similarities between the two crimes. We will introduce a third type of evidence that concerns both specific crimes as well as the similarity between crimes.

The case examples discussed by Evett et al. are viewed from the decision perspective of a prosecutor. The model they present should help to decide whether the prosecutor should charge a suspect with none, one or both crimes. However, Evett et al. do not consider the interesting question of what evidence should be presented when the suspect is charged with only one crime. We will show that it is wrong to select a subset of cases from a group of possibly linked cases and present only the evidence obtained in these cases. This is because evidence that is relevant in an individual case becomes of interest for the other cases when there exists a link between them.

In practical casework, the degree of similarity between crimes is usually poorly defined and lacks a rigorous mathematical treatment. While not solving this problem, we believe that the Bayesian network framework which we develop in this paper is a step in the right direction. It shows how to draw rational inference given certain assumptions and judgements of similarity (but where these judgements come from, and how they should be assessed is still a difficult question, and the topic of the literature mentioned in Section 8.2.1).

In what follows, we extend the work of Evett et al. by developing a generic Bayesian network. While they presented the necessary probabilities and relatedness structure needed for a Bayesian network they did not actually model a

---

1For simplicity, we shall assume that ‘guilty’ and ‘being the offender’ are equivalent even though in practice they are not. For instance, when a 4-year old kills someone, he may be the offender but he is not guilty of murder.
Bayesian networks when there are two linked crimes.

We further extend their work to situations with more than two crimes and present a type of evidence that they did not recognize in their paper, namely evidence supporting the claim that there is one offender responsible for multiple cases while simultaneously supporting the claim that the suspect is this offender. We will use an example to introduce and explain how different situations can be modelled using a Bayesian network. Most importantly, we show that it is not possible to ‘unlink’ crimes. When you have evidence that crimes are linked, all cases should be presented in court even when the suspect is charged for only a selection of them.

8.3 Using Bayesian networks when there are two linked crimes

In this section we introduce as a “thought experiment” the simplest example of two linked cases. In order to focus on the essence of crime linkage, we ignore in this paper important issues like the relevance of the trace, transfer-, persistence-, and recover probabilities, and background levels (see [15] for more realistic models). Also, we ignore all details in assessing the degree of similarity of observations, and simply say they ‘match’ or not, although we are aware that from a scientific point of view this is a problematic concept. We emphasize that in practice, these issues cannot be ignored.

8.3.1 The basic assumptions

Suppose that two crimes - each involving a single (but not necessarily the same) offender - have occurred and are investigated separately. In each case a piece of trace evidence, assumed to have been left by the offender, is secured. Our notion of a ‘trace’ is very general (in the sense described in [7]). It includes biological specimens like blood, hair and semen (from which e.g. a full or partial DNA profile can be determined), marks made (such as fingermarks, footmarks) or physical features as seen by an eye witness (such as height, hair colour or tattoos). In each case, the police has a suspect that ‘matches’ the trace. We label them as suspect 1 and suspect 2 for the suspects in crime 1 and 2 respectively.

The Bayesian networks for these cases are as in Figure 8.1. The (yellow) offender in case i nodes (i = 1, 2) have two states, ‘suspect i’ and ‘unknown’. The (pink) evidence nodes are conditionally dependent of the offender in case i nodes. They have two states, ‘match’ and ‘no match’. The probability tables for the offender in case i nodes are based on the possible offender population. Suppose that this possible offender population consists of 1000 men for each of the two crimes.\(^2\)

Assuming that every person is equally likely to be the offender when no other evidence is available gives a prior probability of 0.001 for the suspect being the offender in each case. For the (pink) evidence case i nodes, the probability that

\(^2\)The number of men in the possible offender population only sets the prior on all the hypotheses of interest. Using another number of men will give another outcome but the conclusions we draw still hold.
a random person matches determine the probability tables (for example, random match probabilities when the evidence concerns DNA profiles). Suppose that the random match probability\(^3\) for the evidence in case 1 is 0.0002 and for case 2 is 0.0003. Here, we assume that no errors occurred in the analysis of the evidential pieces and that the offender matches with certainty. So, the probability tables for the *evidence case* \(i\) nodes are as in Table 8.1.

\[
\begin{array}{cc}
\text{offender in case 1} & \\
\text{unknown} & \text{match} \\
0.9998 & 0.0002 \\
\hline
\text{offender in case 2} & \\
\text{unknown} & \text{match} \\
0.9997 & 0.0003 \\
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>offender in case 1</th>
<th>offender in case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>no match</td>
<td>0.9998</td>
<td>0.9997</td>
</tr>
<tr>
<td>match</td>
<td>0.0002</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

\[\text{Table 8.1: Probability tables for the evidence case 1 and evidence case 2 nodes}\]

The Bayesian network shows what inserting evidence does to the probability that the suspect is the offender. By setting the state of the *evidence case* \(i\) nodes to ‘match’ we get the posterior probability that suspect \(i\) is the offender, given the evidence. In this example, the posterior probability that suspect 1 is the offender in case 1 is 0.83 and the posterior probability that suspect 2 is the offender in case 2 is 0.77. The difference in posterior probability occurs because the random match probability in case 1 is lower than the random match probability in case 2. The probative value of the evidence in case 1 is therefore stronger. The same result is easily obtained by using formulas, see [5].

### 8.3.2 Similarity evidence

Now, suppose that suspect 1 and suspect 2 are the same person. Since this suspect matches with the evidence obtained in both cases, it appears that the cases are

\(^3\)We ignore here all practical difficulties in estimating these frequencies
linked by a common offender, the suspect. In what follows, we will construct a Bayesian network that models these (possibly) linked cases.

Next suppose that, in addition to evidence of similarity of offender, there is other evidence of similarity of the crimes. In contrast to evidence of ‘similarity of offender’ (which is human trace evidence in the sense explained in Section 8.3.1), evidence of ‘similarity of crime’ is not necessarily a human biological trace. This evidence could be, for example: a similar modus operandi, the time span between the two crimes, the distance between the two crime scenes, etc. In this example, we will use the evidence that fibres were recovered from the crime scenes that “matched” each other. In the second case, a balaclava is found at the crime scene. In the first case, fibres that match with the fibres from this balaclava are found. Since it is more likely to observe these matching fibres when the same person committed both burglaries than when two different persons did, the prosecution believe that there might be one person responsible for both crimes. Therefore, they want to link the crimes.

The network follows the description of the probability tables given in Evett et al.[6]. They discuss a crime linkage problem with two cases and use matching fibres as similarity evidence. However, they do use different individual crime evidence.

With two crimes, there are five possible scenarios regarding the offender configuration, namely:

1. The suspect is the offender in both cases.
2. The suspect is the offender in the first case; an unknown person is the offender in the second case.
3. An unknown person is the offender in the first case; the suspect is the offender in the second case.
4. An unknown person is the offender in both cases.
5. An unknown person is the offender in the first case; another unknown person is the offender in the second case.

The new Bayesian network, where we include the matching fibres evidence, is given in Figure 8.2. Note that this BN implies that the evidence from the individual cases is conditionally independent given the offender(s). We will examine the influence of this assumption in Section 8.3.3.

Again, the probability table for the offender configuration node is based on the assumption that the potential offender population consists of 1000 men. We added a (yellow) node, same offender 1&2. This node summarizes the scenarios in which the offender in the first case is the same person as the offender in the second case. The conditional probability table of the node is given in Table 8.2. Also, two (pink) evidence nodes are added, the fibres case i evidence nodes. These nodes have two states ‘type A’ (the type that is found on the crime scene and

---

*We do not distinguish between related and unrelated ‘unknowns’. Obviously, that does affect the random match probabilities but we are ignoring that for simplicity. Also, it does not affect the main argument of this paper.*
Figure 8.2: Bayesian network for linking two cases with similarity evidence

the balaclava) and ‘other’. To get the probability tables for these nodes, we need to determine how probable it is to observe fibres of type A. Suppose that the probability of observing this type of fibres in case 1 is 0.0001. We assume that if one person is responsible for both crimes, we will observe the same type of fibres in both cases. So, the probability tables are as in Table 8.3.

<table>
<thead>
<tr>
<th>Same offender 1&amp;2</th>
<th>Both</th>
<th>Suspect</th>
<th>First</th>
<th>Suspect</th>
<th>Second</th>
<th>Same</th>
<th>Unknown</th>
<th>Different</th>
<th>Unknowns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 8.2: Probability table for the same offender 1&2 node

Now, by inserting the matching fibres evidence and the matching evidence from the individual cases 1 and 2, we can compute the posterior probabilities for the suspect being the offender. The probability that the suspect is the offender in case 1, given the evidence, has increased to 0.9999. In case 2, this posterior probability also increased to 0.9999. The probability that the suspect is the offender in both

---

5In a more realistic setting, these numbers could be obtained by using a database of fibres. Also, other probabilities are involved, like the probability that the same balaclava was used by two different offenders. However, for this example the actual numbers are not that important, and we have chosen to follow the approach of Evett et al.[6].
8.3. Using Bayesian networks when there are two linked crimes

<table>
<thead>
<tr>
<th>fibres case 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>other</td>
</tr>
<tr>
<td>0.9999</td>
</tr>
<tr>
<td>type A</td>
</tr>
<tr>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>fibres case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>same offender 1&amp;2 no yes</td>
</tr>
<tr>
<td>fibres case 1 other type A</td>
</tr>
<tr>
<td>other</td>
</tr>
<tr>
<td>0.9999</td>
</tr>
<tr>
<td>type A</td>
</tr>
<tr>
<td>0.0001</td>
</tr>
</tbody>
</table>

Table 8.3: Probability tables for the fibres case i nodes

case 1 and case 2 follows from the offender configuration node. This posterior probability is 0.99988.

The example shows that, by including evidence that increases the belief that the offenders in case 1 and 2 are the same, this also increases the belief that the suspect is the offender in the individual cases. The similarity evidence makes it possible that the value of evidence obtained in one case is ‘transferred’ to another case. The simple line of reasoning is as follows. There is evidence that the two crimes are committed by the same person. There is evidence that crime 1 is committed by the suspect. The combination of these two pieces of evidence increases our belief that the suspect is the offender in crime 2, even without including the evidence found in crime 2. This also works the other way around, from crime 2 to crime 1.

Note that the likelihood ratio (LR), which is nowadays commonly reported by forensic experts, depends on the assumptions made about the prior probabilities of scenarios 1-5, see [2],[14]. This poses interesting reporting problems. However, this is not the main argument of this paper. In the following, we will focus on the posterior probabilities.

It is important to note that the use of matching fibres as similarity evidence is provided just for convenience. As mentioned in Section 8.2.1, the distance between crime scenes, the time between crimes, the modus operandi or certain behaviour of the offender can provide very strong evidence that two crimes are committed by the same person. We could include any combination of these as similarity of crime evidence, but one can imagine that it is harder to come up with the probability for observing a certain modus operandi in a case. We could also include the work of Taroni[16] that concentrates on the question of the strength of the link between the cases.

8.3.3 “Dependent” evidence

In the last example, we assumed that the evidence obtained in the individual crimes is independent of each other, given the offender(s). However, if the pieces of evidence are of the same type (DNA, footmarks, eyewitness descriptions), knowing that the offender in both crimes is the same person makes them conditionally
dependent. If one person is responsible for both burglaries, and we know that his DNA profile matches with the DNA profile obtained from the crime stain in case 1, it is certain (ignoring all considerations of relevance and various types of errors) that his DNA profile will also match the crime stain in case 2. For our example, we will concentrate on a situation where the evidence in the individual cases consists of two pieces, one of a type that is also found in the other case and one of a ‘case individual’ type.

In case 1, the evidence consists of a fingermark and a footmark of size 12. In case 2, the evidence consists of a partial DNA profile and a footmark of size 12. The suspect’s DNA profile matches with the partial DNA profile, his shoe size is 12 and his fingermark matches the fingerprint from case 1. Clearly, the footmarks from the individual cases are conditionally dependent given whether or not the offender in both cases is the same person. We assume that shoes with size 12 have a population frequency of 0.01. The random match probability of the fingerprint from case 1 is 0.02. The random match probabilities of the partial DNA profile from case 2 is 0.03. Using these numbers, the combined evidential value of the evidential pieces in an individual case (which we assume to be conditionally independent) is the same as in the situations of Figure 8.1 and 8.2. The Bayesian network describing this situation is given in Figure 8.3.

![Figure 8.3: Bayesian network for linking two cases with dependent DNA evidence](image)

The probability tables for the partial DNA profile and the fingerprint evidence nodes are similar to the ones given in Table 8.1 (with different random match probabilities). Only the probability table for the (pink) evidence node footmark size 12, case 2 is different. This node has three parents. It depends on who is the offender in case 2, but it also depends on whether there is one person who is the offender in both cases and the state of footmark size 12, case 1. The probability table for this node is given in Table 8.4.6

The important difference in Table 8.4, when we compare them to the probability tables given in Table 8.1, is that when we know that an unknown person is

---

6The situation offender in case 2 = suspect, same offender 1&2 = yes, footmark size 12, case 1 = no cannot occur. If the suspect is the offender in the second case and the offender in both cases is the same person, we know that he is also the offender in the first case. Hence, the footmark size 12 evidence in case 1 will be a match (assuming no mistakes). This is not a problem because the Bayesian network will never use these numbers in the computation.
8.4 Three linked cases

Now suppose that a third burglary comes up which is similar to the first two. Naturally, the prosecution would like to add this case in the link. With three crimes the number of possible scenarios for the offender configurations grows from 5 to 15, namely:

1. The suspect is the offender in all three crimes.
2. The suspect is the offender in the first two crimes. An unknown person is the offender in the third crime.
3. The suspect is the offender in the first and the third crime. An unknown person is the offender in the second crime.

:                     :

14. An unknown person is the offender in the first crime. Another unknown person is the offender in the second and the third crime.
15. Three different unknown persons are the offenders in the three crimes.

In Appendix 8.A we discuss the number of scenarios, given an arbitrary number, \( n \), of cases.

8.4.1 Assumptions about the evidence

Again, the evidence in this case consists of footmark size 12. The same fibres as in case 1 and 2 are found at the crime scene. The new same offender node summarises

<table>
<thead>
<tr>
<th></th>
<th>footmark size 12, case 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>offender in case 2</td>
</tr>
<tr>
<td>same offender 1&amp;2</td>
<td>no</td>
</tr>
<tr>
<td>footmark size 12, case 1</td>
<td>no match</td>
</tr>
<tr>
<td>no match</td>
<td>0.99</td>
</tr>
<tr>
<td>match</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 8.4: Probability table for the footmark size 12, case 2 node

the offender in both cases, the footmark size 12, case 1 evidence must match with the footmark size 12, case 2 evidence.

If we insert all the evidence (DNA, fingermark, footmark and fibres evidence), the posterior probability that the suspect is the offender in the first case is 0.99402. The posterior probability that the suspect is the offender in the second case is 0.99401, and the posterior probability that the suspect is the offender in both cases is 0.99399. Compared to the previous situation, we are slightly less confident that the suspect is the offender in both cases. This happens because the two items of footmark evidence are conditionally dependent on each other.

Although the posterior probabilities are slightly lower, it is still very likely that, given the evidence, the suspect is the offender in both cases.
which cases have a common offender and has 5 states; (1) one offender for all cases, (2) one offender for the first two cases, another for the third, (3) one offender for the first and third case, another for the second, (4) one offender for the first case, another for the second and third and (5) three different offenders. The probability tables of the nodes fibre evidence case \( i \) are similar to those in Table 8.3 and are also based on the assumption that the fibre type occurs with probability 0.0001.

The Bayesian network for this situation is given in Figure 8.4. The prior probabilities for the offender configuration have changed. Again we assume that the potential offender population consists of 1000 men. Under the assumption that each of these men is equally likely to be the offender in the individual cases, independently from each other we can compute the prior probabilities for all the scenarios. These are given in the fourth column of Table 8.5.

If we insert the evidence, matching DNA profile, matching fingermark, footmarks of size 12 and matching fibres between the cases, we get the posterior probabilities for the offender configuration, given the evidence. These are given in the last column of Table 4. The distribution of posterior probabilities shows that it is very likely that the suspect is the offender in all cases (with probability 0.99294). For the individual cases, the posterior probabilities that the suspect is the offender are 0.99399, 0.99397 and 0.99299 respectively.
8.4. Three linked cases

Table 8.5: Prior and posterior probabilities for the offender configuration node, given that the offender population consists of 1000 men. The posterior probabilities are obtained by inserting the evidence. X represents the suspect, 1, 2 and 3 are other unknown men. The configuration 1, 2, X stands for: An unknown man is the offender in the first case, another unknown man is the offender in the second case and the suspect is the offender in the third case.

<table>
<thead>
<tr>
<th>offender configuration</th>
<th>prior probability</th>
<th>posterior probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>X X X</td>
<td>1.00 · 10^{-9}</td>
<td>0.99</td>
</tr>
<tr>
<td>X X 1</td>
<td>9.99 · 10^{-7}</td>
<td>9.92 · 10^{-4}</td>
</tr>
<tr>
<td>X 1 X</td>
<td>9.99 · 10^{-7}</td>
<td>2.98 · 10^{-5}</td>
</tr>
<tr>
<td>1 X X</td>
<td>9.99 · 10^{-7}</td>
<td>1.98 · 10^{-5}</td>
</tr>
<tr>
<td>X 1 1</td>
<td>9.99 · 10^{-7}</td>
<td>2.98 · 10^{-5}</td>
</tr>
<tr>
<td>X 1 2</td>
<td>9.97 · 10^{-4}</td>
<td>2.97 · 10^{-6}</td>
</tr>
<tr>
<td>1 X 1</td>
<td>9.99 · 10^{-7}</td>
<td>1.98 · 10^{-5}</td>
</tr>
<tr>
<td>1 X 2</td>
<td>9.97 · 10^{-4}</td>
<td>1.98 · 10^{-6}</td>
</tr>
<tr>
<td>1 1 X</td>
<td>9.99 · 10^{-7}</td>
<td>5.95 · 10^{-7}</td>
</tr>
<tr>
<td>1 1 1</td>
<td>9.99 · 10^{-7}</td>
<td>5.95 · 10^{-3}</td>
</tr>
<tr>
<td>1 1 2</td>
<td>9.97 · 10^{-4}</td>
<td>5.94 · 10^{-6}</td>
</tr>
<tr>
<td>1 2 1</td>
<td>9.97 · 10^{-4}</td>
<td>5.94 · 10^{-6}</td>
</tr>
<tr>
<td>2 1 1</td>
<td>9.97 · 10^{-4}</td>
<td>5.94 · 10^{-6}</td>
</tr>
<tr>
<td>1 2 3</td>
<td>0.99</td>
<td>5.92 · 10^{-7}</td>
</tr>
</tbody>
</table>

8.4.2 Evidence proving innocence in the third case

A piece of exculpatory evidence is found in the third case. In our example, an eyewitness description of the offender states that the offender has a permanent tattoo on his left arm. If the suspect does not have a tattoo on his left arm, and we assume that the eyewitness description is correct, i.e. the actual offender has a permanent tattoo, it is certain that the suspect is not the offender in the third case. Now, the prosecution can do two things, (1) drop the third case and go to court with the first two cases, where they have strong evidence of the suspect’s guilt, or (2) go to court with all three cases. The prosecution could argue that both options amount to the same outcome. In the first one, they drop the third case and use the evidence of the first two cases. In the second option, the prosecution uses all three but, since they have evidence that the suspect is innocent in the third crime, they are only interested in whether the suspect is guilty in the first two cases. We will show that the first option is wrong since it withholds exculpatory evidence from the court for the first two cases.

When linking crimes, one needs to be aware that the sword cuts both ways. As we saw, if there is evidence in a case suggesting that there is one person responsible for both cases, evidence in one case is of interest for the question whether or not a suspect is guilty in another case. This means that if there is evidence in the first
case that increases your belief that the suspect is the offender in the first case, it will also increase your belief that the suspect is the offender in the second case. This also works the other way around and is just as relevant: if there is evidence that a suspect is innocent in one case, this should also increase your belief that the suspect is innocent in the second case. This is illustrated by the example.

Suppose that it is known that the proportion of men with a tattoo on their left arm is $1/25.$ The Bayesian network representing the situation is given in Figure 8.5. Remember that if we do not include the third case, we are in the situation of Figure 8.3.

![Bayesian network for linking three cases with exculpatory evidence in the third case](image)

**Figure 8.5:** Bayesian network for linking three cases with exculpatory evidence in the third case

To compare the outcome in terms of the posterior probabilities when one drops or includes the third case, we compare the posterior probabilities of the offender configuration node of the models from Figure 8.3 and 8.5. This is done in Table 8.6. The posterior probabilities for the suspect being the offender in the individual cases under the situation where the third case is dropped and the situation where the third case is included are given in Table 8.7.

---

7In this case, where we insert as evidence that there is no match with the suspect, the probability is irrelevant since it is impossible to observe no match when the suspect is the donor (assuming no errors were made). In a situation where the evidence does not directly show that the suspect is innocent but where it only increases one’s belief that he is innocent, the random match probability is relevant.
8.4. Three linked cases

<table>
<thead>
<tr>
<th>offender 1</th>
<th>offender 2</th>
<th>offender 3</th>
<th>consider 2 cases</th>
<th>consider 3 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0.994</td>
<td>0.00</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>1</td>
<td>0.141</td>
<td></td>
</tr>
</tbody>
</table>

| X          | 1          | X          |                  | 0.00             |
| X          | 1          | 1          | 2.979 \cdot 10^{-5} | 0.004            |
| X          | 1          | 2          |                  | 4.236 \cdot 10^{-6} |

| 1          | X          | X          | 0.00             |                  |
| 1          | X          | 1          | 1.986 \cdot 10^{-5} | 0.003            |
| 1          | X          | 2          |                  | 2.824 \cdot 10^{-6} |

| 1          | 1          | X          | 0.00             |                  |
| 1          | 1          | 1          | 0.006            | 0.849            |
| 1          | 1          | 2          |                  | 8.472 \cdot 10^{-4} |

| 1          | 2          | X          | 0.00             |                  |
| 1          | 2          | 1          | 5.946 \cdot 10^{-6} | 8.472 \cdot 10^{-4} |
| 2          | 1          | 1          |                  | 8.472 \cdot 10^{-4} |
| 1          | 2          | 3          |                  | 8.447 \cdot 10^{-7} |

Table 8.6: Posterior probabilities for the offender configuration node, for a situation where the third case is dropped and a situation where the third case is included.

<table>
<thead>
<tr>
<th>consider 2 cases</th>
<th>consider 3 cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>case 1</td>
<td>0.992</td>
</tr>
<tr>
<td>case 2</td>
<td>0.991</td>
</tr>
</tbody>
</table>

Table 8.7: Posterior probabilities for the suspect being the offender in case 1 and 2 for a situation where the third case is dropped and a situation where the third case in included.

The tables show that excluding or including the third case has serious consequences for the posterior probabilities, and thus, the outcome of a possible trial. When the third case is dropped, one can confidently state that it is very likely that the suspect is the offender in the first two cases. If we use the following two hypotheses,

\( H_p \): The suspect is the offender in case 1 and 2.

\( H_d \): The suspect is not the offender in case 1 nor in case 2.

The posterior odds can be computed as

\[
\frac{\mathbb{P}(H_p | E)}{\mathbb{P}(H_d | E)} = \frac{0.994}{5.958 \cdot 10^{-3} + 5.946 \cdot 10^{-6}} = 167
\]
The conclusion would be: Based on the observed evidence and the prior assumptions, it is 167 times more likely that the suspect is the offender in case 1 and 2 than that he is not the offender in case 1 nor in case 2. When we include the third case and use the same hypothesis pair, the posterior odds become,

\[
\frac{P(H_p|E)}{P(H_d|E)} = \frac{0.14148}{0.849 + 8.472 \cdot 10^{-4} + 8.472 \cdot 10^{-4} + 8.472 \cdot 10^{-4} + 8.447 \cdot 10^{-7}} = 0.17
\]

Now, the conclusion would be: Based on the observed evidence and the prior assumptions, it is 6 times more likely that the suspect is not the offender in case 1 nor in case 2 than that he is the offender in case 1 and 2.

It is important to understand that the probability tables and assumptions under both models are the same. The only thing we changed is including the third case. The underlying reasoning is as follows. There is evidence that there is a common offender in the three cases. Both the matching fibres and the footmark evidence account for this. Also, there is evidence that the suspect is the offender in the first and in the second case. This is the partial DNA profile, the footmark and the fingermark evidence. Due to the similarity evidence, this not only increases our belief that the suspect is the offender in the cases where this evidence was found, it also increases the belief that the suspect is the offender in the other cases. However, the evidence in case 3 shows that the suspect is innocent in the third case. So, there is another unknown person responsible for the third crime. Due to the similarity evidence, this also increases our belief that this unknown person is responsible for the first and the second crime. This is the double-edged sword when linking crimes.

The example shows that one cannot ‘unlink’ a case because of evidence suggesting that the suspect is not the offender. Although this is clear from the example shown, it seems likely that in practice people might simply drop a case without being aware of the consequences it has on the validity of their conclusions. One might even think that it benefits the suspect to drop the case, since the number of cases in which he can be found guilty decreases.

8.5 Discussion

In reality, crime linkage is very complex. We would like to emphasize again that issues like relevance of trace material and many other uncertainties are essential to consider. The presented crime linkage model simplifies the reality and does not capture all the problems that play a role when linking crimes. Hence, we do not recommend the use of the model presented here in actual casework (although more detailed BN models can be used to incorporate many of the other relevant issues also). We do think it is useful for uncovering the interesting aspects of the reasoning underlying crime linkage, and as such assists in understanding and dealing with it in practice.

We have shown, using a simple example, how a Bayesian network can help us understand and interpret evidence in cases where crimes are linked. Also, we have shown how to model cases with “dependent” evidence. It is possible to categorize evidence in a crime linkage problem into three categories.
1. Evidence relevant for the question of who the offender is in a specific case.

2. Evidence relevant for the question of whether the offender in two cases is the same person.

3. A combination of 1 and 2: Evidence relevant for both questions.

The first two categories are mentioned in Evett et al. [6]. The third category is a combination of the first two. For example if the similarity evidence is a match between fibres found at the different crime scenes, it falls in the second category. If, in addition, a sweater is found at the house of the suspect which fibres match with the fibres found at the crime scene, it belongs in the third category.

When linking more than two crimes, the combined effect of different pieces of evidence, i.e. how one observation influences another, rapidly becomes more complex. The use of Bayesian networks helps us understand the relations between observations. In our example we have shown a model where three crimes are linked. Using a Bayesian network the problem breaks down to filling the entries of some very straightforward probability tables. We have only shown Bayesian networks for two and three crimes.

The number of possible offender configurations grows exponentially when the number of linked cases increases. Although we could use a computer to build a Bayesian network linking, e.g., twenty crimes and to fill the probability tables, the computation time will also increase according to the increase in offender configurations. The number of offender configurations with twenty crimes is 474,869,816,156,751 [10]. When the number of linked crimes is not that high (say less than 10), the method to present and understand the relations between evidence described should help provide insight into the problem. For large numbers of linked crimes, further research needs to be done. More on this can be found in Appendix 8.A.

Most importantly, we have shown that one cannot ‘unlink’ cases. When there exists a link between cases, so there is evidence that there is one offender responsible for both cases, the cases should be treated simultaneously. In (forensic) practice, a similar thing occurs when multiple traces are secured of which the location of the traces suggest that they belong to one person, i.e. in a situation where fingermarks are recovered from an object. If these traces lay close to each other and form a grip pattern, it is likely that they belong to the same hand. Now, if only some of the fingermarks are similar to finger prints obtained from a suspect while the others are not, it is wrong to focus on the similarity evidence only.

It would be interesting to see how judges and police investigators deal in practice with cases that appear to be linked, where evidence in one case points towards a suspect whereas in the other case the evidence suggests that the suspect is innocent. Our own limited experience is that the relevance of exculpatory evidence found in one case for other similar cases is underestimated. This hypothesis can be tested in properly designed experiments.

Besides modelling linking of a large number of crimes, future research could study more complex situations of linked crimes. The research presented here could be expanded to situations where there is a group of criminals that e.g. rob houses together in various group compositions. In these cases it is possible to have a very
similar and distinctive modus operandi, while the evidence in the different cases could point to different suspects.

8.A The number of scenarios given the number of cases

The number of possible offender configurations increase very rapidly when the number of cases increases. For one case the number of configurations is 2 (suspect is the offender, unknown is the offender). For two cases it is 5 and for three cases it is 15. The number of scenarios given the number of cases is an example of so-called Bell numbers [10]. For 1 to 10 cases, the number of configurations is given in Table 8.8.

<table>
<thead>
<tr>
<th>cases</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>configurations</td>
<td>2</td>
<td>5</td>
<td>15</td>
<td>52</td>
<td>203</td>
<td>877</td>
<td>4140</td>
<td>21147</td>
<td>115975</td>
<td>678570</td>
</tr>
</tbody>
</table>

Table 8.8: The number of possible offender configurations for 1 to 10 cases

The \( n \)th Bell number represents the number of partitions of a set with \( n \) members, or equivalently, the number of equivalence relations on it. Bell numbers satisfy the recursion formula in 8.1.

\[
B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k
\]  

(8.1)

The \( n \)th Bell number corresponds with the number of offender configurations for \( n - 1 \) cases.

We see that the number of possible offender configurations grows rapidly when we increase the number of cases. Therefore, drawing conclusions becomes more difficult when the number of cases increases. Not regarding all possible offender configurations to decrease the number of scenarios might not be the solution to overcome this problem. Every offender configuration represents an interesting different situation. Especially when the number of cases is large (say 50), using less offender configurations will most likely mean that the new number of scenarios is too limited or still too large. For example when we would only include the suspect is the offender in all cases or another unknown person is the offender in all cases, we are disregarding too many situations. A solution might be limiting the number of different offenders we allow. In a situation where there are 50 similar crimes, it may not be necessary to allow for the possibility that all crimes are committed by different men. Either way, the modelling of very large numbers of possibly linked cases is interesting for further research.

Bibliography


Evaluating evidence in crime linkage scenarios with multiple offenders

Abstract

In de Zoete et al. [4] a framework for the evaluation of evidence when an individual is a suspect of two separate offenses (based on Evett et al. [6]) is implemented using a Bayesian network. Here, we extend this to situations with multiple offenders. These situations differ with respect to two things: (1) whether or not it is possible to distinguish between the offenders, and (2) whether or not one can group several pieces of evidence as originating from the same person. With the aid of a mock case example, we show that evidence evaluation soon becomes complex. Also, these subtle differences between situations can lead to substantially different conclusions.

Although we find it undesirable that Bayesian networks are demonstrated in court, they can be very helpful in guiding expert and legal reasoning, identifying pitfalls and assist in preventing them. Bayesian networks can be used as a tool to understand how the different pieces of evidence in multiple offender crime linkage influence each others evidential value, and the probabilities of the hypotheses of interest.

9.1 Introduction

In legal casework, it is not uncommon that an individual is a suspect of multiple (similar) offenses. In these situations, the evaluation of the evidence becomes rather complex. Most importantly, when considering the culpability of a suspect for a crime, one needs to consider the evidential value from observations of other, similar, crimes. In Evett et al. [6] a method is suggested to evaluate the evidence when a person is a suspect in two separate offenses. The authors show that, in such a situation, the evidence can be grouped into two categories: (1) evidence at the time of this dissertation’s submission to the examination committee, the work in this chapter is under review as a manuscript.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

which is only relevant for a specific crime, and (2) evidence which is relevant for
the connection between crimes. De Zoete et al. [4] extend the analysis of Evett et
al. [6] by recognizing another category of evidence: (3) evidence relevant for both
the link between crimes and for a specific crime. This occurs when similar pieces of
evidence that match characteristics of the suspect are obtained in different crimes.
Furthermore, in de Zoete et al. [4] Bayesian networks are introduced as a reasoning
tool for crime linkage. Both Evett et al. [6] and de Zoete et al. [4] only consider
situations with a single offender.

In this work, the focus lies on evidential reasoning in crime linkage problems
with multiple offenders. We identify different types of multiple offender crime
linkage situations. For each situation we present a Bayesian network.

The paper is structured as follows. Literature on (multiple offender) crime
linkage is discussed in Section 9.2. In Section 9.3 different situations, each resulting
in a different Bayesian network, are presented. Because crime linkage with multiple
offenders becomes rather complex, even with a small number of cases, we use a
simple mock case example. For this example, we make simplifying assumptions,
for example that all the evidence was left by the offenders. The different situations
are distinguishable by (1) whether or not it is possible to distinguish between the
offenders and (2) whether or not it is possible to ‘group’ the evidence. Studying
the behavior of posterior probabilities\footnote{In forensic statistics, it is common to work with likelihood ratios. However, likelihood ra-
tios for hypotheses that are collections of multiple sub-hypotheses can only be determined when
reasonable priors for these sub-hypotheses are available, something which is often not the case.
Nevertheless, for demonstration purposes we will choose priors and work with posterior proba-
bilities.} for the mock example results in general
lessons for these type of problems. These results are presented in Section 9.4. In
Section 9.5 we summarize the general lessons following from the mock example.
We also identify pitfalls in probabilistic reasoning. We conclude that in forensic
casework, these Bayesian networks can be very valuable, not as a tool to compute
probabilities that can be reported, but as a tool that can assist in supporting our
reasoning and in preventing pitfalls.

9.2 Literature overview

In this section we review some aspects of linked crimes in the literature. This sketch
serves as the background against which we present our probabilistic contributions
in the sections to follow. A broad selection on papers regarding crime linkage can
be found on the website of Crime Linkage International NetworK [3].

9.2.1 Identifying linked crimes

In Grubin et al. [9] it is noted that the most reliable means of establishing a link
between offenses is by ‘physical’ evidence (e.g. DNA profiles/footwear evidence).
For example, in de Zoete et al. [4], an example is presented in which the link be-
tween cases consists of matching fibre evidence. However, Winter et al. [26] notes
that such evidence is often absent. In these situations, non-physical evidence can
assist in determining the strength of the link between two crimes. In Woodhams et al. [27], linkage analysis is defined as a process that aims to identify crimes that are likely to have been committed by the same perpetrator. In this, and many other papers [20, 2], the authors are interested in the *behavioural* similarity between two crimes.

Different methods for different types of crimes have been suggested to assess whether there exists a link between crimes. The focus in these papers often concerns the question whether it is possible to determine whether two crimes share a common offender. For investigative purposes, regarding two crimes as ‘linked’ could assist in identifying suspects. However, in a criminal trial, one cannot disregard the uncertainty associated with this link. In [16], a statistical approach is introduced that computes Bayes factors that reflect the strength of the evidence that two cases are linked. Bennell and Canter [1] study the predictive value of different features associated with commercial burglaries for determining whether two crimes have (a) common offender(s). These can assist in estimating how likely it is that two burglaries have a common offender given (for example) the distance between the two crime scenes, the stolen property and/or entry behaviours. And, likewise, in Ribaux et al. [19] characteristics from shoe marks, tool marks and/or glove marks observed in different crimes are used to identify possibly linked burglaries. Summarized, Ormerod and Stirman [15] mention that linkage analysis can be pivotal in assessing whether the same individual is responsible for similar offenses.

Apart from the investigative phase of possibly linked crimes, one should also keep in mind that the crime linkage procedure should be admissible in court. Labuschagne [12] states that linkage analysis is especially important when there is evidence that suggest culpability for some, but not all offenses. However, in Fawcett and Clark [7], the authors note that the linkage analysis is, at present, unrepresentative evidence that cannot *independently* indicate a defendant’s culpability. In other words, linkage analysis ‘alone’ is insufficient evidence. One needs crime specific evidence to make it legally admissible. The same is recognized in de Zoete et al. [5] in which the differences between a Dutch legal and a probabilistic perspective are discussed. The authors of [5] show that, although it is common to require such crime specific evidence when linking crimes in practice, it is unnecessary from a probabilistic point of view.

In Evett et al. [6], a framework is introduced that can assist in evaluating the evidence when an individual is a suspect in two separate offenses. De Zoete et al. [4] extend this analysis with the aid of Bayesian networks. The emphasis in these papers lies on understanding the combined evidential value and acknowledging the (potentially large) set of alternative explanations. Most notably, it is recognized that the combination of these two can result in situations in which the evaluation of the evidence becomes very complex. With multiple offenders, the set of alternative explanations that should be regarded grows even more rapidly. Furthermore, one should recognize that there are several situations that one can regard when linking crimes with multiple offenders with only subtle differences between them. A different situation can result in a substantial different belief of the culpability of the suspect(s), see Section 9.3.
9.2.2 Consistency and behaviour of groups in serial crime

A substantial part of the literature on crime linkage focuses on a situation with only one offender per crime. However, Burrell et al. [2] examines the differences in behavioural similarity for different crime linkage situations. The authors did not observe statistical differences between the behavioural similarity for a link between two lone offender offenses and two group offender offenses. In other words, there is little reason to expect substantial differences in the behavioural similarity when dealing with multiple offender crime linkage compared to single offender crime linkage. Furthermore, Burrell et al. [2] investigate the differences in behavioural similarity when one of the crimes was performed by a lone offender and the other by a group (in which the lone offender was present). The authors advise to take some caution when linking such crimes.

Regarding the size of offender groups, Reiss and Farrington [18] and Walsh [23] identified that the majority of multiple offender crimes had two offenders. Co-offender stability, i.e. the tendency to select the same co-offender for two consecutive crimes is discussed in multiple papers. In [24] youth surveys are used to examine this co-offender stability. The authors conclude that, if offenders commit multiple robberies in a short span of time, they are more likely to select from the same group of companions. For different types of crimes, there was less chance that offenders would select from the same group of companions. Warr [24] believes that this could be due to some kind of group specialisation. Apart from [24], also Klein and Crawford [11], Reiss [17] and Short [21] note that delinquents commonly have a larger network of co-offenders than would be expected from the size of their offending groups. Contrary to Warr [24], Weerman [25] and McGloin et al. [13] conclude that the same offender composition is unlikely to emerge on more than one occasion. Reiss and Farrington [18] state that co-offending pairs tend to be short lived. Nonetheless, in either situation, an interpretation framework like a Bayesian network can assist in evaluating the evidence.

9.3 Modelling crime linkage with multiple offenders

In this section we identify different situations when linking multiple offender crimes. The different situations are introduced with the aid of a mock case example. For each situation we construct a Bayesian networks. The posterior probabilities following from these networks are computed in Section 9.4.

Mock case example Two burglaries occur shortly after each other. In both crimes, there were two offenders. The evidence in the first crime consists of a partial fingermark (showing an arch pattern), two shoe marks (possibly from two different shoes) and a single red hair. In the second crime, a partial fingermark (showing an arch pattern), a single red hair, a shoe mark and a partial DNA profile are obtained. The police has two suspects who, together, ‘match’ with all the evidence, see Table 9.2. It is assumed that all the evidence was left by the offenders. An overview of the pieces of evidence and the associated random match probabilities can be found in Table 9.1.
We do not believe that constructing a ‘complete’ probabilistic model for such a case is feasible. Nevertheless, a probabilistic model for a simplified situation can still provide valuable lessons in terms of the validity and implication of certain assumptions. Additionally, it can serve as a reasoning tool to find out which alternative explanations are reasonable or even likely. Assumptions regarding random match probabilities are made for illustrative purposes and are not necessarily realistic.

<table>
<thead>
<tr>
<th>number</th>
<th>evidence</th>
<th>random match probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>partial fingermark (arch pattern)</td>
<td>0.01</td>
</tr>
<tr>
<td>2.</td>
<td>shoe mark (size 11)</td>
<td>0.02</td>
</tr>
<tr>
<td>3.</td>
<td>shoe mark (size 11)</td>
<td>0.02</td>
</tr>
<tr>
<td>4.</td>
<td>red hair</td>
<td>0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>number</th>
<th>evidence</th>
<th>random match probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.</td>
<td>partial fingermark (arch pattern)</td>
<td>0.01</td>
</tr>
<tr>
<td>6.</td>
<td>red hair</td>
<td>0.03</td>
</tr>
<tr>
<td>7.</td>
<td>shoe mark (size 11)</td>
<td>0.02</td>
</tr>
<tr>
<td>8.</td>
<td>partial DNA profile</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 9.1: Overview of the obtained evidence from the mock case example.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

<table>
<thead>
<tr>
<th>evidence</th>
<th>match with suspect X</th>
<th>match with suspect Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>partial DNA profile</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>partial fingermark (arch pattern)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>red hair</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>shoe mark, size 11</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 9.2: Overview of the evidence related characteristics of the suspects.

9.3.1 Different situations

We identify several different situations. Each situation comes with another Bayesian network structure and, therefore, with a different conclusion in terms of posterior probabilities. Situations differ from each other based on (1) whether it is possible to distinguish between the offenders and (2), whether it is possible to link different pieces of evidence in terms of belonging to the same offender.

Situation 1: distinguish between offenders, evidence can be linked to offenders

In this situation, it is possible to distinguish between offenders. For instance, it may be known that one of the offenders (in both cases) is a man and the other is a woman, or one is tall and the other is short, or offenders with different skin colours, etcetera. The number of alternative explanations (regarding the set of offenders) differs from situations where it is not possible to distinguish between offenders, see Section 9.3.2. For example, in this situation where one of the offenders is a man and the other is a woman, it is impossible that the male suspect committed the first crime with an unknown person while the second crime had this same unknown person as an offender, together with the female suspect (the unknown cannot be both male and female). Furthermore, in this situation we assume that it is possible to link the evidence to the different offenders. For example, when the modus operandi in these burglaries was that the female offender waited outside while the male offender went inside, evidence that is obtained from inside can be linked to the male offender whereas evidence from outside can be linked to the female offender. A pictorial representation of this situation can be found in Figure 9.1.

Note that the assumption that it is possible to distinguish between the offenders must come from background information like, for instance, a witness who stated that a man entered the place and a woman stayed outside. One could take the trustworthiness of this witness into account, which would add a layer of complexity to the model. Alternatively, one could assume that the information provided by the witness is correct and evaluate the simplified model. In Situation 1, 2 and 3 it is assumed that this assumption can be made. Furthermore, within this mock case example, it is assumed that the evidence obtained in the different cases (shoe marks, fingerprints, red hairs and the partial DNA profile) are uninformative about the distinction between the offenders. In reality, this need not be the case,
9.3. Modelling crime linkage with multiple offenders

for instance since men generally have a larger shoe size or when the partial DNA profile contains information on the sex of the donor. Once again, one could model this, but since our interest is in reasoning and not in calculus, we opted for the simpler model.

Figure 9.1: Pictorial representation of Situation 1 in which it is possible to distinguish between the offenders (male/female) and in which the different pieces of evidence can be linked to the different offenders. The blue symbols represent evidential pieces that can be associated with the first (male) offender. The orange symbols represent evidential pieces that can be associated with the second (female) offender.

Situation 2: distinguish between offenders, evidence can be linked to each other

Compared to the first situation, it is still possible to distinguish between offenders (male/female), but there is no direct link between offenders and evidence. However, in this situation, it is possible to ‘link’ different pieces of evidence. In other words, it is not known which offender left what evidence, but it is known which sets of pieces were left by the same offender. For example, it is known that both crimes had a male and a female offender and that, in both crimes, one of them entered the building, but we do not know who, and the other stayed outside. Therefore, it is known that all the evidence outside belongs to one offender where the evidence inside belongs to the other offender. A pictorial representation of this situation can be found in Figure 9.2.

Situation 3: distinguish between offenders, no link between evidence

Also in this situation, it is possible to distinguish between offenders, but it is unknown who left what piece of evidence or what pieces of evidence are ‘grouped’. This may happen, for instance, when both offenders entered the building. A pictorial representation of this situation can be found in Figure 9.3.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

**Figure 9.2:** Pictorial representation of Situation 2. Evidence of which it is known or assumed that it was left by the same offender has the same colour. Furthermore, it is assumed that one group of evidence was left by the first offender and the other group by the second offender.

**Figure 9.3:** Pictorial representation of Situation 3. It is possible to distinguish between offenders. There is no link between the pieces of evidence, i.e. it impossible to group the evidential pieces as in the second situation.

**Situation 4: cannot distinguish between offenders, evidence can be linked to each other**

In this situation, it is not possible to distinguish between offenders, but it is nevertheless possible to group the evidence as in Situation 2. A pictorial representation of this situation can be found in Figure 9.4.
9.3. Modelling crime linkage with multiple offenders

**Figure 9.4:** Pictorial representation of Situation 4. Two indistinguishable offenders and evidence that can be grouped.

**Situation 5: cannot distinguish between offenders, no link between evidence**

In this situation, it is impossible to distinguish between offenders and in addition there is no way to group the pieces of evidence regarding whether they were left by the same offender. This is a situation with the least amount of background information. A pictorial representation of this situation can be found in Figure 9.5.

**Figure 9.5:** Pictorial representation of Situation 5. Two indistinguishable offenders and evidence that cannot be grouped.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

Situation 6: unknown number of offenders

Until now we assumed that it was known that both crimes were committed by two criminals. In this last situation, we deviate slightly from this assumption in the sense that we assume that both crimes were committed by at most two offenders. A pictorial representation of this situation can be found in Figure 9.6

![Figure 9.6: Pictorial representation of Situation 6. An unknown number of offenders, the evidence can not be grouped or linked to any offender.](image)

9.3.2 Alternative hypotheses: offender configurations

The complete set of hypotheses that one can consider was dubbed the set of offender configurations in de Zoete et al. [4]. In the same paper it was observed that the number of such hypotheses for crime linkage with one offender increases very rapidly with the number of crimes. This effect is even stronger in multiple offender crime linkage situations. As an example, when analyzing a case of four possibly linked crimes with two ‘indistinguishable’ offenders and two suspects, the number of possible hypotheses is 4900. Next we identify the hypotheses for Situations 1-6 above.

For Situation 1, Situation 2 and Situation 3, it is possible to distinguish between offenders (for example a male and a female offender). Hence, for instance the alternative explanation

<table>
<thead>
<tr>
<th>crime</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>suspect X</td>
<td>unknown female 1</td>
<td></td>
</tr>
<tr>
<td>unknown male 1</td>
<td>suspect Y</td>
<td></td>
</tr>
</tbody>
</table>

should be considered, whereas,
9.3. Modelling crime linkage with multiple offenders

is an impossible alternative explanation (both offenses have one male and one female offender). The complete set of hypotheses for these situations with 2 crimes, 2 offenders and 2 suspects consists of 25 elements which can be found in the appendix, Table 9.12. Although we write “offender 1” and “offender 2” in the table, one should think of these as unordered. We only label them 1 and 2 because this will be convenient when we construct Bayesian networks below.

By removing the assumption that one can distinguish between offenders (Situation 4 and 5) the number of hypotheses increases. The set of hypotheses for our example with two crimes and two suspects consists of 29 elements which can be found in the appendix, Table 9.13, with the same comment as for the previous table. A table for Situation 6 can be given analogously.

9.3.3 Bayesian networks representing the different situations

In this section, we present Bayesian networks for all six situations. In Section 9.4, we present the posterior probabilities obtained from these networks.

The Bayesian networks we consider all contain the following sub-network.

The **offender configuration** node has no parents, and it can take any of the 25 respectively 29 values from the above-mentioned tables of hypotheses in the Appendix, depending on the situation. We already commented on the difficulty, or perhaps sometimes impossibility, to assign reasonable prior probabilities. But we also mentioned the fact that this article is more about reasoning than about calculus, so we do choose prior probabilities for this node. To keep things as simple as possible, we assume that the number of people in the possible offender population is known, and in the situations that it is possible to distinguish between the offenders, we assume that it is known how many of each group are present in the offender population. We impose uniform priors on these populations.

The other nodes in this basic block, **same offenders** and (**offender i, crime j**) are, strictly speaking not really needed, but they are included because they are
9. Evaluating evidence in crime linkage scenarios with multiple offenders

Convenient. The \((\text{offender } i, \text{crime } j)\) nodes simply repeat the information of \textit{offender configuration} in the (arbitrary) order given in the table, and the \textit{same offenders} node tells us which offenders are the same in the appropriate offender configuration. In the situations where the offenders can be distinguished as say, male and female, this node takes the values “both”, “none”, “male”, and “female”, to indicate which pair(s) of criminals is (are) the same. In all other situations, this node simply expresses which of the offenders are the same.

**Situation 1, distinguish between offenders, evidence can be linked to offenders**

In Situation 1 (see Figure 9.1), it is possible to distinguish between offenders and the different pieces of evidence can be linked to the different offenders. See Figure 9.8 for the Bayesian network for this situation.

![Bayesian network](image)

**Figure 9.8:** Bayesian network representing Situation 1, see Figure 9.1.

The conditional dependence of pieces of evidence of the same type (shoe marks, fingerprints) is ‘linearly’ modeled, i.e. from left to right. There are other ways to model this (i.e. from right to left or by using a symmetric structure; all of these are mathematically equivalent. The probability tables of the evidence nodes are as in Table 9.3 and 9.4.
### 9.3. Modelling crime linkage with multiple offenders

<table>
<thead>
<tr>
<th>offender 1, crime 1</th>
<th>suspect X</th>
<th>unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>arch pattern</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>no arch pattern</td>
<td>0</td>
<td>0.99</td>
</tr>
</tbody>
</table>

**Table 9.3:** Conditional probability table for fingermark 1 node, Situation 1. Nodes with only one parent node have a similar structure.

<table>
<thead>
<tr>
<th>off 1, crime 2</th>
<th>suspect X</th>
</tr>
</thead>
<tbody>
<tr>
<td>same offenders</td>
<td>both</td>
</tr>
<tr>
<td>fingermark 1</td>
<td>arch</td>
</tr>
<tr>
<td>arch pattern</td>
<td>1</td>
</tr>
<tr>
<td>no arch pattern</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>off 1, crime 2</th>
<th>unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>same offenders</td>
<td>both</td>
</tr>
<tr>
<td>fingermark 1</td>
<td>arch</td>
</tr>
<tr>
<td>arch pattern</td>
<td>1</td>
</tr>
<tr>
<td>no arch pattern</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 9.4:** Conditional probability table for fingermark 2 node, Situation 1. Nodes with multiple parent nodes have a similar structure. Note that some entries represent impossible states. These are marked with an asterisk, *.

**Situation 2, distinguish between offenders, evidence can be linked to each other**

Since in this situation (see Figure 9.2) it is unknown which offender (the man or the woman) left which pieces of evidence, there are more conditional dependencies compared to Situation 1. For example, in the previous situation, it was certain that the ‘female’ offender in the first crime and the ‘male’ offender in the second both had red hair. In this situation, this is no longer certain. We present a Bayesian network representing this situation in Figure 9.9.

**Donor** $i \& j$ nodes are parental nodes. They are added to determine which pieces of evidence are left by the same offender. These also determine which offender this is. For example, **donor 3&4** has two states **male offender** and **female offender**. When it is initiated in **male offender**, **shoe mark 2** and **red hair 1**, are attributed to the male offender. The constraint nodes make sure that both the male and the female offender left a part of the evidence. Hence, when **donor 3&4** is initiated as **male offender**, **donor 1&2** will be **female offender**. Again, there are several ways we can model the conditional dependence relation between pieces of evidence that are of the same type. As in Situation 1, we opted to model this ‘linearly’, i.e. from left to right. Hence, evidence nodes belonging to crime 2 have more parental nodes than the evidence nodes from crime 1. Examples of the conditional probability tables can be found in Table 9.5, 9.6 and 9.7. Note that the constraint nodes are always set to the **constraint** state.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

Figure 9.9: Network for Situation 2: Possible to distinguish between offenders, evidence can be linked to each other.

| male offender | 0.5 |
| female offender | 0.5 |

Table 9.5: Conditional probability table for **donor 1&2** node, Situation 2 (see Figure 9.9). **donor 3&4**, **donor 5&6** and **donor 7&8** have similar structures

<table>
<thead>
<tr>
<th>donor 1&amp;2</th>
<th>male offender</th>
<th>female offender</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>donor 3&amp;4</strong></td>
<td>male offender</td>
<td>female offender</td>
</tr>
<tr>
<td>constraint</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>complement</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 9.6: Conditional probability table for **constraint 1** node, Situation 2 (see Figure 9.9). The **constraint 2** node has a similar structure.
9.3. Modelling crime linkage with multiple offenders

<table>
<thead>
<tr>
<th>donor 1&amp;2</th>
<th>male offender</th>
<th>female offender</th>
</tr>
</thead>
<tbody>
<tr>
<td>offender 1</td>
<td>susp X</td>
<td>unknown</td>
</tr>
<tr>
<td>offender 2</td>
<td>susp Y</td>
<td>unknown</td>
</tr>
<tr>
<td>arch pattern</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>no arch pattern</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 9.7: Conditional probability table for fingermark 1 node, Situation 2 (see Figure 9.9). The other evidence nodes have a similar structure.

**Situation 3, distinguish between offenders, no link between evidence**

The structure of the network for this situation is very similar to the network of Situation 2 (Figure 9.9), and it can be found in Figure 9.10. Every piece of evidence now has at least three parents. Either of the offenders can be the donor of a piece of evidence (two parents) and one node determines which of these offenders is the donor (one parent). Evidence nodes of the same type as other evidence nodes (for example, shoe mark evidence nodes), can have more than three parents. Conditional probability tables are similar to the ones presented in Table 9.4 and 9.7.

**Situation 4, cannot distinguish between offenders, evidence can be linked to each other**

For this situation, the network structure is exactly the same as the one from Situation 2, see Figure 9.9. The only difference is the set of states in the offender configuration node, see Table 9.13.

**Situation 5, cannot distinguish between offenders, no link between evidence**

For this situation, the network structure is exactly the same as the one from Situation 3, see Figure 9.10. The only difference is the set of states in the offender configuration node, see Table 9.13.

**Situation 6, unknown number of offenders**

We draw the Bayesian network corresponding to this situation in Figure 9.11. To account for the uncertainty regarding the number of offenders, two nodes are added to the network: number of offenders crime 1 and number of offenders crime 2. Both these nodes have two states, one and two.
9. Evaluating evidence in crime linkage scenarios with multiple offenders

9.4 Computations and results

9.4.1 Posterior probabilities for different situations

To examine the influence of the assumptions underlying the different situations (Is it possible to distinguish between the offenders? Is it possible to group evidence?) we compare the posterior probabilities that follow from the various networks.

For all situations, we assume that the set of possible offenders consists of 1000 people. In the situation where it is possible to distinguish between offenders, i.e. a man and a woman, it is assumed that the potential offender population consists of 500 men and 500 women. Furthermore, it is assumed that prior to inserting the evidence, every individual is equally likely to be the offender in the various crimes. Note that these assumptions, together with the random match probabilities, should be regarded in the light of the illustrative purpose of the example.

We now proceed as follows. We condition on all the evidence (see Table 9.1) and compute the conditional probabilities of the states of the yellow nodes. These

![Network for Situation 3](image-url)

**Figure 9.10:** Network for Situation 3: not possible to distinguish between offenders, evidence cannot be linked to each other.
conditional probabilities are our posterior probabilities. An overview of these posterior probabilities can be found in Table 9.8. From Table 9.8 we see that

<table>
<thead>
<tr>
<th>situation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>graphical representation, Figure</td>
<td>9.1</td>
<td>9.2</td>
<td>9.3</td>
<td>9.4</td>
<td>9.5</td>
</tr>
<tr>
<td>Bayesian network, Figure</td>
<td>9.8</td>
<td>9.9</td>
<td>9.10</td>
<td>9.9</td>
<td>9.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>posterior probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>suspect X off in crime 1</td>
</tr>
<tr>
<td>suspect Y off in crime 1</td>
</tr>
<tr>
<td>suspect X off in crime 2</td>
</tr>
<tr>
<td>suspect Y off in crime 2</td>
</tr>
</tbody>
</table>

**Table 9.8**: Posterior probabilities after inserting the matching partial DNA profile evidence

for this example, the posterior probabilities that our suspects actually are (one of) the offenders differ substantially among the considered situations. Especially
suspect $Y$ has a very broad range in posterior probabilities (0.017 - 0.765 for crime 1, 0.159-0.710 for crime 2). This is mostly due to the special circumstances of this example. Indeed, all the evidence that is associated with suspect $Y$ in Situation 1 (see Figure 9.1) also matches suspect $X$. In other words, if it is uncertain what evidence was left by which offender, the majority of the evidence previously linked with suspect $Y$ will be associated with suspect $X$. By comparing Situation 1 with Situation 2 and 3 or comparing Situation 4 with Situation 5, the influence of uncertainty regarding the distribution of the evidence over the offenders can be examined. Comparing Situation 2 with Situation 4 and Situation 3 with Situation 5 corresponds to examining the influence of the assumption of being able to distinguish between the offenders.

Posterior probabilities for Situation 6 with different numbers of assumed offenders per crime can be found in Table 9.9. For illustrative purposes, equal prior probabilities are assumed for the states of the number of offenders crime 1 and number of offenders crime 2. We would like to warn the reader that these prior probabilities can be very influential for the final conclusion. Nonetheless, introducing this uncertainty in a simplified model to examine its potential influence can be quite informative. When this influence is substantial, one should have substantial background information that implicates this assumption.

<table>
<thead>
<tr>
<th>number of offenders crime 1/</th>
<th>1/1</th>
<th>1/2</th>
<th>2/1</th>
<th>2/2</th>
<th>unknown/</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of offenders crime 2</td>
<td>unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suspect $X$ offender in crime 1</td>
<td>0.499</td>
<td>0.985</td>
<td>0.496</td>
<td>0.960</td>
<td>0.953</td>
</tr>
<tr>
<td>suspect $Y$ offender in crime 1</td>
<td>0</td>
<td>0</td>
<td>0.012</td>
<td>0.017</td>
<td>0.002</td>
</tr>
<tr>
<td>suspect $X$ offender in crime 2</td>
<td>0</td>
<td>0.777</td>
<td>0</td>
<td>0.775</td>
<td>0.730</td>
</tr>
<tr>
<td>suspect $Y$ offender in crime 2</td>
<td>0</td>
<td>0.250</td>
<td>0</td>
<td>0.251</td>
<td>0.235</td>
</tr>
</tbody>
</table>

Table 9.9: Posterior probabilities for different settings regarding the number of offenders in each crime.

From Table 9.9 we conclude that making assumptions on the number of offenders can influence the resulting posterior probabilities substantially. The posterior probability that suspect $X$ is an offender in crime 1 is almost two times higher when assuming that there are two offenders in crime 2 than when it is assumed that there is only one offender in crime 2. In other words, the posterior probability that a suspect is an offender in a crime can be substantially different given the number of assumed offenders in another crime. This has to do with the fact that suspect $X$ cannot be the only offender in crime 2 since his DNA profile does not match the partial DNA profile, see Table 9.2. Hence, if there is only one offender in crime 2, this cannot be suspect $X$. This ensures that there is another unknown person responsible for crime 2. This unknown apparently has red hair, a similar fingermark and a shoe size 11. Hence, this unknown person ‘matches’ with all the evidence from crime 1. This ‘explains away’ this evidence, resulting in a situation that is less incriminating for suspect $X$. It is remarkable, and perhaps counter-
intuitive, that an assumption on the number of offenders in a crime influences the posterior probability of a suspect being an offender in another crime. Bayesian networks can be used to find and understand this effect in this or other crime linkage problems.

9.4.2 Same offender evidence

It is possible that there is (typically non-physical) evidence that does not link to either of the suspects but that does influence our belief that there is a common offender (or a common offender couple) in two crimes. For example, one can think of a similar modus operandi or a situation in which the crimes were burglaries of two side-by-side houses. This type of evidence can be incorporated in the network by adding a node same offender evidence which is a child of the same offenders node.

![Diagram of Bayesian network with nodes same offenders evidence, same offenders, and offender configuration]

**Figure 9.12:** The position of the node same offenders evidence in the networks.

The same offenders evidence node has two states, observed and not observed. The literature as presented in Section 9.2 gives some examples of behavioural features and statistical methods that can aid in assigning these probabilities. Needless to say however, in actual casework it will be close to impossible to assign meaningful probabilities here, but we can still be interested in the theoretical influence of different probability assignments on the overall conclusions. For instance, we can decide that the same offenders evidence node has value observed with probability $p$, whenever the same offenders node tells us that the two crimes have at least one common criminal, and has this value with probability $1 - p$ otherwise.

We examine the influence of the parameter $p$ on the posterior probabilities that suspect $X$ and suspect $Y$ are offenders in crime 1 and 2, by conditioning not only on the evidence, but also on the node same offenders evidence to take the value observed. Figure 9.13 shows the influence of the parameter $p$ for Situation 5 (see Figure 9.5).

Some observations can now be made. First of all, if the evidence regarding same offenders becomes stronger, the posterior probability that suspect $X$ is an offender...
Figure 9.13: Influence of the parameter \( p \) on posterior probabilities of interest, Situation 5. The number of offenders in both crimes is assumed to be two. It is not possible to distinguish between offenders and the evidence cannot be grouped.

...
9.5 Discussion

We have shown that, in comparison with single offender crime linkage networks (see [4]), there are different scenarios for crime linkage with multiple offenders. Each situation requires a different Bayesian network. If the evidence consists of several pieces of evidence, of which a certain selection matches with one suspect and the remaining evidence with the other suspect, it is not necessarily true that in the alternative scenario in which two unknowns were the offenders the evidence is ‘grouped’ in the same way. An unfounded association between a cell type result and a donor has been labeled an ‘association fallacy’ by Gill [8]. With individual pieces of evidence, a similar fallacy occurs easily: assuming that different pieces of evidence originate from the same individual simply because they match with a known individual can lead to an overestimation of the evidence, i.e. an ‘evidence association fallacy’.

The number of possible hypotheses, even in situations with only a few crimes, is already so large that it is rather challenging to get an understanding about the value of the evidence without the aid of a network. On the other hand, it is unlikely, and in our opinion even undesirable that networks of the type presented in this paper are presented in court as an analysis of a complete case [22], since this will most likely result in a very large network in which the probabilities used in any node can and will be subject of discussion (see [14, 10]). Sensitivity analyses regarding the validity and the implication of these probabilities can also be performed.

However, networks such as the ones presented in this paper can be of assistance to both legal representatives and forensic scientists in identifying and studying various likely alternative explanations, or to identify the most important nodes and probabilities for the overall conclusions. Besides that, these networks help identifying pitfalls in reasoning and assist in preventing them.

We also examined the influence of the number of assumed offenders in both crimes. In casework, often the true number of offenders is unknown. It is still possible to construct a Bayesian network in such a case, but it is necessary to set an upper bound on the number of offenders per case. We have constructed networks in which there could be either one or two offenders per case. We observed some interesting phenomena. For example, the number of offenders that one assumes to be present in one case can substantially influence the posterior probabilities in another case. This is a surprising finding that was discovered by using the networks. The network also explains why this effect occurs.

9.5.1 Alternative distribution of evidence

The mock case example we presented gave us some insight in phenomena that we believe most people will find counter-intuitive. However, one should keep in mind that conclusions are very sensible to case specific details. We illustrate this with a case in which the suspects have slightly different characteristics - see Table 9.10 - compared to the mock case example of Table 9.2. The posterior probabilities in this particular situation are given in Table 9.11.

The assumption that some of the pieces of evidence can be grouped rules out
9. Evaluating evidence in crime linkage scenarios with multiple offenders

<table>
<thead>
<tr>
<th>evidence</th>
<th>match with suspect X</th>
<th>match with suspect Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>partial DNA profile</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>partial fingerprint (arch pattern)</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>red hair</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>shoe mark, size 11</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 9.10: Overview of the evidence related characteristics of the suspects, alternative example.

<table>
<thead>
<tr>
<th>situation</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>graphical representation</em>, Figure</td>
<td>9.1</td>
<td>9.2</td>
<td>9.3</td>
<td>9.4</td>
<td>9.5</td>
</tr>
<tr>
<td><em>Bayesian network</em>, Figure</td>
<td>9.8</td>
<td>9.9</td>
<td>9.10</td>
<td>9.9</td>
<td>9.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>posterior probability</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>suspect X off in crime 1</td>
<td>0.893</td>
<td>0.813</td>
<td>0.209</td>
<td>0.814</td>
<td>0.175</td>
</tr>
<tr>
<td>suspect Y off in crime 1</td>
<td>0</td>
<td>0</td>
<td>0.015</td>
<td>0</td>
<td>0.008</td>
</tr>
<tr>
<td>suspect X off in crime 2</td>
<td>0</td>
<td>0</td>
<td>0.211</td>
<td>0</td>
<td>0.151</td>
</tr>
<tr>
<td>suspect Y off in crime 2</td>
<td>0</td>
<td>0</td>
<td>0.141</td>
<td>0</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Table 9.11: Posterior probabilities after inserting the matching partial DNA profile evidence, alternative example.

The possibility that the suspects are offenders of the crimes in many situations. Furthermore, it is remarkable that removing the assumption that one can group the evidence as being left by the same offender substantially decreases the probability that suspect X is the offender in crime 1 (Situation 2 versus Situation 3 and Situation 4 versus Situation 5). In the previous example, dropping this assumption increased the probability that suspect X was an offender in both crimes.

This research shows that evidential reasoning in crime linkage with multiple offenders is complex and very subtle. There are seemingly unimportant assumptions that can have a substantial influence on the probabilities regarding the culpability of suspects. Bayesian networks have proved to be a useful tool in exploring and understanding these factors.

Bibliography


9. Evaluating evidence in crime linkage scenarios with multiple offenders


9. A Offender configurations

<table>
<thead>
<tr>
<th>index</th>
<th>offender 1</th>
<th>offender 2</th>
<th>crime 1</th>
<th>crime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>suspect X</td>
<td>suspect Y</td>
<td>suspect X</td>
<td>suspect Y</td>
</tr>
<tr>
<td>2</td>
<td>suspect X</td>
<td>suspect Y</td>
<td>suspect X</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>3</td>
<td>suspect X</td>
<td>suspect Y</td>
<td>unk. male 1</td>
<td>suspect Y</td>
</tr>
<tr>
<td>4</td>
<td>suspect X</td>
<td>suspect Y</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>5</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>suspect Y</td>
</tr>
<tr>
<td>6</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>7</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>unk. female 2</td>
</tr>
<tr>
<td>8</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>suspect Y</td>
</tr>
<tr>
<td>9</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>10</td>
<td>suspect X</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>unk. female 2</td>
</tr>
<tr>
<td>11</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>suspect X</td>
<td>suspect Y</td>
</tr>
<tr>
<td>12</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>suspect X</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>13</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>unk. male 1</td>
<td>suspect Y</td>
</tr>
<tr>
<td>14</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>15</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>unk. male 2</td>
<td>suspect Y</td>
</tr>
<tr>
<td>16</td>
<td>unk. male 1</td>
<td>suspect Y</td>
<td>unk. male 2</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>17</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>suspect Y</td>
</tr>
<tr>
<td>18</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>19</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>suspect X</td>
<td>unk. female 2</td>
</tr>
<tr>
<td>20</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>suspect Y</td>
</tr>
<tr>
<td>21</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>22</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 1</td>
<td>unk. female 2</td>
</tr>
<tr>
<td>23</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 2</td>
<td>suspect Y</td>
</tr>
<tr>
<td>24</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 2</td>
<td>unk. female 1</td>
</tr>
<tr>
<td>25</td>
<td>unk. male 1</td>
<td>unk. female 1</td>
<td>unk. male 2</td>
<td>unk. female 2</td>
</tr>
</tbody>
</table>

**Table 9.12:** Offender configurations for a situation with 2 crimes, 2 offenders of which one is a man and the other a woman, and 2 suspects, X (male suspect) and Y (female suspect).
Table 9.13: Offender configurations for a situation with 2 crimes, 2 offenders that cannot be distinguished and 2 suspects, suspect X and suspect Y.
Summary

In this thesis I consider the evaluation of a combination of different pieces of evidence in a legal and a forensic context. Evaluation of forensic evidence is the main topic of a research area called forensic statistics. In forensic statistics, the likelihood ratio framework is regarded as the standard for evaluating evidence. In legal practice, it is common that two competing propositions are presented to the trier of fact. The trier of fact needs to establish whether the proposition presented by the prosecution can be proven to the extent that there could be no ‘reasonable doubt’ in the mind of a ‘reasonable person’ that the defendant is guilty. The presented evidence should rule out any reasonable doubt. The likelihood ratio framework is based on probabilistic inference by applying Bayes’ Theorem. It allows for the transition from prior (initial) belief regarding the presented propositions to posterior (final) beliefs. This transition is based on the conditional probabilities to observe the evidence given the propositions (the likelihood ratio).

In forensic casework, it is common that multiple pieces of evidence that need to be evaluated in terms of their support regarding the presented propositions are available. The most straightforward way of doing this for a forensic expert is by presenting a separate likelihood ratio for each individual piece of evidence. However, when doing so, one needs to be confident that the individual reports are optimally combined by the trier of fact. When forensic experts believe that their knowledge regarding the dependency structure between pieces of evidence is lost by presenting the likelihood ratios separately, one should strive to combine this evidence before it is sent to the trier of fact. Especially in situations where the pieces of evidence are of the same type (e.g. two shoe marks), one usually cannot regard them as conditionally independent observations and a combined evaluation is needed to prevent unnecessarily independent misconceptions.

When considering how different pieces of evidence influence each other and the probabilities of the propositions of interest, it is necessary to consider several different situations. Pieces of evidence relevant for the same question (e.g. who was present at the crime scene?) are easier to combine in terms of evaluation of the evidence than when they are relevant for different questions (e.g. who was present at the crime scene? and what happened at the crime scene?). Furthermore, when one piece of evidence is only relevant for the distinguishing between propositions given another piece of evidence, yet another method is needed to evaluate the evidence combined. In this thesis, methods for combining evidence are presented for these different situations, using common scenarios from forensic practice. As DNA is currently one of the most important types of forensic evidence, we focus on this area for practical examples. However, we want to emphasize that the application of our methods to other areas (even outside forensic science) is rather straightforward.

Apart from different situations, the dependency structure between pieces of evi-
Summary

dence is important when combining evidence. Various types of dependencies are examined for several types of evidence. Recommendations on how to combine the evidence are presented. Thus, we explore a wide range of problem types starting from combining two kinds of DNA analyses of a single sample and ending at combining several types and pieces of evidence linking multiple cases with multiple offenders.

For common situations in forensic practice, Bayesian networks have been developed in the forensic statistical literature as a tool to apply the likelihood ratio framework. A Bayesian network is a graphical representation of the dependency structure between a set of random variables. Random variables are represented with nodes, conditional dependencies with directed edges between nodes. For discrete variables, each node consists of a set of mutually exclusive states which represent the set of possible values of the random variable. Complex situations from forensic practice can be modelled with the aid of a Bayesian networks. With the aid of software, which basically repeatedly applies Bayes’ theorem, one is still able to examine the impact of observations on other random variables for these complex problems.

In this thesis, Bayesian networks have been developed to study a broad selection of forensic topics. One Bayesian network considers the situation in which a crime stain is obtained from a piece of tape. In these situations, it is common that the suspect is not disputed to be the source of the crime stain. The activity that resulted in the crime stain is disputed: was it transferred during the crime or was it transferred in an innocent way? By reconstructing the order of the pieces of tape to the roll of tape, one can infer the most likely layer of the piece of tape on which the crime stain was found. The crime stain becomes much more incriminating when the crime stain is on a piece of tape that was on an ‘inner’ layer. For a stain on the ‘outer’ layer, one cannot disregard the possibility that a suspect only held the tape prior to the crime. The Bayesian network identifies which data should be collected in order to properly evaluate such a situation.

Another situation in which the discriminative value of a DNA profile obtained from a crime stain depends on other evidence is when the type of cell material is crucial for evaluating what happened. In sexual offense cases, evidence supporting that the crime stain consists of semen cells will result in a stronger belief that the suspect is the offender than when it is likely that the crime stain consists of skin cells. In this thesis, probabilistic models for both traditional methods for cell type determination and RNA profiles are developed. Existing methods for the interpretation of RNA profiles as evidence for the presence of certain cell types are compared with these probabilistic methods. It is recognized that these probabilistic methods are an important improvement of the existing methods. Besides a better performance, they are flexible and can be adapted to other situations. For example, they can assist in the combination of RNA with DNA evidence.

The combination of DNA evidence with cell type evidence is evaluated using a Bayesian network for traditional methods for cell type determination. A litera-
ture overview is presented that summarizes the specificity and the sensitivity of three common traditional cell type tests. The developed Bayesian network combines the results from the traditional cell type tests with a software package for the evaluation of DNA profiles obtained from crime stains with multiple donors. The Bayesian network identifies the added value of performing additional specificity/sensitivity tests. Furthermore, it is shown that the common assumption that an individual (e.g. the victim) is one of the donors in a mixed DNA profile can have undesirable consequences for the association between donors and cell types.

In sexual offense cases, it regularly happens that the autosomal DNA profile obtained from e.g. a vaginal swab is not very incriminating regarding a suspect. The amount of material left by the male offender is often minimal compared to the amount of material left by the female victim. In such a situation, a Y-chromosomal DNA profile of the same crime stain can be informative. A Y-chromosomal DNA profile is a male specific profile only of the Y-chromosome. Because Y-chromosomal DNA profiles are paternally inherited, Y-chromosomal DNA profiles usually bear substantially less discriminative value than autosomal DNA profiles. Hence, it might be necessary to combine the evidential value of the autosomal and the Y-chromosomal DNA profile. In this thesis, the dependency relation between these two DNA profiles is examined, both with a simulation model and a database containing 2085 men. It is concluded that if the Y-chromosomal DNA profiles match, one can still regard the autosomal DNA profile as independent from the Y-chromosomal DNA profile if the matching person is not a descendant of the father of the donor of the (crime) sample. The combined evidential value can, in that case, be computed by multiplying the random match probabilities of the individual profiles.

The last three chapters consider situations in which a person is suspected of being the offender in multiple crimes. Firstly, crime linkage is modelled with Bayesian networks. A general structure is introduced that allows the interpretation of evidence in simplified crime linkage situations, in which it is assumed that each crime only had one offender. This simplified model allows for the examination of the validity of intuitive reasoning. Most importantly, it is highlighted that crime linkage is a double edged sword, it may be used to prove guilt but also innocence. An important practical consequence of this is that in a series of crimes of which it is believed that they have a common offender one cannot simply drop one of the crimes because there is exculpatory evidence regarding the suspect. This can result in a substantial overestimation of the evidence in the remaining cases.

Lastly, how mathematical models can aid legal practitioners in their reasoning is explored. ‘Schakelbewijs’ is the term used within Dutch legal practice when several cases of which it is believed that they have been committed by the same offender are linked to reach a verdict, even in cases in which the evidence would be insufficient when regarding the case individually. The underlying idea is that, when there is an ‘anchor’ case, i.e. ‘sufficient’ evidence to reach a verdict in the first crime, and the modus operandi of that crime is very similar to a second
Summary

crime, the prior belief that the same suspect is the offender in this second crime is stronger. Hence, less evidence is needed to reach a verdict in this second crime. The logic behind such a ‘schakelbewijs’ methodology is approached from both a Dutch legal and a probabilistic point of view. These two are compared, leading to insights on how probabilistic reasoning can aid in schakelbewijs casework. It is shown that the idea of schakelbewijs is logical, and in fact some legal requirements are too strict from a purely probabilistic point of view: an ‘anchor’ case is not necessary and the link between the cases may be other types of evidence than a similar modus operandi.

The last chapter extends the Bayesian network approach for crime linkage with one offender to situations with multiple offenders. Several situations are distinguished for multiple offender crime linkage problems. For a simplified example, it is recognized that the assumed situation can have a substantial influence on the posterior probabilities of interest. Due to the complexity of such multiple offender crime linkage problems, probabilistic models in which the dependency structure of observations can be specified are valuable reasoning tools, for legal and forensic experts.
Samenvatting

Dit proefschrift behandelt de evaluatie van de combinatie van verschillende stukken bewijsmateriaal in een juridische en forensische context. De evaluatie van forensisch bewijsmateriaal is het belangrijkste thema van het onderzoeksgebied forensische statistiek. In dit gebied is het likelihood ratio raamwerk de standaard voor het evalueren van bewijsmateriaal. In een juridische context is het gebruikelijk dat twee proposities worden gepresenteerd aan de rechter. Deze moet bepalen hoeveel twijfel er bestaat over de proppositie gepresenteerd door het openbaar ministerie. Het gepresenteerde bewijs wordt gebruikt om twijfel te verkleinen of te vergroten. Het likelihood raamwerk is gebaseerd op de stelling van Bayes. De stelling van Bayes geeft de mogelijkheid a-priori kansen om te zetten in a-posteriori kansen op basis van geobserveerde bewijsstukken. Deze omzetting is gebaseerd op de conditionele kansen op het observeren van het bewijs, gegeven de gepresenteerde proposities (de likelihood ratio).

In forensisch zaakwerk komt het vaak voor dat meerdere bewijsstukken geëvalueerd dienen te worden om hun bewijswaarde voor de gepresenteerde proposities te onderzoeken. De makkelijkste manier om dit te doen voor een forensisch expert is om voor elk bewijsstuk apart een likelihood ratio te geven. Echter, als men dit doet moet men er wel van overtuigd zijn dat de individuele bewijsstukken optimaal gecombineerd worden door de rechter. Als de forensische expert denkt dat zijn kennis over de afhankelijkheidsstructuur tussen de bewijsstukken verloren gaat door de likelihood ratios van de bewijsstukken apart te rapporteren is het gewenst om de gecombineerde bewijswaarde te rapporteren. Specifiek voor gevallen waarin de bewijsstukken van hetzelfde type zijn (bijvoorbeeld twee schoensporen) is het normaal gesproken niet het geval dat de bewijsstukken conditioneel onafhankelijk zijn en is een gecombineerde evaluatie nodig om onnodige misverstanden te voorkomen.

Om te bepalen hoe de verschillende bewijsstukken elkaar en de kansen behorende bij de gepresenteerde proposities beïnvloeden, is het noodzakelijk om verschillende scenarios te beschouwen. Bewijsstukken die relevant zijn voor hetzelfde vraagstuk (bijvoorbeeld wie was er op het plaats delict) zijn gemakkelijker te combineren, in termen van de evaluatie van het bewijs, dan bewijsstukken die relevant zijn voor verschillende vragen (bijvoorbeeld wie was er op het plaats delict en wat heeft er plaatsgevonden op het plaats delict). Wanneer een bewijsstuk alleen relevant is voor het onderscheid tussen twee proposities gegeven een ander bewijsstuk, is weer een andere methode noodzakelijk om de gezamenlijke bewijswaarde van deze stukken te bepalen. Dit proefschrift presenteert methoden voor het combineren van bewijsstukken in deze verschillende situaties aan de hand van veelvoorkomende type zaken. Aangezien DNA op dit moment een van de belangrijkste typen forensisch bewijs is ligt hierop de focus in de praktische voorbeelden. Echter, ik wil benadrukken dat de gepresenteerde methoden relatief eenvoudig zijn toe te passen op andere gebieden (ook buiten de forensische wetenschap).

Afgezien van deze verschillende situaties is ook de afhankelijkheidsstructuur tussen...
bewijsstukken belangrijk wanneer men bewijs combineert. Verschillende typen afhankelijkheden worden onderzocht voor verschillende typen bewijsmateriaal. Aanbevelingen voor het combineren van dit bewijs worden gegeven. Hierbij bestuderen we een brede set van problemen beginnende met het combineren van twee soorten DNA-profielen verkregen van hetzelfde spoor tot het combineren van verschillende soorten en stukken bewijs wanneer men zaken linkt met meerdere daders.

Voor veelvoorkomende situaties in de forensische praktijk zijn Bayesianse netwerken ontwikkeld in de forensische literatuur om het likelihood ratio raamwerk toe te kunnen passen. Een Bayesianse netwerk is een grafische representatie van de afhankelijkheidsstructuur tussen een set van random variabelen. Deze random variabelen worden als knopen weergegeven en de afhankelijkheidsrelatie daartussen door pijlen tussen de knopen. Voor discrete variabelen bestaan de knopen uit een set van elkaar uitsluitende toestanden. Deze toestanden komen overeen met de set van mogelijke waarden behorende bij de random variabele. Complex situaties uit de forensische praktijk kunnen met Bayesianse modellen worden gemanageerd. Met behulp van software pakketten, die herhaaldelijk de stelling van Bayes toepassen, kan de impact van observaties op random variabelen onderzocht worden voor deze complexe problemen.

In dit proefschrift worden Bayesianse netwerken bestudeerd die zijn ontwikkeld voor de evaluatie voor een selectie aan typen forensische zaken. Eén van deze netwerken behandelt de situatie waarin een biologisch spoor op een stuk tape wordt gevonden. In dit soort zaken is het gebruikelijk dat de verdachte als donor van het spoor niet wordt betwist. De activiteit die leidde tot het spoor wordt wel betwist. Heeft, bijvoorbeeld, het contact tussen verdachte en tape plaatsgevonden tijdens de misdaad of reeds daarvoor? Forensisch specialisten kunnen de verkregen stukken tape reconstrueren en nagaan hoe ze oorspronkelijk aan de rol vastzaten. Door dit te doen is het mogelijk te bepalen in welke ‘tapelaag’ het spoor zich op de rol bevond. Het spoor is vaak meer belastend wanneer het van een ‘binnenste’ tapelaag komt. Als de verdachte de rol heeft aangeraakt voordat de misdaad werd begaan is het waarschijnlijker dat het spoor op een stuk tape zit dat oorspronkelijk behoorde tot de buitenste laag van de rol tape. Het Bayesianse netwerk assisteert in het identificeren van welke informatie verzameld dient te worden om dit soort zaken correct te evalueren.

Een andere situatie waarin de bewijswaarde van een gevonden DNA-profiel afhankelijk is van andere gevonden bewijzen, is wanneer het celttype bepalend is voor de evaluatie van wat er is gebeurd. In zedenzaken is het bijvoorbeeld veelal aanneemelijker dat er een misdaad is begaan wanneer er spermacellen zijn gevonden dan wanneer er alleen huidcellen zijn aangetroffen. In dit proefschrift worden kansmodellen voor zowel traditionele methodes voor de bepaling van celtypes, alswel voor RNA-profielen ontwikkeld. Bestaande methoden voor de interpretatie van RNA-profielen als bewijsmateriaal voor de aanwezigheid voor bepaalde celtypes worden vergeleken met deze kansmodellen. We concluderen dat deze methoden een belangrijke verbetering zijn van de bestaande methoden. Naast een beter
performance zijn ze ook flexibeler en kunnen worden aangepast aan andere situaties. Ze kunnen bijvoorbeeld worden ingezet voor de combinatie van RNA- en DNA-bewijsmateriaal.

De combinatie van DNA-bewijs en celtype bewijs wordt geëvalueerd met een Bayesiaans netwerk voor traditionele methoden voor de bepaling van celtypen. In een literatuuroverzicht wordt een samenvatting van specificiteit en sensitiviteit van drie traditionele celtype tests gepresenteerd. Het ontwikkelde Bayesiaanse netwerk combineert de resultaten van deze traditionele celtype tests met een software pakket voor de evaluatie van gemengde DNA-profielen die op een crime scene worden gevonden. Het Bayesiaanse netwerk identificeert de toegevoegde waarde van extra specificiteits-/sensitiviteitstests. Ook wordt aangegeven dat de gebruikelijke aanname dat een individu (bijvoorbeeld het slachtoffer) één van de donors in een gemengd DNA-profiel is, ongewenste consequenties kan hebben voor de associatie tussen donors en celtypen.

In zedenzaken is het niet ongebruikelijk dat het autosomale DNA-profiel verkregen uit bijvoorbeeld een vaginaal uitstrijkje een lage bewijswaarde heeft. De hoeveelheid materiaal afkomstig van de mannelijke dader is vaak minimaal vergeleken met de hoeveelheid materiaal behorende bij het vrouwelijke slachtoffer. Wanneer dit het geval is wordt vaak een Y-chromosomaal DNA-profiel van het spoor gemaakt. Een Y-chromosomaal profiel is een profiel alleen van het Y-chromosoom. Omdat alleen mannen een Y-chromosoom hebben is het maken van een Y-chromosomaal DNA-profiel alleen mogelijk voor mannen. Het Y-chromosoom wordt van vader op zoon doorgegeven. Derhalve is een Y-chromosomaal DNA-profiel minder onderscheidend dan een autosomaal DNA-profiel en is het wenselijk om de bewijswaarde van het Y-chromosomaal DNA-profiel te combineren met de bewijswaarde van het autosomale DNA-profiel. In dit proefschrift wordt de afhankelijkheidsrelatie tussen deze twee DNA-profielen onderzocht middels een simulatiemodel en een DNA-databank bestaande uit DNA-profielen van 2085 mannen. Er wordt geconcludeerd dat wanneer er matchende Y-profielen worden geobserveerd, men de autosomale profielen nog steeds als onafhankelijk kan beschouwen van de Y-chromosomale DNA-profielen wanneer de verdachte geen afstammeling van de vader van de donor van het zaakspoor is. In zulke gevallen kan de gecombineerde bewijswaarde worden bepaald door de random match-kansen van de autosomale en Y-chromosomale DNA-profielen met elkaar te vermenigvuldigen.

In de laatste drie hoofdstukken worden situaties beschouwd waarin een verdachte wordt verdacht van het plegen van meerdere vergrijpen. Het linken van zaken wordt gemodelleerd met behulp van een Bayesiaans netwerk. Een algemene structuur wordt geïntroduceerd voor een vereenvoudigde situatie waarin wordt aangenomen dat elk vergrijp werd gepleegd door één dader. Dit versimpelde model geeft de mogelijkheid de validiteit van een intuïtieve redenatie te testen. Middels dit model wordt bijvoorbeeld gevonden dat het linken van zaken een mes is dat snijdt aan twee kanten. Het kan gebruikt worden om zowel schuld als onschuld te bewijzen. Een belangrijk gevolg hiervan is dat het niet mogelijk is om een zaak
weg te nemen uit een set geschakelde zaken omdat er bewijs in die zaak bestaat
dat de onschuld van een verdachte aantoont. Doet men dit wel dan kan dit leiden
tot een substantiële overschatting van het bewijs in de overige zaken.

In de Nederlandse juridische praktijk wordt een schakelbewijsconstructie gebruikt
om meerdere zaken waarvan men denkt dat ze dezelfde dader hebben gezamenlijk
te behandelen. Deze bewijsconstructie kan gebruikt worden om tot een veroorde-
ling te komen in zaken waarbij het bewijs, wanneer de zaak individueel behandeld
zou worden, onvoldoende zou zijn geweest om tot een veroordeling te komen. Het
achterliggende idee is dat wanneer er voldoende bewijs is om tot een veroordeling
te komen in de eerste zaak, een zogenaamd ‘anker’, en een tweede zaak een over-
eenkomstige modus operandi heeft, er voor het tweede bewijs reeds een sterker
a-priory geloof is dat de verdachte de dader is. Daardoor is er in de tweede zaak
minder zaak-specifieke bewijs nodig om tot een veroordeling te komen. De logica
achter deze schakelbewijsconstructie wordt in dit proefschrift benaderd vanuit het
gezichtspunt van het Nederlandse rechtssysteem en de kansrekening. Deze wor-
den met elkaar vergeleken wat leidt tot inzichten over hoe de kansrekening kan
assisteren in het redeneren in schakelbewijszaken. We laten zien dat de achterlig-
gende gedachte bij schakelbewijs klopt. Echter, vanuit de kansrekening gezien zijn
sommige vereisten voor het gebruik van schakelbewijs te strikt. Een ‘anker’ is bij-
voorbeeld niet noodzakelijk en ook ander bewijs dan een overeenkomstige modus
operandi kan gebruikt worden om zaken te schakelen.

Het laatste hoofdstuk houdt zich bezig met uitbreiding van het Bayesiaanse net-
werk voor het schakelen van zaken met één dader naar situaties met meerdere
daders. Voor het schakelen van zaken met meerdere daders worden meerdere on-
derscheidbare situaties beschouwd. Middels een voorbeeld wordt aangetoond dat
de veronderstelde situatie een substantiële invloed kan hebben op de a-posteriori
kansen. Vanwege de complexiteit behorende bij geschakelde zaken met meerdere
daders hebben kansmodellen die dit beschrijven een toegevoegde waarde als hulp-
middel voor de waarschijnlijkheidsredenatie, voor zowel juristen als forensische
experts.
Overview of author contributions

Chapter 1: Introduction

J. de Zoete (PhD Candidate) Authored text of this chapter.
M. Sjerps Reviewed the final manuscript version.
R. Meester Reviewed the final manuscript version.

Chapter 2: The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample

This chapter was originally published as: J. de Zoete, M. Sjerps, R. Meester and E. Cator. The combined evidential value of autosomal and Y-chromosomal DNA profiles obtained from the same sample. International Journal of Legal Medicine, Volume 128, Number 6, pp 897–904, 2014.

J. de Zoete (PhD Candidate) Wrote original MATLAB™ source code. Drafted full text of original manuscript. Performed computational work that generated all results presented in the manuscript. Performed revisions to satisfy reviewer concerns
M. Sjerps Discussion and development of the concept and supervision of the research.
R. Meester Provided editorial input in pre-submission manuscript.
E. Cator Discussion and development of the concept and supervision of the research.
P. de Knijff Provided database consisting of 2085 men to investigate the dependence between autosomal and Y-chromosomal DNA profiles.

Chapter 3: The interpretation of traces found on adhesive tapes

This chapter was originally published as: R. Wieten, J. de Zoete, B. Blankers and B. Kokshoorn. The interpretation of traces found on adhesive tapes. Law, probability and risk, (14), pp 305 – 322, 2015.

R. Wieten Co-authored text of this chapter. Constructed Bayesian Network.
J. de Zoete (PhD Candidate) Co-authored text of this chapter. Discussion and development of the concept and supervision of the research. Performed revisions to satisfy reviewer concerns.
B. Blankers Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.
B. Kokshoorn Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.
M. Sjerps Discussion and development of the concept.
Overview of author contributions

Chapter 4: Categorical methods for the interpretation of RNA profiles as cell type evidence and their limitations

This chapter was originally published as: J. de Zoete, J. Curran and M. Sjerps. *Categorical methods for the interpretation of RNA profiles as cell type evidence and their limitations.* Forensic Science International: Genetics Supplement Series, (5), e305-e307, 2015.

**J. de Zoete (PhD Candidate)** Drafted full text of original manuscript. Performed literature review of all papers cited in Section 2. Performed revisions to satisfy reviewer concerns.

**J. Curran** Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

**M. Sjerps** Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

**P. Maaskant** Discussion and development of the concept.

Chapter 5: A probabilistic approach for the interpretation of RNA profiles as cell type evidence

This was chapter originally published as: J. de Zoete, J. Curran and M. Sjerps. *A probabilistic approach for the interpretation of RNA profiles as cell type evidence.* Forensic Science International: Genetics, (20), pp 30 – 44, 2016.

**J. de Zoete (PhD Candidate)** Drafted full text of original manuscript. Performed literature review of all papers cited in Section 2. Wrote original MATLAB™ source code. Performed revisions to satisfy reviewer concerns.

**J. Curran** Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

**M. Sjerps** Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

**A. Roeder and C. Haas** Provided a database consisting of RNA profiles from samples of various cell types for training and testing the probabilistic methods.

**A. Lindenbergh, P. Maaskant and T. Sijen** Provided a database consisting of RNA profiles from samples of various cell types for training and testing the probabilistic methods.

Chapter 6: Cell type determination and association with the DNA donor

This was chapter originally published as: J. de Zoete, W. Oosterman, B. Kokshoorn and M. Sjerps. *Cell type determination and association with the DNA donor.* Forensic Science International: Genetics, [http://dx.doi.org/10.1016/j.fsigen.2016.08.004](http://dx.doi.org/10.1016/j.fsigen.2016.08.004), 2016.
Overview of author contributions

J. de Zoete (PhD Candidate) Drafted full text of original manuscript. Wrote original R source code. Performed Sensitivity analysis. Co-constructed the Bayesian network. Performed revisions to satisfy reviewer concerns. Discussion and development of the concept.

W. Oosterman Performed and wrote literature review of all papers cited in Section 2. Co-constructed the Bayesian network. Discussion and development of the concept.

B. Kokshoorn Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

M. Sjerps Provided editorial input in pre-submission manuscript.

Chapter 7: Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten

This chapter was originally published as Jacob de Zoete, Koen Vriend, Menno Dolman, Ronald Meester en Marjan Sjerps. Het gebruik van schakelbewijs; juridische en kanstheoretische gezichtspunten. Expertise en Recht. 2014-5.

J. de Zoete (PhD Candidate) Authored the probabilistic parts of the text of this chapter. Performed revisions to satisfy reviewer concerns

K. Vriend Co-authored the legal practice sections of this chapter. Performed revisions to satisfy reviewer concerns

M. Dolman Co-authored the legal practice sections of this chapter. Performed revisions to satisfy reviewer concerns

R. Meester Discussion and development of the concept and supervision of the research.

M. Sjerps Discussion and development of the concept and supervision of the research.

Chapter 8: Modelling crime linkage with Bayesian networks


J. de Zoete (PhD Candidate) Drafted full text of original manuscript. Constructed Bayesian Networks. Performed computational work that generated all results presented in the manuscript. Performed revisions to satisfy reviewer concerns.

M. Sjerps Discussion and development of the concept and supervision of the research.

D. Lagnado Discussion and development of the concept and supervision of the research.

N. Fenton Provided editorial input in pre-submission manuscript.
Overview of author contributions

Chapter 9: Evaluating evidence in crime linkage scenarios with multiple offenders

At the time of this dissertation’s submission to the Doctorate Committee, the work in this chapter is under review as a manuscript submitted to the Elsevier journal: Science & Justice.

J. de Zoete (PhD Candidate) Drafted full text of original manuscript. Constructed Bayesian Networks. Performed computational work that generated all results presented in the manuscript. Performed literature review of all papers cited in Section 2.

M. Sjerps Discussion and development of the concept and supervision of the research. Provided editorial input in pre-submission manuscript.

R. Meester Discussion and development of the concept. Provided editorial input in pre-submission manuscript.

D. Lagnado Provided suggestion for research regarding situations with an unknown number of offenders. Discussion based on the most important messages of a very early draft of the manuscript.

N. Fenton Discussion based on the most important messages of a very early draft of the manuscript.
Acknowledgements

Marjan, jij was voor mij de perfecte begeleider. In alle beslissingen die mijn werk aangingen was jij altijd in de eerste plaats bezig met wat het beste was voor mij. Dankzij jouw inspanningen heb ik onder andere drie keer een aantal maanden een onderzoeksgroep in het buitenland kunnen bezoeken. Deze bezoeken reken ik tot de meest verrijkende periodes in mijn leven. Ik ben trots op wat we samen hebben klaargespeeld en hoop in de toekomst nog geregeld samen te kunnen werken. Ronald, ik heb veel geleerd van jouw kijk op het vakgebied. Je wees me dikwijls op onduidelijkheden in mijn manuscript en hebt meerdere keren bewezen dat je een meester bent in het bondiger verwoorden van de boodschap die ik probeer over te brengen. Naast mijn dagelijkse begeleiders ben ik Eric, Bert, Koen en Menno dank verschuldigd voor hun inzichten en de prettige samenwerking.

Ik prijs me gelukkig met de verschillende werkplekken die ik heb gehad de afgelopen jaren. De combinatie universiteit en forensisch instituut heeft me veel interessante onderzoeksvragen opgeleverd. De WISK afdeling op het NFI heeft er de afgelopen vijf jaar voor gezorgd dat ik met veel plezier naar mijn werk ging. Annabel, Arian, Ate, Charles, Didier, Ivo, Jan, Peter en Reinoud, dank voor de gezelligheid en de discussies. Sharda en Yvonne, met een gastrolletje voor Sanne, het was goud om samen een kamer te delen. Bas, Bart, Petra, Klaas en Jeannette, jullie waren de hoogtepunten buiten de WISK afdeling. Ik heb genoten van de samenwerking en de gezelligheid, binnen en vooral ook buiten het NFI.

Op de Vrije Universiteit heb ik de afgelopen jaren mijn UvA rethoriek mogen uitleven tegenover Ivan, Maarten, Marcin en Tim. Van alle kantoren waar ik heb gewerkt tijdens mijn PhD was het dinsdagkantoor op de VU de plek waar werk en waanzin de perfecte mix vormden. Maarten, ik hoop in de toekomst nog op veel congressen met jou te kunnen lego. Abhishek, Alice, Apo, Arie, Asma, Bart, Bart, Dirk, Gang, Guusje, Hanung, Jacob, Jan, Kayed, Kirsten, Lance, Marcelo, Marijn, Mariska, Masha, Masoumeh, Nikos, Petr, Reinier, Sasha, Sonja, Suzanne, Wendy and Zhou, you made the institute such an enjoyable place. Hopefully, I will be able to disturb your research at least one more time. Botond, Brendan, Danilo, David, Jarno, Julia, Moritz, Nicos and Peter, you are the worst, no wait, you are the best, ever. Martin, Marjan’s other PhD student, you have an amazing drive to make the most out of everything. Thank you for the ‘excellent’ times. Wessel and Remi, it was a pleasure working together. I am proud of what we achieved within your projects.

Throughout my doctoral studies I have been lucky enough to carry out several research visits to foreign institutes. Dave, my visit to your group in London is an indisputable highlight in my life. I have thoroughly enjoyed working and playing football together and spending time with your group. Adam, Christos, James, Nadine, Neil you made it all so much better. Norman, thanks for the good times, the collaboration and the kind invitation to Cambridge. Alessia, Rens and Sofia, living together with you was the most amazing feature of my time in London. We had the best house and the most awesome of times.
Acknowledgements

James, you are exemplary for me, inside and outside the research field. You are incredibly kind and a wonderful person to be around. I hope that one day, I will be able to join you and your group in Auckland for (at least) a couple of years. Furthermore, I recommend every PhD student to share a water slide experience with their supervisor. Alastair, Chris and Thomas, thanks for the travel suggestions, the Friday drinks and all the lunches. Claudia, David, Jackie and Patricio, thank you for all the non-work related activities. Jeroen, we had the best adventures and you have the superior beard. Andy en Celine, thank you for welcoming me to your place.

Mikkel and Torben, you hosted my third and last research visit during my PhD. As advocates of the R-programming language, you re-inspired me to work in this language. More importantly, you represent the best the field has to offer. Thanks for the good times in Denmark, and more recently, in Cambridge. Apart from the two of you, I should thank Ege, Diego, Nasos, Søren, Svante and Umberto for all the activities and discussions in Aalborg.

Teun en Mark, vanaf het moment dat mijn universiteitsleven begon zijn jullie er altijd geweest. Jullie staan synoniem voor goedheid en zijn een sprekend voorbeeld van synergie. De jongens van De Grasnaaiers, dank voor de afleiding en de voorzetten op maandagavond. De Oopz & Knak Filmclub evenzo voor de verrijking op het gebied van Japanse klassiekers. Rune, Frans, Roel en Daan dank voor de activiteiten.

Chris, Frank, James, Jeroen, Joost, Maarten en Paul, alles wat ik met jullie doe is fantastisch. Ik ben trots op wat we zijn en wat we doen. Al meer dan tien jaar maken jullie mijn leven rijker en mooier. Ik kijk uit naar alle komende momenten vol eindeloze liedjesavonden, discussies met onverdedigbare standpunten, zwarte randjes en ongrijpbare gezelligheid.

Chris, you have the tidiest and most orderly brain, with the greatest capacity for storing facts, of any man living. Jarno, you are the organiser of half that is evil and of nearly all that is undetected in our great faculty. Both of you have proven, time and time again, that you can be quiet and look seemingly unaffected while I am defending disputable statements. Hence, I feel very confident and fortunate having you by my side.


Lieve Charlotte, jij bent de liefste en belangrijkste van allemaal. Ik zie u graag.