UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2013

Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence

Simone N. Gittelson

Simone N. Gittelson 2013 Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence

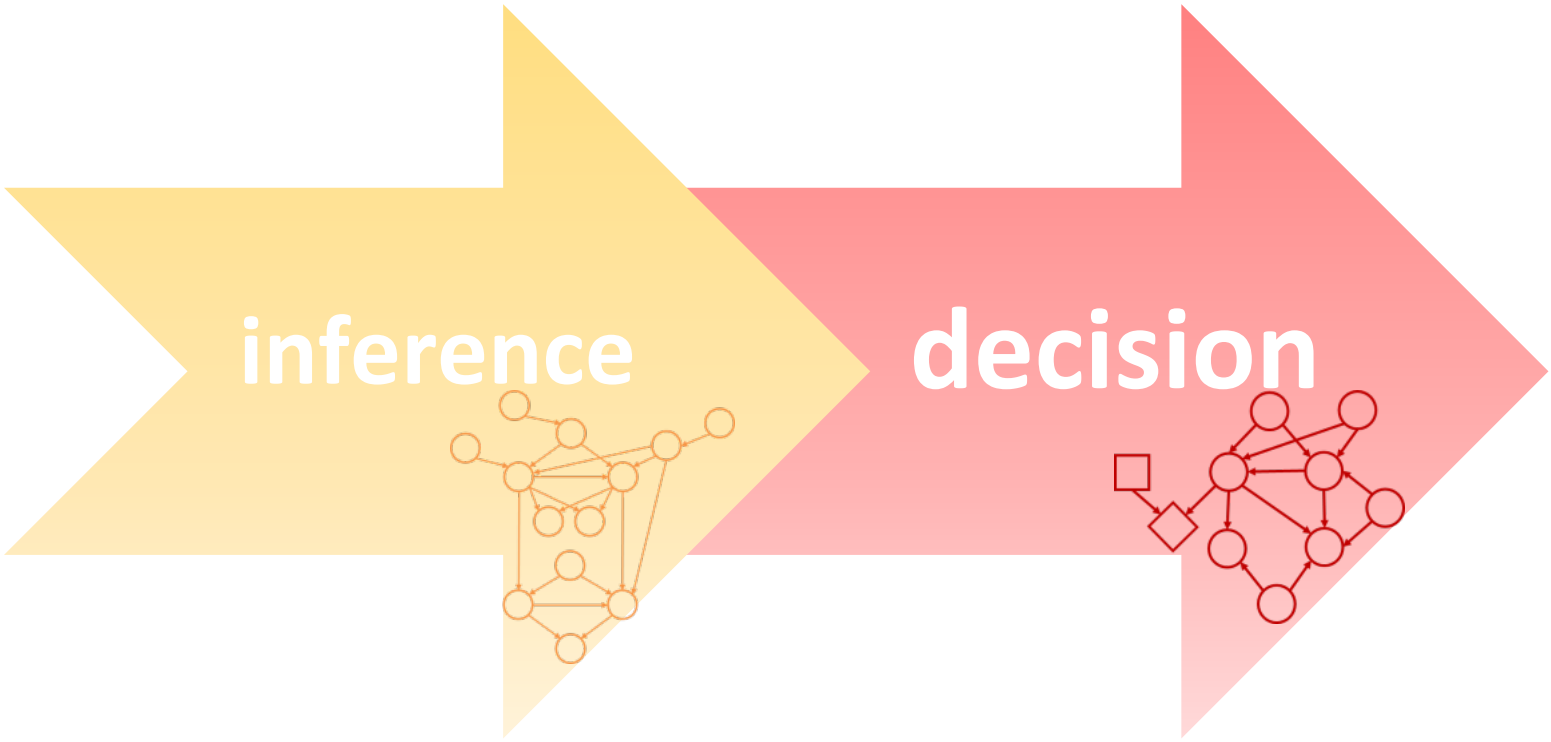
Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive.
http://serval.unil.ch

Droits d'auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence



Thèse de Doctorat

présentée à l'Institut de Police Scientifique
de l'Université de Lausanne

par

Simone N. Gittelson
Master en sciences forensiques, mention identification,
de l'Université de Lausanne

Lausanne 2013

UNIL | Université de Lausanne
Université de Lausanne
Faculté de Droit et des Sciences Criminelles
Ecole des Sciences Criminelles
Institut de Police Scientifique

Série criminalistique LVI
ISBN 2-940098-60-3

UNIL | Université de Lausanne
Ecole des sciences criminelles
bâtiment Batochime
CH-1015 Lausanne

IMPRIMATUR

A l'issue de la soutenance de thèse, le Jury autorise l'impression de la thèse de Mme Simone Gittelson, candidate au doctorat en science forensique, intitulée

« Evolving from Inferences to Decisions in the Interpretation of Scientific Evidence »

Le Président du Jury

Professeur Pierre Margot

Lausanne, le 26 août 2013

A Franco Taroni qui m'a transmis la passion pour le domaine de l'interprétation des indices en science forensique

Acknowledgements

This research was made possible by the support of the Swiss National Science Foundation grants n° 100014-122601 and n° 100014-135340. It was performed at the Institute of Forensic Science of the School of Criminal Justice at the University of Lausanne, under the supervision of Prof. Franco Taroni, Professor in Forensic Statistics in the Faculty of Law and Criminal Justice of the University of Lausanne. The dissertation committee consisted of:

Prof. Pierre Margot (President),
Professor of Forensic Science at the University of Lausanne, and Director of the University of Lausanne's School of Criminal Justice;

Dr. Alex Biedermann,
Lecturer in Forensic Statistics at the University of Lausanne;

Prof. Silvia Bozza,
Assistant Professor in Statistics at the University Ca'Foscari of Venice, Italy;

Prof. Anders Nordgaard,
Reader in Statistics at Linköping University, Sweden, and Senior Statistician at the Swedish National Laboratory of Forensic Science; and

Prof. Niels Morling,
Professor of Forensic Genetics at the University of Copenhagen, Denmark, and the University of Tromsø, Norway, and Director and Chairman of the Department of Forensic Medicine at the University of Copenhagen, Denmark.

Funding for international visits and presentations of this research at international conferences was provided by the Swiss National Science Foundation (as part of the grants listed above), the Institute of Forensic Science's Alumni Association, the Swiss Study Foundation, and the University of Lausanne's Foundation of the 450th anniversary.

In addition to expressing my gratitude to the above institutions for their financial support, I would like to give a special thanks to everyone who has supported and encouraged me in writing this dissertation during these past four and a half years:

- Above all, Prof. Franco Taroni for giving me the opportunity to pursue my dream of writing this dissertation in forensic statistics; for being an outstanding supervisor through the most priceless discussions, encouragement, support and feedback that a Ph.D. student can ask for; and most of all, for pushing me to go beyond what I thought were my limits. The topic of this dissertation was Prof. Franco Taroni's idea, and I therefore dedicate this dissertation to him.
- Prof. Pierre Margot for giving me the opportunity to begin my career as a researcher within the walls of Lausanne's Institute of Forensic Science, where I enjoyed the invaluable freedom of being able to think and write whenever I wanted to.
- Dr. Alex Biedermann, Prof. Silvia Bozza, Prof. Anders Nordgaard and Prof. Niels Morling for their constructive feedback, critical comments and fruitful discussions that resulted from meticulously reading my work, and which have made me become a better researcher. I would also like to thank Prof. Silvia Bozza for being such a kind host in Venice.

- Dr. Tacha Hicks Champod, Prof. Cedric Neumann, Prof. Christophe Champod, Dr. Ian Evett, Prof. Colin Aitken, Amy Wilson, Giulia Cereda, Patrick Juchli, Emanuele Sironi, Matteo Gallidabino, Alex Thiery, Dr. Gunnar Andersson, Dr. Roberto Puch-Solis and Dr. Sue Pope for more insightful discussions in forensic statistics and for their friendship. Also:
  - Prof. Colin Aitken and Amy Wilson for their generous hospitality in Edinburgh.
  - Dr. Ian Evett for his excellent presentations that always motivated me in pursuing my research.
- Dr. Raphaël Coquoz, Dr. Vincent Castella, Christian Gehrig, Carmelo Simili and Dr. Diana Hall for valuable insight in forensic genetics.
- Anne Marville for her invaluable administrative organization and assistance.
- Daniel and Willy for their stories and for making me laugh.
- All my other colleagues at the School of Criminal Justice, who have motivated me in my work. In particular: Martin, Juan, Aurèle, Durdica, Valérie, Flore, Isabelle, Beatrice, Joëlle, Agathe, Sébastien and Denise.
- And SEFE for the chocolates.

J'aimerais encore particulièrement remercier:

l'Ecole et Troupe de Danse UTOPIA, Francesca et Gilles, pour ne pas m'avoir laissé oublier la beauté de la danse;

et ma famille pour son soutien inconditionnel.

Abstract

Forensic science casework involves making a series of choices. The difficulty in making these choices lies in the inevitable presence of uncertainty, the unique context of circumstances surrounding each decision and, in some cases, the complexity due to numerous, interrelated random variables. Given that these decisions can lead to serious consequences in the administration of justice, forensic decision making should be supported by a robust framework that makes inferences under uncertainty and decisions based on these inferences. The objective of this thesis is to respond to this need by presenting a framework for making rational choices in decision problems encountered by scientists in forensic science laboratories.

Bayesian inference and decision theory meets the requirements for such a framework. To attain its objective, this thesis consists of three propositions, advocating the use of (1) decision theory, (2) Bayesian networks, and (3) influence diagrams for handling forensic inference and decision problems.

The results present a uniform and coherent framework for making inferences and decisions in forensic science using the above theoretical concepts. They describe how to organize each type of problem by breaking it down into its different elements, and how to find the most rational course of action by distinguishing between one-stage and two-stage decision problems and applying the principle of expected utility maximization.

To illustrate the framework's application to the problems encountered by scientists in forensic science laboratories, theoretical case studies apply decision theory, Bayesian networks and influence diagrams to a selection of different types of inference and decision problems dealing with different categories of trace evidence. Two studies of the two-trace problem illustrate how the construction of Bayesian networks can handle complex inference problems, and thus overcome the hurdle of complexity that can be present in decision problems. Three studies—one on what to conclude when a database search provides exactly one hit, one on what genotype to search for in a database based on the observations made on DNA typing results, and one on whether to submit a fingermark to the process of comparing it with prints of its potential sources—explain the application of decision theory and influence diagrams to each of these decisions. The results of the theoretical case studies support the thesis's three propositions.

Hence, this thesis presents a uniform framework for organizing and finding the most rational course of action in decision problems encountered by scientists in forensic science laboratories. The proposed framework is an interactive and exploratory tool for better understanding a decision problem so that this understanding may lead to better informed choices.

Résumé

Le travail d'un(e) expert(e) en science forensique exige que ce dernier (cette dernière) prenne une série de décisions. Ces décisions sont difficiles parce qu'elles doivent être prises dans l'inévitable présence d'incertitude, dans le contexte unique des circonstances qui entourent la décision, et, parfois, parce qu'elles sont complexes suite à de nombreuse variables aléatoires et dépendantes les unes des autres. Étant donné que ces décisions peuvent aboutir à des conséquences sérieuses dans l'administration de la justice, la prise de décisions en science forensique devrait être soutenue par un cadre robuste qui fait des inférences en présence d'incertitudes et des décisions sur la base de ces inférences. L'objectif de cette thèse est de répondre à ce besoin en présentant un cadre théorique pour faire des choix rationnels dans des problèmes de décisions rencontrés par les experts dans un laboratoire de science forensique.

L'inférence et la théorie de la décision bayésienne satisfont les conditions nécessaires pour un tel cadre théorique. Pour atteindre son objectif, cette thèse consiste de trois propositions, recommandant l'utilisation (1) de la théorie de la décision, (2) des réseaux bayésiens, et (3) des réseaux bayésiens de décision pour gérer des problèmes d'inférence et de décision forensiques.

Les résultats présentent un cadre uniforme et cohérent pour faire des inférences et des décisions en science forensique qui utilise les concepts théoriques ci-dessus. Ils décrivent comment organiser chaque type de problème en le décomposant dans ses différents éléments, et comment trouver le meilleur plan d'action en faisant la distinction entre des problèmes de décision en une étape et des problèmes de décision en deux étapes et en y appliquant le principe de la maximisation de l'utilité espérée.

Pour illustrer l'application de ce cadre à des problèmes rencontrés par les experts dans un laboratoire de science forensique, des études de cas théoriques appliquent la théorie de la décision, les réseaux bayésiens et les réseaux bayésiens de décision à une séléction de différents types de problèmes d'inférence et de décision impliquant différentes catégories de traces. Deux études du problème des deux traces illustrent comment la construction de réseaux bayésiens permet de gérer des problèmes d'inférence complexes, et ainsi surmonter l'obstacle de la complexité qui peut être présent dans des problèmes de décision. Trois études—une sur ce qu'il faut conclure d'une recherche dans une banque de données qui fournit exactement une correspondance, une sur quel génotype il faut rechercher dans une banque de données sur la base des observations faites sur des résultats de profilage d'ADN, et une sur s'il faut soumettre une trace digitale à un processus qui compare la trace avec des empreintes de sources potentielles—expliquent l'application de la théorie de la décision et des réseaux bayésiens de décision à chacune de ces décisions. Les résultats des études des cas théoriques soutiennent les trois propositions avancées dans cette thèse.

Ainsi, cette thése présente un cadre uniforme pour organiser et trouver le plan d'action le plus rationnel dans des problèmes de décisions rencontrés par les experts dans un laboratoire de science forensique. Le cadre proposé est un outil interactif et exploratoire qui permet de mieux comprendre un probléme de décision afin que cette compréhension puisse aboutir à des choix qui sont mieux informés.

Contents

Chapter 1

Introduction

1.1 General context

A forensic scientist contributes to judicial decision making by presenting the value of scientific observations and analytical results to the court. Through laboratory analyses and comparisons of trace evidence, she[1] seeks to establish associations or disassociations,[2] determine the components of an unknown substance,[3] or evaluate the observations and results in view of different possible activities.[4] Her goal in performing each of these tasks is to help the court form an opinion on the disputed facts.

However, there is one major challenge in her work: forensic science casework inevitably involves uncertainty. This uncertainty is reflected in questions such as:

- What traces did the perpetrator leave behind on a crime scene, victim or item collected by the investigators? Or, what traces did a crime scene or victim leave on the perpetrator?
- What activity caused the transfer of these traces?
- If the traces are latent, where are they located?
- If an analytical method is applied, is the result error-free?

To deal with this uncertainty, the forensic scientist makes inferences based on her knowledge, past experience, observations and the results of any analytical methods she may have applied.

Yet, once she has made these inferences, what does she do with them? Formally, inferences take the form of a probability distribution over a random variable of interest. They give a piece of information. They do not tell the forensic scientist,

- what traces to recover, collect, analyze and compare with reference material;
- which examination techniques and analytical methods to apply;
- what conclusion to draw based on the observations and results of each of the applied methods; or

[1] Or he. For better readability of this thesis, we shall consider a female forensic scientist, and further on, a female factfinder and decision-maker, and a male perpetrator and suspect. There is no sex discrimination whatsoever intended by this choice. It will simply make this thesis easier to read.

[2] According to Edmond Locard's famous **exchange principle**, the author of a crime inevitably leaves behind traces indicating his presence on the crime scene, and, at the same time, takes away material from the scene, also indicating his presence on the crime scene (Locard, 1940).

[3] For example, whether this substance contains an illegal drug, or whether fire residues contain traces of an ignitable liquid.

[4] For example, from the glass fragments on a suspect's shirt, whether this suspect smashed a window or whether he was a witness standing at least 2 meters away, or from semen collected on a victim's body, whether the agressor had sexual intercourse with the victim.

- what profile to search for in a database to try to find a potential source.

Making inferences is the first step in a process that continues on to use the information from the inference to make better informed decisions (Biedermann, 2007; Taroni et al., 2010).

The interpretation of scientific evidence is therefore not only about making inferences: the inferences made by forensic scientists on the presence, nature, relevance, quality, quantity or origin of a trace form the basis for making and justifying the subsequent course of action. Hence:

> [t]hough logically independent, inference and decision are connected because the results of the former are the point of departure of the latter. (Taroni et al., 2012)

Decision-making is a new area of research in forensic science. It is emerging in a wide range of forensic fields, covering decisions investigators make on a crime scene, decisions scientists make in a forensic science laboratory, and decisions based on forensic intelligence. This thesis focuses on decisions that a forensic scientist must make while processing traces in a forensic science laboratory.[5] These decisions range from whether the laboratory should process a particular trace to what conclusions can be drawn from the obtained results.

Several research studies have explored how forensic scientists make such decisions (e.g., Dror et al., 2005, 2006; Dror and Charlton, 2006; Dror and Rosenthal, 2008; Hall and Player, 2008; Langenburg et al., 2009; Charlton et al., 2010; Helsloot and Groenendaal, 2011; Dror et al., 2011; Dror and Hampikian, 2011; Ulery et al., 2012). No doubt incited by notorious errors made by forensic scientists (e.g., the Mayfield[6] case, the American Innocence Project[7]) and the general trend of increasingly questioning their decisions (e.g., National Research Council (NRC) of the National Academies, 2009), these studies have focused on what factors influence forensic scientists' decision-making process. Most of them came to the conclusion that forensic scientists can make inconsistent decisions, in the sense that a same forensic scientist faced with the same data can make different, or even contradicting choices, when the decision is made at two different points in time and/or with different, usually biasing, contextual information.[8] This conclusion is particularly troublesome when one considers the influential, or even biasing (Kassin et al., 2013), effect that forensic conclusions can have on trial outcomes. According to these studies, there is apparently a great need for forensic scientists to make more consistent choices.

At this point, let us make the distinction between two different approaches to studying decision making. What the above cited studies did, was to describe how people make decisions. This is known as the **descriptive** approach to decision making. Another approach, called a **normative** approach, explores what the most rational course of action would be based on logic and within a strict mathematical framework (Lindley, 1985). So, a descriptive approach describes how people act, and a normative approach provides a standard on how people ought to act—this standard should result in better decisions if implemented (Lindley, 2006). The question is, should we study decision making using a descriptive or a normative approach? As Edwards (1991 - 1992) puts it,

> [s]hould we simply accept the fact that unaided human intuition does not conform to formal rules of thought? Or should we provide the thinker with aids based on

[5] For further information on research concerning the other types of decisions, see for example Hazard et al. (2011) for an investigator's decisions on a crime scene, and Ribaux and Margot (2003) for how to use forensic intelligence in view of making decisions.

[6] The FBI erroneously identified Brandon Mayfield as the source of a fingermark recovered on the crime scene of the Madrid bombings in 2004. See, for example, *http://edition.cnn.com/2006/LAW/01/06/mayfield.report/* (last visited on 31.05.2013).

[7] By performing DNA analyses, the Innocence Project has exonerated over 300 wrongfully convicted individuals (*http://www.innocenceproject.org/*, last visited on 31.05.2013). In 2009, more than half of the exonerated individuals had been wrongfully convicted because of unvalidated or improper forensic science.

[8] It is worth noting that most of these inconsistent decisions were made for difficult decision problems, that is, decisions concerning low quality traces. Yet, low quality traces are a reality in forensic science, and are presented in court as evidence, so that these decisions are of the utmost importance because they may determine whether the true perpetrator is found and convicted of the crime.

> those formal rules and instruction about how to use both rules and aids, and thus help the thinker to avoid cognitive illusions? (...) As I understand it, legal inference and decision is an inherently normative enterprise. Its goal is to promote clarity and cogency of thought and decision about real and practical problems, *not* to enshrine the error-inducing impacts of ignorance. (Edwards, 1991 - 1992, p.1056)

Since we want forensic scientists to make consistent choices, this thesis adopts an entirely normative approach. A normative approach provides the tools to aid the decision-maker in organizing her thoughts for making consistent decisions.

Sofar only a handful of studies describe a normative approach to a selection of isolated forensic decision problems (Taroni et al., 2005, 2006b, 2007, 2010; Shen et al., 2006; Biedermann et al., 2008a, 2012a; Mazumder, 2010). Research in forensic decision making using a normative approach is therefore at its very beginnings.

Globally, there are two types of decision problems between which we will distinguish in this thesis: **one-stage decision problems** and **two-stage decision problems**. A one-stage decision problem is a decision for which the decision-maker can directly quantify the satisfaction obtained from this decision's outcome. For example, if the decision-maker has a choice of choosing between three doors, behind which, one hides a prize and the others hide nothing, choosing the one with the prize will make her happy, and choosing one of the other two will make her feel disappointed. The outcome of her choice of door directly leads to her happiness or disappointement. A two-stage decision problem, on the other hand, is a decision for which the decision-maker's satisfaction does not directly depend on this decision's outcome, but depends on the outcome of a subsequent decision. For example, if before choosing a door, our decision-maker has the choice of spending 10$ to obtain a clue that may provide information on which door hides the prize, then her satisfaction from this decision does not depend on whether she bought the clue or not, but again on whether the door she ultimately chooses is the one with the prize. For this decision problem, the question of interest is whether the information provided by the clue is worth the 10$ it costs to obtain it. This thesis will examine both one-stage and two-stage decision problems in forensic science.

1.2 Objective of this thesis

The objective of this thesis is to show how to confront some decision problems encountered by scientists in forensic science laboratories. The focus of this thesis is on how a forensic scientist should make these decisions so that her choices are rational, and not on the description of how a forensic scientist currently makes them. The idea is that once we have a model for making rational choices, it can help forensic scientists make more consistent decisions by providing a structure for organizing the different elements that should have an impact on the scientist's choice.

The range of decisions made in a forensic science laboratory is very broad. They range from whether to perform a particular laboratory analysis on a trace or item collected on a crime scene, to what conclusion to draw from the obtained results. In addition, different laboratory methodologies apply to each different category of trace evidence (e.g., fingermarks, biological traces, gunshot residues), so that each category of trace evidence has its own particular decision problems. Comparing the current publications in forensic science that present a normative approach to some of these individual decision problems amongst each other shows a lack of uniformity on how to address these problems. The aim of this thesis is therefore not to address each of these decisions separately, but to provide a uniform framework that applies to all of them.

Every decision can be categorized as either a one-stage or a two-stage decision problem. This thesis will therefore distinguish between these two types of decision problems, and show how to confront each type in forensic science. For this, three forensic science questions have been chosen to illustrate the theoretical concepts and their applications in forensic science:

- one-stage decision problems:
  - What genotype to search for in a database based on the observations made on electropherograms (EPGs) obtained from DNA analyses of low-template DNA traces?
  - What to conclude when a database search provides exactly one hit?
- two-stage decision problem:
  - Submit a fingermark to the process of comparing it with prints of its potential sources?

The thesis consists of three propositions on how to handle forensic inference and decision problems, notably:

Proposition 1: *Decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence.*

Proposition 2: *Bayesian networks (BNs) provide a means for handling new complex inference problems encountered by forensic scientists.*

Proposition 3: *Influence diagrams (IDs) provide forensic scientists with a practical tool for structuring and providing the elements required for making coherent choices in forensic decision problems.*

Separate studies, in the form of research articles, will try to support each of these propositions.

1.3 Outline

This thesis has three parts: the first considers the theoretical aspects of inference and decision methodologies, the second the practical aspects of implementing these methodologies in forensic science, and the third consists of papers applying these methodologies in case studies.

PART I: THEORETICAL ASPECTS summarizes the background knowledge forming the basis for the practical aspects in Part II. This part contains four chapters:

Chapter 2: Justification for a Bayesian and Decision-theoretic Framework explains the concept of a normative framework and the fundamental notions of a coherent approach to judicial decision making. These notions cover the principles of interpretation of scientific evidence and introduce the concept of decision theory. The mathematical notation in this chapter is limited to probabilities, leaving the mathematical notation of decision theory for Chapter 3.

Chapter 3: Decision Theory presents the fundamental notions of statistical decision theory.

Chapter 4: Graphical Models presents graphical probabilistic models, notably Bayesian networks and influence diagrams.

Chapter 5: Current State of the Art of Decision-theoretic Models in Forensic Science gives an overview and summary of the scientific literature on normative approaches to forensic decision problems.

PART II: PRACTICAL ASPECTS explains the contributions of this thesis. It contains three chapters:

Chapter 6: Thesis presents the three propositions of interest, advocating the use of decision theory, Bayesian networks and influence diagrams for handling forensic inference and decision problems. The objective of the practical aspects of this thesis is to study these propositions.

Chapter 7: Results and Discussion summarizes the results of the studies for the propositions presented in Chapter 6 by explaining how to confront one-stage and two-stage decision problems in forensic science. The separate studies referred to in this chapter figure in Part III.

Chapter 8: Conclusions summarizes the dissertation's contributions and describes the implications of the results in a larger context of forensic science and of its role in judicial decision making.

PART III: PUBLISHED PAPERS contains the case studies applying decision theory, Bayesian networks and influence diagrams to different forensic inference and decision problems. Its five chapters are in the form of separate research papers, published or in the process of being published, in peer-reviewed journals:

Chapter 9: Modeling the forensic two-trace problem with Bayesian networks presents a generic Bayesian network for a forensic inference problem known as the two-trace problem.

Chapter 10: Bayesian networks and the value of the evidence for the forensic two-trace transfer problem explains how to contruct Bayesian networks for more complex inference problems of a two-trace scenario.

Chapter 11: The database search problem: A question of rational decision making presents the application of decision theory and influence diagrams to the decision of individualizing a person as the trace's source when this person is found as a single hit in a database search.

Chapter 12: Decision analysis for the genotype designation in *low-template-DNA* profiles presents the application of decision theory and influence diagrams to the decision of choosing the genotype to search for in a database of DNA profiles from potential sources.

Chapter 13: Decision-theoretic reflections on processing a fingermark presents the application of decision theory and influence diagrams to the decision of submitting a fingermark to the comparison process with prints of its potential sources.

Part I

Theoretical Aspects

Chapter 2

Justification for a Bayesian and Decision-theoretic Framework

2.1 Decisions in forensic science

The decisions can lead to serious consequences in the administration of justice, such as determining whether the true perpetrator is found and convicted of the crime, or whether an innocent person is acquitted.[1] These decisions are therefore of the utmost importance and must be made based on coherent foundations that can be used to justify the resulting choices in a court of law.

Yet, the decisions that forensic scientists must make inevitably come with the following challenges:

- they must be made in the presence of *uncertainty*;
- each is made in a *unique context*, formed by the circumstances of the particular case;
- they can be *complex* due to numerous pieces of information (say, variables) that are not necessarily independent of each other.

To handle these challenges, forensic decision making should be supported by a robust framework that makes:

(i) inferences under uncertainty, and

(ii) decisions based on these inferences.

As discussed in Section 1, only a normative approach makes sense for such a framework if we want it to help forensic scientists make more consistent, or as Lindley (1985) calls them, coherent decisions.

2.2 A normative framework for ensuring coherence

Due to the inevitable presence of uncertainty when making the decisions,

> it will not be possible to say that a decision is right but only that these decisions cohere, or not. It is the relationships between events or decisions that matter, not the individual events or decisions. (Lindley, 1985, p. 22)

[1] See, for example, footnote 7 on page 2 concerning the American Innocence Project.

A normative framework provides constraints that ensure coherence in decision making. Essentially, these constraints demand that the decision-maker's degrees of belief in uncertain events, as well as her degrees of satisfaction with the choices' possible consequences, obey the laws of probability.[2] It has been shown that making coherent decisions is beneficial to the decision-maker, because it prevents her from obtaining sure losses[3] (Ramsey, 1931; de Finetti, 1980, 1970; Cornfield, 1967; Lindley, 1985, 2006; Parmigiani and Inoue, 2009). By providing a logical foundation, a normative approach therefore offers the potential for improving the decisions made (Lindley, 2006). It gives the forensic scientist a transparent and rational foundation which she can use to organize her thoughts and to justify her choices. Note that the aim of a normative approach is not to replace the forensic scientist as the decision-maker, but rather to provide her with tools, or a standard on how a decision in a particular scenario should be made, that she can use to make coherent decisions.[4]

The next step is to define this normative framework. Statisticians and legal scholars have shown that a **Bayesian**[5] framework provides the most adequate normative model for judicial decision making (e.g., Lindley, 1977a; Kaye, 1979; Fienberg and Schervish, 1986; Robertson and Vignaux, 1993; Redmayne, 2001), and, as a consequence, also for representing the value of scientific evidence (e.g., Finkelstein and Fairley, 1970; Lindley, 1975, 1977b; Evett, 1983; Aitken and Stoney, 1991; Saks and Koehler, 1991; Robertson and Vignaux, 1995a,b; Evett and Weir, 1998; Aitken and Taroni, 2004). A Bayesian framework is adequate for dealing with the uncertainty present in judicial decision making because it provides a coherent way of describing the uncertainty in a particular case, and a means to logically update an individual's degrees of belief in the light of new information (e.g., evidential testimony) (Finkelstein and Fairley, 1970). In addition, it provides a framework not only for making inferences, but also for making decision. The next two sections summarize these ideas.

2.3 Subjective probabilities and Bayes' theorem for making inferences

In a judicial setting, the factfinder (judge or jury) faces two opposing parties, each relating a different story of a past event. The factfinder can never know with certainty what truly

[2]The basic laws of probability are:

1. **Convexity Law:** A probability of an event X, denoted by $Pr(X)$, is a real number $0 \leq Pr(X) \leq 1$. If X is known to be true, $Pr(X) = 1$, and if X is known to be impossible, $Pr(X) = 0$.
2. **Addition Law:** The probability of one of two events occurring, say event X or event Y, is: $Pr(X \text{ or } Y) = Pr(X)+Pr(Y)-Pr(XY)$. If the two events are mutually exclusive, then $Pr(X \text{ or } Y) = Pr(X)+Pr(Y)$. If they are mutually exclusive and exhaustive, then $Pr(X \text{ or } Y) = Pr(X)+Pr(Y) = 1$.
3. **Multiplication Law:** The probability of two events occurring together, say events X and Y, is: $Pr(X \text{ and } Y) = Pr(Y|X) \times Pr(X)$.

[3]Sure losses occur when there is a Dutch book. A Dutch book is a combination of bets whose probabilities do not obey the laws of probability and are therefore incoherent. For example, if events X, Y and Z are mutually exclusive and exhaustive (i.e., not more than one of them can occur, but one of them must occur), then betting odds of 3 to 1 against X (i.e., $Pr(X) = \frac{1}{4}$), 2 to 1 against Y (i.e., $Pr(Y) = \frac{1}{3}$) and evens against Z (i.e., $Pr(Z) = \frac{1}{2}$) constitute a Dutch book because $Pr(X) + Pr(Y) + Pr(Z) = \frac{13}{12}$ is greater than 1 and therefore violates the Addition Law (see footnote 2 in this chapter). A person who bets on each of these events so as to be sure to bet on the winning event, and places the stakes in function of the betting odds so that each should break even in the long run, will always lose a certain amount of money: for example, placing a stake of 3\$ on event X, 4\$ on event Y and 6\$ on event Z means the person pays a total of $3\$+4\$+6\$ = 13\$$, yet, if X occurs, this person wins $4 \times 3\$ = 12\$$, if Y occurs $3 \times 4\$ = 12\$$, and if Z occurs $2 \times 6\$ = 12\$$, producing a sure loss of 1\$ (this numerical example is based on the example presented in Taroni et al. (2001, p. 148)).

[4]When a normative approach is used to provide such a standard for a practical application, literature calls this approach a **prescriptive** approach to decision making (e.g., Lindley, 1985; Kaye, 1986; Lindley, 2006).

[5]The term **Bayesian** describes a method of inference that updates personal beliefs about an event after observing a set of data. This method is named after Reverend Thomas Bayes, who first described this reasoning process in 1763 (Bayes, 1763), establishing what is now known as Bayes' Theorem (Eq. (2.1)).

happened (Kaplan, 1968). The factfinder is uncertain, and her degree of uncertainty is different for each case. Given a case's circumstancial information and the stories presented by the parties, the factfinder or the parties formulate **propositions**[6] regarding the disputed issue. For the evaluations that follow to be coherent, these propositions must be **mutually exclusive**[7]. For example, in a criminal trial where the uncertain event is whether the defendant committed the crime, these propositions are:

proposition 1: *the defendant committed the crime*,

proposition 2: *the defendant did not commit the crime; someone else in the population of potential suspects committed the crime.*

The factfinder cannot state that one of these propositions is certain to be true, or certain to be false, for her.[8] Thus she must have a degree of belief in each one that is somewhere inbetween these two extremes.

In the Bayesian framework, **subjective** (or **personal**) probabilities quantify these degrees of belief (de Finetti, 1980, 1993; Ramsey, 1931; Savage, 1972; Jeffrey, 2004). These probabilities should by no means be understood as *arbitrary*. By definition, they are conditional on an individual's knowledge and experience, and describe a personal relationship between this individual and her outside world (e.g., Lindley, 1978). Their key advantage over a frequentist interpretation of probabilities (i.e., calculating the frequency of an outcome from a sequence of repetitions of similar events), is that subjective probabilities can describe a person's belief in the truth of events that occur only once, for which a strictly frequentist definition seems inadequate. The particular offense that the defendant is charged with is such a one-time event. The uncertainty surrounding it is therefore a subjective, or personal, notion representing a particular person's state of mind. Only subjective probabilities make sense for quantifying this degree of belief (Finkelstein and Fairley, 1970; Lindley, 1977a; Taroni et al., 2001).

A quantitative measure of an individual's subjective probabilities exists by comparing that individual's belief in an event to the betting conditions which the individual chooses for and against that event (de Finetti, 1993, 1970).[9] If these quantitative assessments are coherent, it has been shown that subjective probabilities obey the laws of probability. This justifies their use in a normative framework for making coherent decisions, where obeying the laws of probability is a requirement (see Section 2.2).

According to the definition of subjective probabilities, all probabilities are conditional because they are conditioned on a person's knowledge. Notationwise, $Pr(\cdot|\cdot)$ will denote a conditional probabiltity. This expression designates the conditional probability of the element(s) to the left of the vertical bar, given the element(s) to the right of the vertical bar. For example, if we denote by I a person's knowledge at a given point in time, by Θ_1 proposition 1, and by Θ_2 proposition 2, $Pr(\Theta_1|I)$ and $Pr(\Theta_2|I)$ denote this person's degrees of belief in Θ_1 and Θ_2 at the point in time described by I.

As the factfinder hears evidential testimony, this testimony will modify the factfinder's knowledge by adding new pieces of information concerning the case. As a result, the factfinder's degrees of belief in the propositions may change. The Bayesian framework's principal property is that of providing a logical means to update an individual's initial degrees of belief to take into account a new piece of information. For this, **Bayes' theorem** (a logical consequence of the laws of probability) describes the relationship between the probability of a proposition without the new information and the probability of this proposition with the new information. Mathematically, if we denote a new piece of information here by E (the mathematical notation will be extended and described in more detail in Section 3.2),

[6] A **proposition** is any statement about an event that is either true or false.

[7] Several propositions are **mutually exclusive** if never more than one is true.

[8] According to Cromwell's rule it is unsatisfactory to speak of absolute certainty or impossibility of events if this is not demonstratable by logic (Lindley, 2006).

[9] Another way of ensuring that the probabilities assigned by an individual reflect that individual's true degrees of belief is by using a quadratic scoring rule to penalize the assigned values that deviate from the true beliefs (e.g., de Finetti, 1970). Biedermann et al. (2013) re-explain this idea in a more recent publication.

Bayes' theorem for discrete probability distributions[10] describes the relationship between $Pr(\Theta_1|E,I)$ and $Pr(\Theta_1|I)$ as:

$$Pr(\Theta_1|E,I) = \frac{Pr(E|\Theta_1,I)Pr(\Theta_1|I)}{Pr(E|I)} \ . \tag{2.1}$$

Likewise the relationship between $Pr(\Theta_2|E,I)$ and $Pr(\Theta_2|I)$ is:

$$Pr(\Theta_2|E,I) = \frac{Pr(E|\Theta_2,I)Pr(\Theta_2|I)}{Pr(E|I)} \ . \tag{2.2}$$

This updating process may be repeated as many times as necessary to take into account every new piece of information.

In both of these equations, one may assign $Pr(E|I)$ by using the rule of the **extension of the conversation**[11]:

$$Pr(E|I) = Pr(E|\Theta_1,I)Pr(\Theta_1|I) + Pr(E|\Theta_2,I)Pr(\Theta_2|I) \ . \tag{2.3}$$

Or, one may avoid assigning $Pr(E|I)$ by dividing Eq. (2.1) by Eq. (2.2). This produces the odds form of Bayes' theorem:

$$\underbrace{\frac{Pr(\Theta_1|I)}{Pr(\Theta_2|I)}}_{\textit{prior odds}} \times \underbrace{\frac{Pr(E|\Theta_1,I)}{Pr(E|\Theta_2,I)}}_{\textit{Bayes factor}} = \underbrace{\frac{Pr(\Theta_1|E,I)}{Pr(\Theta_2|E,I)}}_{\textit{posterior odds}} . \tag{2.4}$$

The posterior odds are equal to the product of the prior odds and the **Bayes factor**[12]. Thus, this equation provides a clear distinction between the contribution of the prior odds and the contribution of the new piece of information to the posterior odds of a pair of propositions. Note that in this equation, Θ_1 and Θ_2 may represent either a simple[13] proposition or a set of multiple[14] propositions. When Θ_1 and Θ_2 each represent a simple proposition, the Bayes factor is equal to the **likelihood ratio** (Berger, 2010), that is, the ratio of the probability of E given that Θ_1 is true (i.e., the likelihood of Θ_1 given E) and the probability of E given that Θ_2 is true (i.e., the likelihood of Θ_2 given E), in light of the knowledge contained in I.[15] When either Θ_1 or Θ_2 is a set of multiple propositions, or both are a set of multiple

[10] Bayes' theorem also applies to continuous probability distributions. However, this thesis considers only discrete probability distributions, so the equations for continuous probability distributions have been ommitted to avoid unnecessary mathematical notation and developments.

[11] The **extension of the conversation** (also referred to as the **law of total probability**) allows one to assign the probability of an event by breaking it down into conditional probabilities including another event. These conditional probabilities are sometimes easier to assign than the probability of the event on its own. For example, the probability of X may be assigned by considering its conditional probabilities on the mutually exclusive and exhaustive events $Y_1,\ldots,Y_n$: $Pr(X) = \sum_{i=1}^{n} Pr(X|Y_i)Pr(Y_i)$.

[12] By definition, the **Bayes factor** is equal to the ratio dividing the posterior odds by the prior odds.

[13] Proposition Θ_j, $j = 1,2$, is a simple proposition when the probability of E is the same for all possible scenarios given I and given that Θ_j is true.

[14] Proposition Θ_j, $j = 1,2$, consists of a subset of multiple propositions when the probability of E is different for different possible scenarios given I and given that Θ_j is true. In this case, each of the different possible scenarios is a different proposition, and $Pr(E|\Theta_j,I)$ must be assigned using the rule of the extension of the conversation (Eq. (2.3)) for these different propositions given Θ_j.

[15] Mathematically, the Bayes factor for simple propositions is equal to:

$$\begin{aligned} \textit{Bayes factor} &= \frac{Pr(\Theta_1|E,I)}{Pr(\Theta_2|E,I)} \times \frac{Pr(\Theta_2|I)}{Pr(\Theta_1|I)} \\ &= \frac{Pr(E|\Theta_1,I)Pr(\Theta_1|I)}{Pr(E|\Theta_2,I)Pr(\Theta_2|I)} \times \frac{Pr(\Theta_2|I)}{Pr(\Theta_1|I)} \\ &= \frac{Pr(E|\Theta_1,I)}{Pr(E|\Theta_2,I)} \ , \end{aligned}$$

which is the likelihood ratio.

propositions, the probability of E for the set of multiple propositions must be assigned using the rule of the extension of the conversation to take into account each of the sub-propositions. In this case, the Bayes factor depends on probabilities of the propositions, in addition to the probabilities of E given each of the propositions.[16]

The odds form of Bayes' theorem highlights what the value of a piece of evidence is with regard to the propositions of interest in a case: the value of the evidence is equal to the Bayes factor. The Bayes factor shows by how much one must modifiy the prior odds to take into account the information denoted by E. A Bayes factor greater than one means that this information supports Θ_1, a Bayes factor less than one means that this information supports Θ_2, and a Bayes factor equal to one means that this information does not support either proposition more than the other.[17] The more the Bayes factor is different from one (in either direction), the more important is the contribution of E for assessing the posterior odds. This valuable insight on the value of the evidence, and on how to combine it with a factfinder's previous knowledge and beliefs, has given the Bayesian approach support among legal scholars, statisticians and forensic scientists (e.g., Lindley, 1975, 1977b; Lempert, 1977; Redmayne, 2001; Kaye, 1979; Evett, 1983, 1984, 1986; Evett et al., 1987; Evett, 1990; Aitken and Stoney, 1991; Saks and Koehler, 1991; Robertson and Vignaux, 1993; Champod and Taroni, 1993, 1994; Robertson and Vignaux, 1995a,b; Evett et al., 1998; Taroni et al., 1998, 2002; Aitken and Taroni, 2004, 2008; Aitken et al., 2010; Nordgaard and Rasmusson, 2012).

In forensic science, the Bayesian framework has formed the basis for a model of case assessement and interpretation that promotes the use of a Bayes factor, in particular a likelihood ratio, for representing the value of scientific evidence (Cook et al., 1998a,b, 1999; Evett et al., 2000b,c). This model shows that a logical approach for interpreting scientific evidence (Evett and Weir, 1998; Jackson, 2000; Evett et al., 2000c):

(i) depends on the conditioning information I, which consists of the case's circumstances, and on the assumptions made in the model;

(ii) must consider at least two mutually exclusive propositions; and

(iii) consists of assigning the probability of the evidence given each of these propositions (i.e., for forming a likelihood ratio).

After providing a sound foundation, this logical approach now draws the attention of forensic scientists to more subtle considerations for implementing it in practice. First of all, the Bayes factor is a function of the propositions Θ_1 and Θ_2, which means that it will vary for different formulations of these propositions. This can be confusing if one is not aware of the different possible formulations and how they affect the evaluation of the Bayes factor. In particular, one can formulate these propositions at different levels in the hierarchy of propositions, and, in a multiple trace problem, they may differ in the number of traces they consider. The hierarchy of propositions distinguishes between (Cook et al., 1998b):

[16] Mathematically, the Bayes factor for a pair of propositions composed each of a set of multiple propositions is equal to:

$$\begin{aligned} \textit{Bayes factor} &= \frac{Pr(E|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E|\Theta_2, I)Pr(\Theta_2|I)} \times \frac{Pr(\Theta_2|I)}{Pr(\Theta_1|I)} \\ &= \frac{\sum_{i=1}^{n} Pr(E|\Theta_{1,i}, I)Pr(\Theta_{1,i}|I)}{\sum_{k=1}^{m} Pr(E|\Theta_{2,k}, I)Pr(\Theta_{2,k}|I)} \times \frac{Pr(\Theta_2|I)}{Pr(\Theta_1|I)} \,. \end{aligned}$$

[17] If the Bayes factor is equal to one, the posterior odds are equal to the prior odds. In this case, E does not provide the court with any useful information, and is considered to be logically irrelevant (Lempert, 1977). Lempert (1977) speaks of **logically relevant** evidence when its Bayes factor is different from one, an idea directly translated from the US Federal Rule of Evidence 401 into the mathematical language of the Bayesian framework:

> "Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

- **source level** propositions stating whether a trace comes from a particular item or person;
- **activity level** propositions describing the activities that may have caused the transfer of the traces;
- **offense level** propositions stating whether an individual committed the crime (e.g., the propositions on page 9 are offense level propositions);

and, for DNA traces (Evett et al., 2002):

- **sub-source level** propositions stating whether DNA comes from a particular person.

In a multiple trace problem, the formulation of the propositions may include either all of the traces or only one trace (Meester and Sjerps, 2003, 2004a). For each different pair of propositions, the Bayes factor will involve different parameters, and therefore lead to different values. Second, the mathematical expressions become increasingly complex and less transparent for non-statisticians as they take into account more and more variables that are not necessarily independent of each other. Third, observations, analytical results and traces rarely come alone so that the evaluation of the value of these results requires combining several observations, analytical results and traces that may not be independent of each other. Up to now, research has not addressed these points in a satisfactory way. Some of these points will be discussed in further detail in Part II. The next section shows how to extend Bayesian reasoning to decision making.

2.4 Maximizing the expected utility for making decisions

A factfinder's ultimate task is to reach a decision. In the criminal case discussed in the previous section, this decision is either to convict or to acquit the defendant: the factfinder wants to convict a guilty defendant and acquit an innocent defendant. Yet, as seen in the previous section, the uncertainty inherently present prevents the factfinder from ever being sure about the defendant's guilt or innocence (Kaplan, 1968). Probability theory, used for describing this uncertainty, does not tell the factfinder how to decide.

The normative theory of decision making that directly follows from probability theory is the principle of **maximizing the expected utility** (e.g., Good, 1952; Savage, 1972; Raiffa and Schlaifer, 1961; Lindley, 1977a; Edwards, 1991 - 1992). According to Good (1952):

> once the theory of probability is taken for granted, the principle of maximizing the expected utility per unit time is the only fundamental principle of rational behavior. (p. 111)

Here **rational** behavior is understood as the **coherent** behavior described in Section 2.2. According to Lindley (1997), the principle of **maximizing the expected utility** is the only decision rule that guarantees coherence. According to Edwards (1991 - 1992):

> any decision rule meaningfully different from expected utility maximization can be shown to be inferior to it, in the sense that any other decision rule will implement the decision maker's values less well in the long run. (p. 1059)

Hence, coherent decision making that implements the decision-maker's values as well as possible is equal to making decisions according to the principle of **maximizing the expected utility**.

The principle of **maximizing the expected utility** is the following: the decision-maker perceives different degrees of satisfaction for the decision's different possible consequences, and wants to make the choice that provides her with the greatest satisfaction. However, not knowing which consequence each of the possible choices will lead to, she can only choose

the course of action she *expects* to provide her with the greatest satisfaction, by using her probabilities for the uncertain event as the weights for the degrees of satisfaction associated with the possible consequences. Hence, she evaluates the expected satisfaction for each of the choices, and chooses the one with the maximum expected satisfaction. This procedure for making choices is the most widely applied principle in **decision theory** (e.g., Peterson, 2009), and is the only one that will be discussed in this thesis. If one measures the degrees of satisfaction using **utilities** (Section 3.3), this idea comes down to **maximizing the expected utility**. Analogously, if one measures the degrees of dissatisfaction using **losses** (Section 12.2.3), we speak of **minimizing the expected loss**. Section 3 presents the mathematical meaning of these notions. For now, it suffices to say that utilities and losses quantify the degrees of satisfaction and dissatisfaction, respectively, and that decision theory advocates maximizing the expected satisfaction, or minimizing the expected dissatisfaction.

Kaplan (1968) recognized that this principle could be applied to judicial decision making. For a criminal trial, he considered the losses associated with convicting and acquitting the defendant: the losses associated with convicting a guilty and acquitting an innocent defendant are set to 0, so that the decision problem reduces to defining the loss of convicting an innocent defendant and the loss of acquitting a guilty defendant (Table 2.1). Applying

Table 2.1: *A regret matrix for the decision of convicting or acquitting the defendant. The columns represent the propositions, the rows the actions, and the table is filled out with the losses associated to each of the possible consequences. The best consequences are those with a loss of 0, here either convicting a guilty defendant or acquitting an innocent person. The loss values associated to the other two possible consequences are positive numbers that reflect the extent of the loss associated with each of the false verdicts.*

	Θ_1: guilty	Θ_2: innocent
convict	0	*loss of convicting innocent*
acquit	*loss of acquitting guilty*	0

the principle of minimizing the expected loss, the factfinder convicts the defendant if the expected loss of acquitting is greater than the expected loss of convicting. Mathematically, if the probability $Pr(\Theta_1|E,I)$ from the previous section denotes the factfinder's final probability that the defendant is guilty, and with $Pr(\Theta_2|E,I) = 1 - Pr(\Theta_1|E,I)$, we obtain the following decision rule:

convict if and only if:

$$Pr(\Theta_1|E,I) \times \textit{loss of acquitting guilty} > [1 - Pr(\Theta_1|E,I)] \times \textit{loss of convicting innocent} \ .$$

Mathematical reformulation of this equation leads to a threshold probability for $Pr(\Theta_1|E,I)$ marking the transition point on a scale from 0 to 1 where the factfinder should switch from acquitting the defendant to convicting the defendant:

convict if and only if:

$$Pr(\Theta_1|E,I) > \frac{1}{1 + \frac{\textit{loss of acquitting the guilty}}{\textit{loss of convicting the innocent}}} \ . \tag{2.5}$$

Hence, the application of the principle of expected utility maximization provides a threshold probability for choosing a verdict based on the relative magnitude of the losses assigned to the possible erroneous verdicts. For example, if these losses have the same value, then the threshold probability (Eq. (2.5)) is 0.5. This corresponds to the "preponderance of evidence" rule applied in civil cases. For criminal cases, the loss of convicting an innocent person has long been considered to be greater than the loss of acquitting a guilty person (e.g., Blackstone (1765-1769), Hale (1847), and Fortescue (F. Grigor transl. 1917) cited by

Kaplan, 1968, p. 1077),[18] leading to a higher threshold probability.[19]

The application of such a threshold minimizes the expected loss. It is important to note that this does not always minimize the *actual* loss (Kaye, 1997, 1999). Yet, as Kaye (1999) puts it,

> it is hard to know what rule would do better. (p. 25)

According to Edwards (1991 - 1992):

> we like the decisions that control our lives to be wise. That is, we want these decisions to implement our values as well as they can be made to do so. In a world of fallible inferences, we can accomplish this only by maximizing expected utility; fact-finding specifies the probabilities that define expectations. (p. 1060)

Given that we want judicial decision making to implement society's values as well as possible, decision theory provides the most adequate normative model for decision problems in law. According to Kaye (1988):

> This Bayesian decision theory, or some less elaborate variant on it, has proved fruitful in the study of the burden of persuasion. As with Bayesian inference, its usefulness in describing the decision-making of jurors and mock jurors has been the subject of many empirical studies. Likewise, (...) many a law review article has drawn on the power of decision theory to analyze the burden of persuasion. (p. 12)

On the one hand, legal scholars have recognized the potential of Bayesian inference and decision theory as a normative model for the legal fact-finding process. On the other hand, forensic scientists have recognized the potential of Bayesian inference for handling forensic inference problems, yet have not yet fully recognized the potential of decision theory as a normative model for making and justifying decisions made by forensic scientists.[20] Yet, by definition, forensic science is the application of scientific knowledge to legal problems.[21] It is therefore only a logical extension that the principles that form a normative model for the legal fact-finding process should also apply to forensic science.

Like the factfinder, forensic scientists must make decisions in the presence of uncertainty. No decision rule can guarantee that the scientist makes the correct choice in every case. As a consequence, the most that a court of law can ask of the forensic scientist is that her decisions are coherent given her state of knowledge and objectives at the moment of making the decision. The application of decision theory assures coherent decision making. Hence, forensic decision problems will be modeled using decision theory.

---

[18] The cited texts of Blackstone, Hale and Fortescue are freely available on the internet at the following websites:

- Blackstone, W. Commentaries of the Laws of England, Book 4, Chapter 27, 1765-1769, made available by the Lonang Institute: *http://www.lonang.com/exlibris/blackstone/bla-427.htm* (last visited on 23.06.2013);
- Hale, M. Historia placitorium coronae: The history of the pleas of the crown, Volume 2, W. Stokes and E. Ingersoll ed., Robert H. Small, Philadelphia, 1847: p. 288, made available by Google Books: *http://books.google.ch/books?id=gB4eAAAAMAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false* (last visited on 23.06.2013);
- Fortescue, J. Commendation of the Laws of England, F. Grigor translation, Sweet & Maxwell, Ltd, London, 1917: p. 45, digitalized version by Microsoft Corporation in cooperation with Cornell University Libraries, 2007, made available by Cornell University Library: *http://archive.org/stream/cu31924021661909#page/n65/mode/2up* (last visited on 23.06.2013).

[19] For further discussions on this threshold probability and its implications for judicial decision making, see for example Kaplan (1968); Kaye (1986); Connolly (1987); Kaye (1997, 1999).

[20] There have been a few publications that go in this direction, yet they are scarce and only apply to a few isolated decision problems (Chapter 5).

[21] Merriam-Webster dictionary, at *http://www.merriam-webster.com/dictionary/forensic* (last visited on 31.05.2013).

Chapter 3

Decision Theory

Decision theory focuses on choosing and justifying a rational course of action on the basis of the inferences made in the presence of uncertainty (e.g., Savage, 1951, 1972; Edwards, 1954; Pratt et al., 1964). The difficulty of decision making under uncertainty is that an unknown condition determines the consequence of the chosen action. This condition, called the **state of nature** (or the **state of the world**), is beyond the control of the decision-maker and unknown to the decision-maker at the time of making the decision (and may remain unknown even after making the decision). If the decision-maker knew the state of nature, she would be able to predict each action's consequence with certainty and choose the action whose consequence brings her the most satisfaction. However, without this perfect knowledge, she can only consider her probabilities for the various states of nature, and not which one of them is true. She must thus act under uncertainty. Decision theory's major feature is that of combining probabilities for measuring this uncertainty with utilities or losses decribing the decision-maker's preferences among the decision's possible consequences. By weighing the utility or loss of each consequence with its probability of occurring, the rational[1] decision-maker makes the choice that maximizes her expected utility (or minimizes her expected loss).

3.1 Origins

The study of rational decision making is nothing new. The first notions of utility date back to Blaise Pascal's argumentation in 1670 (known as "Pascal's wager") on why people should choose to believe in God.[2] Decision theory, or more specifically the theory of **expected utility maximization**, has its roots in the solution to Nicolas Bernoulli's St. Petersburg Paradox (Lengwiler, 2009).[3] Yet, it wasn't until the 1940s that the theory became more

[1] **Rational** is a synonym to **coherent**. **Rational decision making** therefore means making decisions according to the principle of maximizing the expected utility, and a **rational decision-maker** is one who acts according to this principle.

[2] According to Pascal, the disutility of living as if God did not exist when he actually does exist is so great (eternal damnation in hell) that people should live as though God exists, even if the prior probability of God's existence is considered very small, and even though the utility during a person's lifetime is considered greater if the person lives as though God didn't exist, than if the person leads a religious life. Though this argumentation led to many philosophical debates, it was the first description of a decision based on the expected utility of the possible consequences (Lengwiler, 2009).

[3] The St. Petersburg Paradox refers to gamble in which a fair coin is tossed: if it lands on tails, the game is over, but if it lands on heads, the coin is tossed again, and the game is repeated until the coin lands on tails. The gambler's gain depends on the number of tosses, doubling with each toss (for example, if 1 toss gives a gain of 2 CHF, 2 tosses gives a gain of 4 CHF, 3 tosses a gain of 8 CHF, and so on). The question is, how much is the gambler willing to pay to participate in this gamble? The paradox consists of the fact that the expected value of the gamble (obtained by summing the products of the value of the expected gain and the probability of that gain) is infinity, however, no person is willing to bet their entire fortune on this gamble. Gabriel Cramer (in 1728) and Daniel Bernoulli (in 1738) argued that this paradox could be solved

concrete, when von Neumann and Morgenstern (1947) postulated a series of axioms that utilities need to follow (Section 3.3). Savage (1972) then extended these ideas by eliciting a set of axioms that form the basis of Bayesian decision theory. In addition to von Neumann and Morgenstern (*The Theory of Games and Economic Behavior* (1947)) and Savage (*The Foundations of Statistics* (1954)), the major contributions came from Ramsey (*Truth and Probability* (1926)), and Wald (*Statistical Decision Functions* (1950)). A clear explanation of each of these contributions can be found in Parmigiani and Inoue (2009).

3.2 Mathematical notation

There are four distinct spaces (Raiffa and Schlaifer, 1961), denoted here as follows:

- **an action space $\mathcal{A}$:** This space consists of an exhaustive[4] list of mutually exclusive[5] actions denoted $a_1, a_2, \ldots, a_m$.
- **a state space Θ:** This space contains the possible states of nature. Here, these are discrete so that Θ is partitioned into n possible states denoted $\Theta_1, \Theta_2, \ldots, \Theta_n$.
- **a test space $\mathcal{T}$:** This space consists of an exhaustive list of mutually exclusive actions that can be undertaken to obtain more information, $t_1, t_2, \ldots, t_p$, and the possibility of not undertaking any action to obtain more information, denoted t_0.
- **a space of possible test results $\mathcal{E}$:** This space contains the possible results obtained from performing an action in $\mathcal{T}$, and these results are denoted $E_1^{t_h}, E_2^{t_h}, \ldots, E_{q_h}^{t_h}$, $h = 0, 1, 2, \ldots, p$, with q_h denoting the number of possible results obtained from performing action t_h.

The combination of $\mathcal{A}$ with Θ, that is $\mathcal{A} \times \Theta$, produces a fifth space:

- **the space of the actions' possible consequences $\mathcal{C}$:** The consequence $C(a_i, \Theta_j)$, abbreviated by C_{ij}, denotes the consequence of the decision-maker choosing action a_i, $i \in \{1, 2, \ldots, m\}$, when Θ_j, $j \in \{1, 2, \ldots, n\}$, is true.

Further random variables will be denoted with capital roman or greek letters according to the same scheme as Θ: the capital letter without subscripts denotes the variable's set of possible states, and the capital letter with a subscript one of its states. For Boolean variables[6], the capital letter describing the random variable denotes the state "true" of this variable, and this capital letter preceded by the symbol "¬" denotes its negation, that is, the state "false" of this variable. The results presented in this thesis consider only discrete random variables so that each has a predefined number of exhaustive and mutually exclusive states. In some cases, the same capital letter will be used to denote multiple appearances of a particular variable in a problem, and in these cases, prime marks or superscripts will be used to distinguish between them.

Each variable is characterized by a probability distribution representing the degree of belief the decision-maker has in each of the states being true at a given point in time. These degrees of belief are described by subjective (or personal) probabilities (de Finetti, 1980), as described on page 9.

---

by considering expected utilities instead of expected values (Bernoulli, 1954): Cramer proposed to calculate the utility as the square root of the monetary gain, and Bernoulli proposed the logarithm of the gain as the utility function.

[4] The list is **exhaustive** when the decision-maker inevitably chooses one of the actions in the list. Note that if it is possible for the decision-maker to do nothing, then this possibility must be defined as one of the actions for the action space to be exhaustive.

[5] The actions are **mutually exclusive** if the decision-maker can never choose more than one of them at one time.

[6] A Boolean variable is a variable with two possible states: "true" and "false".

To quantify the decision-maker's preferences among the actions' possible consequences, utilities or losses will be used, denoted by the following expressions:

- $u(a_i, \Theta_j) = u(C_{ij})$ is the **utility** of consequence C_{ij}, $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$;
- $l(a_i, \Theta_j) = l(C_{ij})$ is the **loss** of consequence C_{ij}, $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$.

The decision-theoretic analysis of a problem will compute the expected utility $\bar{u}(a_i|\cdot)$, or the expected loss $\bar{l}(a_i|\cdot)$, of each of the possible actions $a_i \in \mathcal{A}$, with the information available to the decision-maker at the time of making the decision to the right of the conditioning bar. For example, $\bar{u}(a_1|I)$ is the decision-maker's expected utility for performing action a_1 given the information contained in I.

The rest of the symbols used follow conventional mathematical notation.

3.3 Utilities

Utilities measure the desirability of each of the possible consequences C_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, based on the decision-maker's objectives and personal preferences. A utility function $u : \mathcal{C} \to \mathbb{R}$ represents the desirability of each possible consequence C_{ij} as a numerical value. It is a subjective notion (again in the sense of personal, not arbitrary), meaning that decision-makers with different objectives and preferences may have different preference orderings of the possible consequences.

If we consider that each possible action a_i, $i \in \{1, 2, \ldots, m\}$, leads to a particular probability distribution over $\mathcal{C}$, then the decision-maker's preference ordering of C_{ij} leads the decision-maker to have a preference ordering of these possible probability distributions. These probability distributions can be seen as gambles specifying the probabilities of obtaining each of the possible consequences C_{ij}.

Utility theory states that a probability distribution (or gamble) 1, denoted here as g_1, is preferred to a probability distribution (or gamble) 2, denoted here as g_2, if and only if the expected utility of the consequence obtained by probability distribution (or gamble) 1 is greater than the expected utility of the consequence obtained by probability distribution (or gamble) 2:

$$g_1 \succ g_2 \;\Leftrightarrow\; \bar{u}(g_1) > \bar{u}(g_2) \;,$$

where the symbol $\succ$ means "is preferred to".

To ensure that there exists a unique utility function possessing this property, the following axioms should be satisfied (e.g., DeGroot, 1970; Berger, 2010):

Axiom 1: *It is possible for the decision-maker to order the possible probability distributions (gambles) from best to worst, or to explicitly state her indifference between two, or several, of them: i.e., for each g_1 and g_2, either $g_1 \succ g_2$ (g_1 is preferred to g_2), $g_1 \prec g_2$ (g_2 is preferred to g_1), or $g_1 \sim g_2$ (the decision-maker is indifferent between g_1 and g_2).*

Axiom 2: *These preferences respect the property of transitivity: i.e., if $g_1 \preceq g_2$ and $g_2 \preceq g_3$, then $g_1 \preceq g_3$ must be true, where the symbol $\preceq$ means "is not preferred to".*

Axiom 3: *The preferences between two probability distributions (gambles) do not change when these are represented as compound gambles*[7]*: i.e., if $g_1 \preceq g_2$, then $pg_1 + (1-p)g \preceq pg_2 + (1-p)g$ for any probability distribution (gamble) g and any probability p.*

[7] A compound gamble between two probability distributions (gambles) means that one of them will occur with a probability p and the other with a probability $(1-p)$.

Axiom 4: *If* $g_1 \preceq g_2 \preceq g_3$, *then there must exist two probabilities* $p_1, p_2 \in (0,1)$ *so that* $p_1 g_1 + (1-p_1)g_3 \preceq g_2$ *and* $g_2 \preceq p_2 g_1 + (1-p_2)g_3$. *This means that all gambles are comparable with each other in the sense that no gambles are infinitely desirable or infinitely undesirable.*

Following from these axioms, one can prove that there exists a unique number $p \in (0,1)$ for any consequences C_1, C_2 and C_3, where $C_1 \preceq C_2 \preceq C_3$, so that the decision-maker is indifferent between a gamble that produces consequence C_2 with certainty and a gamble that produces consequence C_1 with probability p and consequence C_3 with probability $(1-p)$ (DeGroot, 1970):

$$C_2 \sim pC_1 + (1-p)C_3 \ . \tag{3.1}$$

When this is satisfied it can be proved that (DeGroot, 1970):

$$u(C_2) = p \times u(C_1) + (1-p) \times u(C_3) \ . \tag{3.2}$$

Such a utility function is unique up to a linear transformation (e.g., DeGroot, 1970; Berger, 2010).

These axioms thus provided decision theory with a means for quantifiying and unifying the satisfaction obtained from each of the possible consequences on a single scale. Further, the comparison of gambles defines coherent numerical values for a decision-maker's utilities. This is best seen by considering a utility scale from 0 to 1, where the most desirable consequence (which we shall denote by C^+) has a utility of 1:

$$u(C^+) = 1 \ .$$

Analogously, the least desirable consequence (denoted by C^-) has a utility of 0:

$$u(C^-) = 0 \ .$$

The utilities of the remaining consequences $u(C_{ij})$, $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$, are somewhere between 0 and 1. The decision-maker can find their numerical values by comparing the following two gambles (e.g., Pratt et al., 1964):

gamble 1: *obtain consequence* C_{ij} *for sure*,

gamble 2: *obtain the most desirable consequence* C^+ *with a probability of* $Pr(C^+)$ *and the least desirable consequence* C^- *with a probability of* $Pr(C^-) = 1 - Pr(C^+)$.

Which gamble does the decision-maker prefer? If $Pr(C^+)$ is very small, the decision-maker will choose gamble 1. If $Pr(C^+)$ is very large, she shall choose gamble 2. Thus, there must be a single value for $Pr(C^+)$ marking the turning point somewhere between these two probabilities, where the decision-maker will be indifferent between gambles 1 and 2. This value for the probability of $Pr(C^+)$ is equal to the decision-maker's utility for consequence C_{ij}. In other words:

$$\begin{aligned} u(C_{ij}) &= u(C^+) \times Pr(C^+) + u(C^-) \times (1 - Pr(C^+)) \\ &= 1 \times Pr(C^+) + 0 \times (1 - Pr(C^+)) \\ &= Pr(C^+) \ . \end{aligned} \tag{3.3}$$

For example, a utility of $u(C_{ij}) = 0.7$ signifies that the decision-maker is indifferent between obtaining consequence C_{ij} for sure and a gamble that gives her the best possible consequence with a probability of 0.7 and the worst possible consequence with a probability of 0.3. A decision-maker's utilities must satisfy all possible pairs of gambles for these utilities to be coherent and lead to coherent choices of actions.

3.4 Losses

A loss is defined as a non-negative number, equal to the difference between the utility of the best possible consequence C^+ and the utility of the consequence for which the loss is being evaluated C_{ij} (e.g., Berger, 2010):

$$l(C_{ij}) = u(C^+) - u(C_{ij}) \ . \tag{3.4}$$

Translating the utility scale into losses according to Eq. (11.11), produces a scale ranging from 0 to describe the *most* desirable consequence:

$$\begin{aligned} l(C^+) &= u(C^+) - u(C^+) \\ &= 1 - 1 \\ &= 0 \ , \end{aligned} \tag{3.5}$$

to 1 to describe the *least* desirable consequence:

$$\begin{aligned} l(C^-) &= u(C^+) - u(C^-) \\ &= 1 - 0 \\ &= 1 \ . \end{aligned}$$

From Eq. (11.11) and (3.5), it follows that the loss assigned to consequence C_{ij} is equal to:

$$l(C_{ij}) = 1 - u(C_{ij}),$$

which, according to Eq. (3.3), is equal to

$$\begin{aligned} l(C_{ij}) &= 1 - Pr(C^+) \ , \\ &= Pr(C^-) \ . \end{aligned}$$

Thus, the loss assigned to consequence C_{ij} is equal to the probability $Pr(C^-)$ which makes the decision-maker indifferent between obtaining consequence C_{ij} for sure, and a gamble leading to the least desirable consequence C^- with a probability of $Pr(C^-)$ and to the most desirable consequence C^+ with a probability of $Pr(C^+) = 1 - Pr(C^-)$. For example, a loss of $l(C_{ij}) = 0.7$ means that the decision-maker is indifferent between obtaining consequence C_{ij} for sure and obtaining the worst possible consequence with a probability of 0.7 and a gamble that gives her the best possible consequence with a probability of 0.3.

3.5 One-stage decision problems

In a one-stage decision problem, the decision consists of choosing an action in $\mathcal{A}$ based on the decision-maker's current knowledge of the state of nature. In this case, the decision-maker has a utility or loss function over all of the possible consequences in $\mathcal{C}$. A rational decision-maker will choose the action that maximizes the satisfaction she expects to obtain. In terms of utilities, this corresponds to choosing the action that maximizes the expected utility:

$$\underset{i}{arg\ max}\ \bar{u}(a_i|\cdot) = \underset{i}{arg\ max}\ \sum_{j=1}^{n} u(C_{ij}) Pr(\Theta_j|\cdot) \ ,$$

and in terms of losses, the action that minimizes the expected loss:

$$\underset{i}{arg\ min}\ \bar{l}(a_i|\cdot) = \underset{i}{arg\ min}\ \sum_{j=1}^{n} l(C_{ij}) Pr(\Theta_j|\cdot) \ .$$

This action is called the **Bayes action** (e.g., Berger, 2010).

3.6 Two-stage decision problems

The probability distribution over Θ is conditioned on the decision-maker's information. In Section 2.3, the letter I denoted the information the decision-maker has at a given point in time. In this section, let this point in time be when the decision-maker is first faced with the decision of choosing an action in $\mathcal{A}$. In a two-stage decision problem, the question of interest is, "Should the decision-maker make this decision with her current knowledge I, or should she acquire an additional piece of information before making this decision?" The decision-maker therefore faces a sequence of two decisions:

(1) the preliminary decision of choosing an action in $\mathcal{T}$, and

(2) the terminal decision of choosing an action in $\mathcal{A}$.

Note that making the terminal decision is a one-stage decision problem, because when the decision-maker comes to this point (after having made the preliminary decision), she will effectively be facing a one-stage decision problem. Thus, solving the terminal decision problem follows the principle of maximizing the expected utility (minimizing the expected loss), as described in Section 3.5. The focus of a two-stage decision problem is on how to make the preliminary decision of obtaining an additional piece of information.

A rational decision-maker (i.e., a decision-maker wanting to maximize the satisfaction she expects to obtain from choosing an action in $\mathcal{A}$) will acquire an additional piece of information if its expected value is greater than the cost of aquiring it. This requires a quantification of the expected value of the additional piece of information.

Decision theory defines the **expected value of information**, denoted $EVOI(t_h)$, as the difference between the maximum expected utility *with* the information obtained from performing action t_h and the maximum expected utility *without* this information (or the difference between the minimum expected loss *without* the information obtained from performing action t_h and the minimum expected loss *with* this information) with regard to the terminal decision.

Consider for example the preliminary decision of choosing between t_0 (not obtaining any additional information) and t_1 (performing a particular action that produces an additional piece of information). For a decision-maker who defined her preferences in terms of utilities, the maximum expected utility if she chooses t_0 is:[8]

$$\begin{aligned} max_i\ \bar{u}(a_i|E_1^{t_0}, I) &= max_i \sum_{j=1}^{n} u(C_{ij}) Pr(\Theta_j|E_1^{t_0}, I) \\ &= max_i \sum_{j=1}^{n} u(C_{ij}) Pr(\Theta_j|I)\ , \end{aligned} \tag{3.6}$$

since $E_1^{t_0}$ is a dummy variable that does not contain any information (i.e., it is the outcome of choosing action t_0, which is to do nothing, and nothing comes from doing nothing). If the decision-maker chooses t_1, the maximum expected utility becomes the weighted average of the maximum expected utilities for each of the different possible realizations of this new piece of information (which is still unknown at the moment when the decision-maker must decide to acquire or not acquire this information):

$$\sum_{k=1}^{q} {}_1 max_i\ \bar{u}(a_i|E_k^{t_1}, I) Pr(E_k|I) = \sum_{k=1}^{q} {}_1 max_i \sum_{j=1}^{n} u(C_{ij}) Pr(\Theta_j|E_k^{t_1}, I) Pr(E_k^{t_1}|I)\ , \tag{3.7}$$

where $Pr(\Theta_j|E_k^{t_1}, I)$ is the updated probability of Θ_j upon learning $E_k^{t_1}$, that is, the posterior probability of Θ_j. The mathematical relationship between the posterior probability $Pr(\Theta_j|E_k^{t_1}, I)$ and the prior probability $Pr(\Theta_j|I)$ is given by Bayes' theorem (Eq. (2.1)).

[8] For preferences defined in losses, the reasoning that leads to the $EVOI(t_1)$ is analogous to the reasoning presented here for utilities. The formula for losses is presented hereafter in Eq. (3.10).

Inserting Eq. (2.1) into Eq. (3.7) produces:

$$\begin{aligned}\sum_{k=1}^{q} {}_1max_i\ \bar{u}(a_i|E_k^{t_1},I)Pr(E_k^{t_1}|I) &= \sum_{k=1}^{q} {}_1max_i\ \sum_{j=1}^{n} u(C_{ij})\frac{Pr(E_k^{t_1}|\Theta_j,I)Pr(\Theta_j|I)}{Pr(E_k^{t_1}|I)}Pr(E_k^{t_1}|I) \\ &= \sum_{k=1}^{q} {}_1max_i\ \sum_{j=1}^{n} u(C_{ij})Pr(E_k^{t_1}|\Theta_j,I)Pr(\Theta_j|I)\ . \qquad (3.8)\end{aligned}$$

The $EVOI(t_1)$ is the difference between Eqs. (3.8) and (3.6):

$$\begin{aligned}EVOI(t_1) &= \sum_{k=1}^{q} {}_1max_i\ \bar{u}(a_i|E_k^{t_1},I)Pr(E_k^{t_1}|I) - max_i\ \bar{u}(a_i|I) \qquad (3.9)\\ &= \sum_{k=1}^{q} {}_1max_i\ \sum_{j=1}^{n} u(C_{ij})Pr(E_k^{t_1}|\Theta_j,I)Pr(\Theta_j|I) - max_i\ \sum_{j=1}^{n} u(C_{ij})Pr(\Theta_j|I)\ .\end{aligned}$$

For a decision-maker who defined her preferences in terms of losses, an analogous line of reasoning leads to the following expression for the $EVOI(t_1)$:

$$\begin{aligned}EVOI(t_1) &= min_i\ \bar{l}(a_i|I) - \sum_{k=1}^{q} {}_1min_i\ \bar{l}(a_i|E_k^{t_1},I)Pr(E_k^{t_1}|I) \qquad (3.10)\\ &= min_i\ \sum_{j=1}^{n} l(C_{ij})Pr(\Theta_j|I) - \sum_{k=1}^{q} {}_1min_i\ \sum_{j=1}^{n} l(C_{ij})Pr(E_k^{t_1}|\Theta_j,I)Pr(\Theta_j|I)\ .\end{aligned}$$

The $EVOI(t_1)$ is always greater than or equal to zero, reflecting the informative value of additional information (e.g., Raiffa and Schlaifer, 1961; Good, 1967; Watson and Brown, 1978; DeGroot, 1984). If this value is greater than the cost of performing t_1, the decision-maker should acquire the additional information. If this value is less than the cost, the rational decision-maker would choose not to acquire the information. In other words, the decision-maker in this decision problem seeks to maximize the **expected net gain** ($ENG(t_h),\ h = 0,1$) (Raiffa and Schlaifer, 1961) (sometimes also referred to as the **expected profit** (Jensen and Nielsen, 2007)). She should therefore choose:

$$\underset{h}{arg\ max}\ ENG(t_h) = \underset{h}{arg\ max}\ [EVOI(t_h) - c(t_h)]\ ,\ h = 0,1\ , \qquad (3.11)$$

where $c(t_h)$ denotes the absolute value of the cost of choosing action t_h. The cost $c(t_0)$ is equal to 0, and the cost $c(t_1)$ to the cost of performing t_1 that will produce information $E_k^{t_1}$. Note that to evaluate the $ENG(t_h)$, $c(t_h)$ and the $EVOI(t_h)$ must be quantified in the same units (Raiffa and Schlaifer, 1961; Lindley, 1997). Equation (3.11) is readily extendable to $h = 0,1,2,\ldots,p$ for solving a preliminary decision with multiple possible tests[9] producing different types of information.

3.6.1 Optimal sample size

Raiffa and Schlaifer (1961) originally developed the above equations (Eqs. (3.9)-(3.11)) for determining the optimal sample size. In a sampling problem, the decision-maker wishes to determine the proportion of units in a confined entity that have a particular characteristic. The classic case is that there is a group of N discrete units, all identical except for the presence or absence of the trait of interest (i.e., the trait for which the decision-maker wishes to determine the proportion in the group). The decision-maker has the possibility of examining individual units. Assuming that no errors are possible in this examination process, she will determine with certainty whether or not the trait is present in these examined units. Thus, examining every single unit in the group allows the decision-maker to determine the

[9]Note that the word "test" is used here in a decision-theoretic sense to refer to any possible action belonging to space $\mathcal{T}$ (see Section 3.2). In the context of forensic science, these "tests" take the form of examinations and analyses.

proportion with certainty. However, a cost associated with the examination of a unit, and the decision-maker's limited resources do not allow her to examine every single one. The question is therefore, "How many units should the decision-maker examine?"

This is a two-stage decision problem where making a statement on the proportion is the terminal decision, and choosing the sample size is the preliminary decision (Lindley, 1997). Let $h = 0, 1, \ldots, N$ denote the possible sample sizes (i.e., the number of units to examine), and t_h the action of examining h items. First, as h increases, the $ENG(t_h)$ will increase because the information obtained from the samples will allow the decision-maker to make a better informed terminal decision with a greater expected utility (or a smaller expected loss). However, at some point, this increase in expected utility (or decrease in expected loss) will be smaller than the cost of examining an additional unit. At this point, the $ENG(t_h)$ decreases with increasing h, due to the sampling cost. The optimal sample size is the h for which the $ENG(t_h)$ is maximal (Raiffa and Schlaifer, 1961; Lindley, 1997).

This line of reasoning applies to both fixed sample size problems (i.e., choosing the sample size before sampling any items) and sequential analysis problems (redoing the calculation after each sample to decide whether to stop sampling and make the terminal decision, or continue sampling). For further explanations concerning these two approaches, see Taroni et al. (2010, Section 7.4).

3.6.2 The special case of perfect information

If the information provided by $E_k^{t_1}$ tells the decision-maker with certainty what the true state of nature is, then we speak of **perfect information**, and the $EVOI(t_1)$ reduces to:

$$\sum_{j=1}^{n} max_i\ u(C_{ij})Pr(\Theta_j|I) - max_i\ \sum_{j=1}^{n} u(C_{ij})Pr(\Theta_j|I)\ ,$$

for utilities, and to:

$$min_i\ \sum_{j=1}^{n} l(C_{ij})Pr(\Theta_j|I) - \sum_{j=1}^{n} min_i\ l(C_{ij})Pr(\Theta_j|I)\ ,$$

for losses. This value is called the **expected value of perfect information**. It quantifies how much perfect knowledge of the state of nature is worth in the decision problem. For a rational decision-maker, it corresponds to the maximum amount that she would pay to obtain additional information before making the terminal decision (Raiffa and Schlaifer, 1961).

However, in practice information is never perfect, that is, it will always leave some doubt on the true state of nature. To differentiate between the expected value of perfect information and the $EVOI(t_h)$ that is not perfect, literature calls the $EVOI(t_h)$ in Eqs. (3.9) and (3.10) for $h = 1$ the **expected value of partial information** (Lindley, 1985; Taroni et al., 2010), the **expected value of sample information** (Raiffa and Schlaifer, 1961), and the **expected benefit** (Jensen and Nielsen, 2007; Korb and Nicholson, 2011). Because in practice information is never perfect, the $EVOI(t_h)$ will refer to this value in the rest of this thesis.

3.7 Sensitivity analyses

The most rational course of action defined by the above equations depends on: (i) the decision-maker's utility (or loss) function, and (ii) the probability distributions over Θ and $\mathcal{E}$. In one-stage decision problems, the utility (or loss) function determines a threshold probability of Θ_j marking the transition point between two different Bayes actions. Equation (2.5) in Section 2.4 is an example for such a threshold in judicial decision making. Choosing a particular action is therefore equivalent to assigning a particular probability distribution over the possible states of nature (Edwards, 1988). If this probability distribution, for example

$Pr(\Theta_1|E_{k_h}^{t_h})$, $k_h = 1, \ldots, q_h$, depends on multiple parameters that define the relationship between Θ_1 and $E_{k_h}^{t_h}$, then the Bayes action also depends on the values assigned to these parameters. In both cases, variations in these probability assignments and parameter values may lead to different Bayes actions. For this reason,

> [a] decision analysis that does not include a set of sensitivity analyses would be regarded by virtually any practitioner as woefully incomplete if the issue at hand were of any importance or difficulty. (Edwards, 1988, p. 338)

Sensitivity analyses study the impact of the values and probabilities in a decision model on the Bayes action. The aim is to determine the range of values a particular parameter or probability can take and still lead to the same Bayes action. These ranges of values highlight the critical thresholds that will cause the Bayes action to change. These thresholds make up a decision strategy, that states under what circumstances to choose which action. This information makes the decision-maker aware of when she must assign the value of a parameter or probability with high numerical precision, and when a simple order of magnitude suffices for a coherent decision analysis. The aim of a decision-theoretic analysis of a problem is therefore to determine the decision strategy.

Chapter 4

Graphical Models

Applying probability calculus and decision theory to real problems can become complicated and time consuming when the problem contains many variables, and/or intricate dependence relationships among these variables. Graphical models provide a means to organize a problem by representing it in a qualitative structure (e.g., Cowell et al., 2007a). Bayesian networks are graphical models for making inferences according to the laws of probability, and influence diagrams are graphical models that represent both probability distributions and utilities (or losses) for computing expected utilities (losses) in addition to making inferences (e.g., Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008).

4.1 Bayesian networks

A **Bayesian network (BN)**, sometimes referred to as a normative expert system, implements probability theory in graphical models composed of (Kim and Pearl, 1983):

- **nodes** in the form of circles (○) to represent random variables; and
- **arrows** to denote the probabilistic relationships between these nodes to form a **directed acyclic graph (DAG)**[1].

The nodes can represent discrete and continuous random variables, yet this thesis considers only discrete variables. Hence, each node contains the variable's exhaustive list of mutually exclusive states (Section 3.2). Each node has a probability distribution over its states. A probability table associated with each node allows the user to define these probability distributions:

- if the node has no arrows pointing to it, then this node is called a **root** node and the user specifies the probability of each of its possible states; and
- if the node has one or several arrows pointing to it from other nodes, then this node is called a **child** node, with the node or nodes having an arrow pointing towards it called **parent** node(s), and the user specifies the conditional probabilities of each of the child node's states given each state, or combination of states, of its parent node(s).

Hence, the graphical structure defines the probabilistic relationships between the variables by conditioning the probability distribution of each child node upon its parent(s). In this way, a BN decomposes the joint probability distribution of a set of random variables $X^1, \ldots, X^n$ into the product of their probabilities conditioned upon their parents:

$$Pr(X^1, ..., X^n) = \prod_{i=1}^{n} Pr(X^i | parents(X^i)) \ . \tag{4.1}$$

[1] A **directed acyclic graph (DAG)** connects the nodes using arrows (i.e., directed edges) to form a structure of connected nodes that never forms a loop when following the paths defined by the arrows.

This expression, known as the **chain rule**, is what gives BNs their invaluable capacity of splitting up a complex inference problem into its different components (e.g., Kjaerulff and Madsen, 2008). For example, imagine a scenario that can be modeled with six Boolean random variables (i.e., variables with the two states "true" and "false" (Section 3.2)) connected as shown in Fig. 4.1(a). According to Eq. (4.1), the joint probability of them all being true

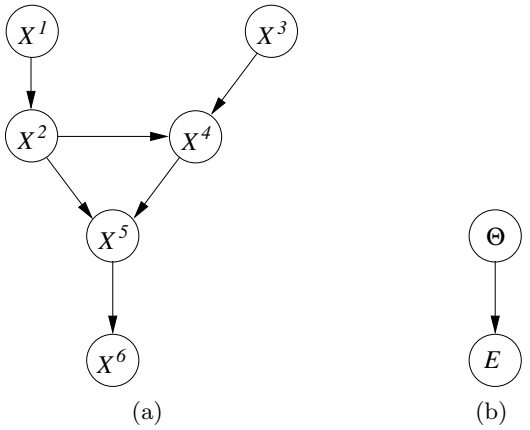


Figure 4.1: *Examples of two Bayesian networks (BNs).* **(a)** *A BN with six random variables, denoted by* $X^1, X^2, \ldots, X^6$. **(b)** *A generic structure for representing Eq. (2.4) as a BN, with* Θ *containing the propositions of interest in a case (e.g.,* Θ_1*: the defendant committed the crime and* Θ_2*: the defendant did not commit the crime, someone else in the population of potential suspects committed the crime), and E specifying the observed evidence.*

is equal to:

$$\begin{aligned} &Pr(X^1, X^2, X^3, X^4, X^5, X^6) \\ &= Pr(X^1) \times Pr(X^2|X^1) \times Pr(X^3) \times Pr(X^4|X^3, X^2) \times Pr(X^5|X^4, X^2) \times Pr(X^6|X^5) \ . \end{aligned}$$

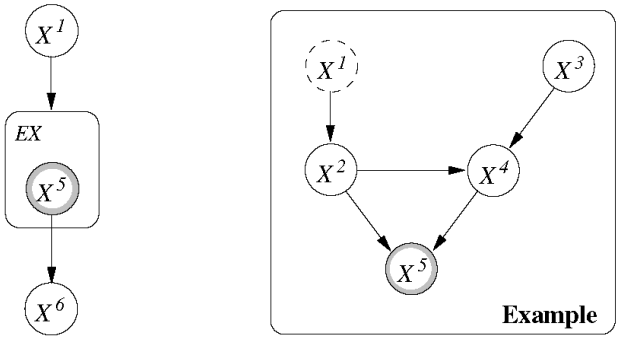
This means that instead of computing and storing $2^6 = 64$ probabilities, this model only needs to compute 24 probabilities.

BNs also allow the user to obtain the numerical values of any conditional probabilities for one or more given states of other variables in the model: for this, the user instantiates[2] the variables of the states figuring to the right of the probability's conditioning bar to these known states.[3] For example, to obtain the value of $Pr(X^2|X^1)$ the user instantiates X^1 to the state "true", and reads the value computed for the state "true" of X^2. In this way, the user can use a Bayesian network to obtain numerical values for the probabilities forming the likelihood ratio, Bayes factor, or posterior odds seen in Section 2.3 (Eq. (2.4)). Figure 4.1(b) shows the generic structure for representing Eq. (2.4) as a BN, where Θ represents the propositions of interest in a case and E specifies the observed evidence ($E_1^{t_1}$, figuring in Eq. (2.4), is one of E's states).

An object-orientated approach makes it possible to decompose a BN into substructures, and combine these on several hierarchical levels to form an **object oriented Bayesian network (OOBN)**. A substructure is referred to as an **object** or **instance** of a network class, which may be a single random variable, or a complex model of its own (such as another BN) (e.g., Koller and Pfeffer, 1997). Fig. 4.2 illustrates this concept by representing the BN in Fig. 4.1(a) as an OOBN: Fig. 4.2(a) represents the entire network. By combining different instances of one or several network classes in a hierarchical structure, an OOBN allows its

[2] **Instantiating** a variable means setting the probability of one of its states to certainty (i.e., assigning a probability of 1).

[3] This propagation of information in a BN invokes additional concepts and computations that are not discussed in this thesis. They concern the technical aspects of implementating the probabilistic models defined by BNs and are independent of the probabilistic model defined by a given BN.



(a) (b)

Figure 4.2: *A translation of the BN in Fig. 4.1(a) into an object-oriented Bayesian network (OOBN).* **(a)** *The entire structure, with the instance node EX belonging to the network class called* ***Example****.* **(b)** *The network class* ***Example****. In figure (b), node X^1 has a dashed contour to indicate that it is an input node in the network class* ***Example*** *(i.e., a node that is not part of the network class, but is a parent of at least one node in the network class). Node X^5 has a shaded contour in both figures to indicate that it is an output node of the network class* ***Example*** *(i.e., a node that is visible when an instance of the network class is inserted into another model).*

user to construct models for evaluating more complex problems (Koller and Pfeffer, 1997). That is, the underlying structure of the model is still a BN, and may also be represented as a BN, but the object-orientation allows the user to reduce the visual complexity of the interface, differentiate between several hierarchical levels, and combine a set of nodes from different models. This proves to be very useful for combining a set of identical network fragments that form a repetitive pattern in a regular BN (Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008).

4.2 Influence diagrams

An **influence diagram (ID)**, sometimes referred to as a normative expert system, combines probability and decision theory in a graphical model (e.g., Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008). IDs consist of (Shachter, 1986):

- **nodes** to represent
  - random variables in the form of circles (○),
  - the decisions that need to be made in the form of squares (□), and
  - the utilities or losses in the form of diamonds (◇); and
- **arrows** to denote
  - the functional relationships between these nodes in the form of arrows with continuous lines (⟶), and
  - precedence links between two decision nodes to define the sequence of decisions, and from a random variable to a decision node to indicate that the state of that variable is known when making the decision it points to, in the form of arrows with non-continuous lines (⇢).

  Again these arrows form a DAG.

An ID is a translation of the spaces in the decision problem into a graphical structure. The basic structure of an ID combines the three types of nodes as shown in Fig. 4.3(a): the utility

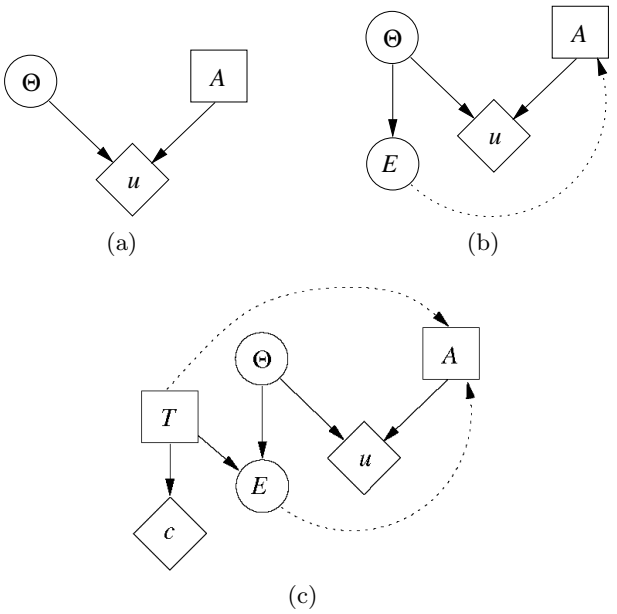


Figure 4.3: *Generic structures of influence diagrams (IDs). Arrows with continuous lines (⟶) represent functional relationships between the nodes, and arrows with non-continuous lines (⇢) represent precedence links for decision nodes to indicate that the parent node is instantiated when making the decision in the child decision node.* **(a)** *The general structure of an ID for a one-stage decision problem, combining the actions (node A), the states of nature (node Θ), and the utilities (node u).* **(b)** *The combination of the ID shown in (a) with the model in Fig. 4.1(b), producing an ID for choosing an action in A given a new piece of information (node E).* **(c)** *Extending the ID in (b) to include the preliminary test decision (node T) and the cost of performing this test (node c) produces the generic structure of an ID for a two-stage decision problem (Korb and Nicholson, 2011). Note that these IDs presented here with a utility function in node u have identical structures when the decision-maker defines her preferences in terms of losses: in this case, a diamond-shaped node called l containing this loss function replaces node u.*

or loss of a consequence (◇) depends on the action (□) and on the state of nature (○). This is the generic structure for solving a one-stage decision problem. If a piece of information has updated the decision-maker's probability distribution over Θ (the state of nature), then the extended influence diagram shown in Fig. 4.3(b) can take into account this information (modeled in node E). If obtaining this information is the object of a preliminary decision, then the addition of a test decision node and a utility node containing the cost of obtaining this information produces the generic influence diagram structure for a two-stage decision problem (Fig. 4.3(c)) (Korb and Nicholson, 2011).

The main advantage of an ID is its practical capacity of performing probabilistic and decision-theoretic calculations in a complex decision problem (Shachter, 1986; Howard and Matheson, 2005; Jensen and Nielsen, 2007). With this capacity they have successfully met the needs of several fields faced with important problems of decision making.

4.2.1 Practical implementations

The practical application of IDs ranges from medicine to environmental sciences[4] and industry[5]. In particular, the domain of medical decision analysis concerned with finding the optimal treatment by balancing the costs and the benefits of different possible treatments

[4]For example, decisions concerning the seeding of an approaching hurricane Howard et al. (1972), the collection of data for assessing nuclear waste (Heger and Hill, 1993), and the treatment of a stream (Heger and White, 1997).

[5]For example, finding the optimal maintenance schedule for the components of a production system (Vatn et al., 1996).

(e.g., Table 4.1) has accepted a decision-theoretic approach as a valuable tool for rational decision making:

> The strength of decision-theoretic ideas in medicine is that of providing a structure and an end to the process of gathering, organizing, and integrating the quantitative information that is relevant to a decision. In this sense, despite the limitations of its axiomatic foundations, and the cognitive difficulties with representing and communicating values and utilities to the general public, decision-oriented quantification is an almost indispensable component of good medical decision making. (Parmigiani, 2002, p. 51)

This domain sees IDs as a useful aid for problems with complex probabilistic relationships (Owens et al., 1997), and values in particular the IDs' capacity of compactly representing complex decisions and highlighting probabilistic relationships (Nease and Owens, 1997). In addition, the IDs' capability of combining qualitative, causal knowledge with quantitative, probabilistic knowledge gives them an advantage over a rule-based classification approach (Lucas et al., 2000).

Table 4.1 lists only a few examples of medical decision problems which have been modeled and analyzed through influence diagrams. These are all concerned with applying a particular treatment or surgical procedure on a patient with a particular set of symptoms and/or diagnostic test results. They model the decision problem according to the generic structure in Fig. 4.3(c), by defining:

- action space $\mathcal{A}$: the possible treatments (e.g., perform surgery or not perform surgery);
- state space Θ: the patient's current health status (e.g., having, or not having, a particular disease), and/or any other inherent trait of the patient that will influence the outcome of the treatment (e.g., having a genetic disposition that makes the patient responsive or non-responsive to the treatment in question);
- test space $\mathcal{T}$: the possible diagnostic tests;
- space of possible test results $\mathcal{E}$: the results of diagnostic tests;
- consequences $\mathcal{C}$: the patient's possible future health histories (e.g., excellent health, minor stroke, major stroke, death).

The trickiest aspect in a medical decision problem is defining the utility (or loss) function. In this decision problem, the decision-maker is the patient or a person, such as a physician, acting on behalf of the patient's interests. The probabilities and utilities therefore represent the patient's probabilities and utilities. The utility values represent the patient's benefit, in terms of her health. This benefit is measured by the patient's **quality-adjusted life expectancy (QALE)** in years. For each of the possible consequences, this value quantifies the period of time in full health that the patient considers equivalant to a year in the health state of the consequence (Parmigiani, 2002). The most rational decision is the one that maximizes the QALE.

The construction of influence diagrams in medicine and other domains reveal the capability of influence diagrams to deal with complex decision problems.

Table 4.1: *Some examples where IDs model medical decision problems.*

Decision problems	
Test and treat an infant born to a mother infected with HIV	(Owens et al. (1997))
Perform a bacterial culture and treat with antibiotics a patient with a sore throat	(Nease and Owens (1997))
Perform diagnostic tests and thoracotomy on a patient with a lung-cancer tumor	(Nease and Owens (1997))
Perform a tonsillectomy (surgical procedure) on a patient with a sore throat	(Renooij and Gaag (1998))
Perform surgery and treat with chemotherapy and/or radiotherapy a patient with primary gastric non-Hodgkin lymphoma	(Lucas et al. (1998))
Select antibiotic drugs for a patient with signs of pneumonia in the intensive-care unit	(Lucas et al. (2000))

Chapter 5

Current State of the Art of Decision-theoretic Models in Forensic Science

Some papers have addressed forensic decision problems from a normative point of view. This chapter presents the major publications with regard to the questions of performing or not performing a test, how many tests to perform, what tests to perform, and what conclusion to draw. Note that in forensic science, the word "test" should be understood as the application of forensic examination and/or analytical methods.

5.1 To perform or not to perform a test

5.1.1 Case pre-assessment (Cook et al., 1998a)

In the 1990's, the UK's Forensic Science Service (FSS)'s business-like approach to managing forensic science casework[1] led to the desire of improving the cost-effectiveness of forensic examinations (Blakey, 1995; Cook et al., 1998a). Until then, the forensic scientists at the FSS exhaustively performed all possible tests on every item that they received. To improve cost-effectiveness, the FSS proposed to think about the test's expected results *before* performing that test: they called this preliminary stage of reflection **case pre-assessment** (Cook et al., 1998a). This stage was part of a model whose aim was to better meet the needs of the customer (i.e., the policeman or agency asking for the expertise in forensic science) by implementing an approach that represents the value of scientific evidence by the Bayes factor (Cook et al., 1998a,b, 1999; Evett et al., 2000b,c).

For each proposition of interest to the customer, case pre-assessment formalizes what results the scientist expects to obtain from a particular test given the case's circumstances, and given that that proposition is true. Expressing these expectations in terms of probabilities, case pre-assessment produces a probability distribution over the possible likelihood ratio values for each proposition given that that proposition is true.Case pre-assessment therefore introduces transparency on the expected contribution of a forensic examination or analytical method for discriminating between two propositions. Cook et al. (1998a) described these expectations as

> pav[ing] the way for sound decision making during the service delivery phase (p. 153).

[1] The Home Office's FSS became an executive agency in 1991. One of the consequences of this transition was that the FSS began to manage its budget separately, so that its income and expenses became transparent. As a result, the costs of its services also became transparent to the policemen asking for forensic science expertise, because they were now charged directly for the work they asked for (Cook et al., 1998a).

However, *how* the decisions were to be made based on these expectations was not described in this model. A probability distribution over the expected results is a key ingredient for making rational decisions, but as Taroni et al. (2005) pointed out,

> it does not offer clear criteria for the decision as to whether or not to perform a test (p. 895).

The decision does not only depend on the results expected from the test, but also on what the decision-maker gains (or loses) from these results.

5.1.2 Decision analysis (Taroni et al., 2005)

Taroni et al. (2005) filled this gap by combining the probability distributions over the possible results with a utility function over these possible results. The combination of these two produces the expected utility which provides the decision-making criteria of expected utility maximization (Section 3.5).

Taroni et al. (2005) illustrated this idea by applying it to the decision of performing or not performing a DNA analysis in contested kinship scenarios (the first row of Table 5.1 summarizes the decision-theoretic framework used). They modeled the posterior probability of kinship as the decision problem's state of nature, where 'posterior' refers to the probability of kinship after having made the decision and observed the DNA analysis's results if the analysis was performed. According to this definition, the decision's possible consequences are the possible posterior probabilities of kinship if the test is performed, and a consequence of *status quo* (i.e., my knowledge on the two propositions of interest remains as it is) if the test is not performed. To create discrete states for the first case, Taroni et al. (2005) used *Hummel's scale*, a scale commonly used for legal kinship decisions, to divide the range of possible posterior probabilities into six intervals. This produced seven possible consequences: six for performing the analysis and obtaining a posterior probability in one of the intervals, and one for not performing the analysis and not obtaining a posterior probability. By defining a utility function over these seven consequences, the authors evaluated the expected utilities of performing and not performing a DNA analysis for pairwise combinations of the propositions of full sibship, half-sibship and unrelated individuals for different prior probabilities of kinship. They then applied the expected utility maximization criteria to justify performing or not performing a DNA analysis.

The major contribution of this paper was to recognize that forensic scientists make decisions, and that a normative approach to these decisions requires the application of decision theory and the use of the expected utility maximization criteria.

5.2 How many tests?

5.2.1 Using utilities to measure the added value expected from an additional test (Taroni et al., 2007)

When the scientist decides to perform a DNA analysis, one of the follow-up questions is "How many markers to type?". Taroni et al. (2007) proposed to answer this question by evaluating how much is gained by an additional marker in terms of utility. For this, they defined the number of markers as the possible actions, and the value of the scientific evidence (i.e., the likelihood ratio) obtained from typing these markers as the state of nature (the second row in Table 5.1 summarizes the decision-theoretic framework used). To obtain a set of discrete states of nature, the authors defined nine likelihood ratio intervals, denoted here Θ_j, $j = 1, 2, \ldots, 9$. They considered this decision in the context of a criminal case with a biological trace that needs to be compared with a suspect's sample, and evaluated likelihood ratios for the following two pairs of propositions:

H_p - *the trace comes from the suspect,*

Table 5.1: *One-stage decision problems*

Reference	Decision	Action space $\mathcal{A}$	State space Θ
Taroni et al. (2005)	Perform or not perform a DNA analysis?	a_1 = perform test a_2 = not perform test	intervals of the posterior probability of kinship (according to *Hummel's scale*): $\Theta_1 = (0.9979, 1]$ *practically proven* $\Theta_2 = [0.9910, 0.9979]$ *extremely likely* $\Theta_3 = [0.9500, 0.9910)$ *very likely* $\Theta_4 = [0.9000, 0.9500)$ *likely* $\Theta_5 = [0.8000, 0.9000)$ *undecided* $\Theta_6 = [0, 0.8000)$ *not useful*
Taroni et al. (2007)	How many markers to type?	a_1 = type 1 marker a_2 = type 2 markers a_3 = type 3 markers a_4 = type 4 markers a_5 = type 5 markers a_6 = type 6 markers $\vdots = \vdots$ a_{15} = type 15 markers	intervals of the likelihood ratio obtained from typing the markers: $\Theta_1 = [0, 1)$ $\Theta_2 = [1, 10)$ $\Theta_3 = [10, 10^2)$ $\vdots = \vdots$ $\Theta_8 = [10^6, 10^7)$ $\Theta_9 = [10^7, \infty)$
Biedermann et al. (2008a)	What to conclude?	a_1 = individualization a_2 = inconclusive a_3 = exclusion	Θ_1 = the trace and the reference sample come from the same source Θ_2 = the trace and the reference sample come from different sources
Taroni et al. (2006b)	Consider two cases as linked?	source level: a_1 = state that the traces come from the same source a_2 = state that the traces come from different sources crime level: a_1 = state that the same perpetrator committed the crimes a_2 = state that different perpetrators committed the crimes	source level: Θ_1 = the traces come from the same source Θ_2 = the traces come from different sources crime level: Θ_1 = the same perpetrator committed the crimes Θ_2 = different perpetrators committed the crimes
Taroni et al. (2010)	What to conclude?	concluding statements	values or intervals of values of the parameter of interest

H_{d1} - the trace does not come from the suspect, it comes from an individual unrelated to the suspect;

and

H_p - the trace comes from the suspect
H_{d2} - the trace does not come from the suspect, it comes from a full sibling of the suspect.

In this framework, the probability of obtaining a likelihood ratio in interval Θ_j is a function of the number of markers typed, and of whether the trace actually comes from the suspect:

$$Pr(\Theta_j|a_i) = Pr(\Theta_j|a_i, H_p)Pr(H_p) + Pr(\Theta_j|a_i, H_{dk})Pr(H_{dk}) \ , \quad k \in \{1, 2\} \ .$$

The authors then defined utility functions over the likelihood ratio intervals. Because the utilities vary according to the decision-maker's objectives, they defined three different utility functions to represent the objectives of three different actors in the trial: the prosecutor with a high utility for high likelihood ratios, the defense lawyer with a high utility mostly for state Θ_1 (a likelihood ratio between 0 and 1), and the judge with a high utility for both extremes of the likelihood ratio scale (since he is interested in obtaining a powerful likelihood ratio, but does not care in which direction it points). The combination of each of these individual's utilities with $Pr(\Theta_j|a_i)$, for all $i = 1, 2, \ldots, 15$ and $j = 1, 2, \ldots, 9$, produced that individual's expected utilities $\bar{u}(a_i)$, for all $i = 1, 2, \ldots, 15$. The authors then evaluated the $EVOI$ of an additional marker as:

$$\bar{u}(a_i|\cdot) - \bar{u}(a_{i-1}|\cdot) \ , \quad \text{for all } i = 2, 3, \ldots, 15 \ ,$$

and found that an additional marker's relative $EVOI$ in percentage reduced to less than 0.05% after the 7th marker for the prosecutor's and judge's utility functions, and after the 8th marker for the defense lawyer's utility function.

The major contribution of this paper is to highlight that typing a maximum number of markers is not necessarily a rational choice from the point of view of maximizing the expected utility. From a decision-theoretic perspective, the optimal number of markers to type depends on the decision-maker's objectives, which differ among the actors in a criminal trial, so that there is no single optimal number. Hence, this paper illustrates how the optimal course of action is subjective (in the sense of personal) because it reflects the decision-maker's goals and preferences.

5.2.2 Using the mutual information to measure the added value expected from an additional test (Mazumder, 2010)

Mazumder (2010) answered the question of "How many markers to type?" by modeling this decision as a two-stage decision problem (the first row in Table 5.2 summarizes the decision-theoretic framework used). In this framework, the decision "To type or not to type a marker?" is a preliminary decision for the terminal decision "What is the probability distribution to report for the propositions of interest in the case?". The DNA typing results are the test results obtained after making the preliminary decision. The state of nature is the true proposition, and the state space consists of the different propositions formulated in the case. Thus, in this framework, the action space of the terminal decision is $\mathcal{A} = [0, 1]$, where $a_i \in \mathcal{A}$ is the probability of proposition $\Theta_j \in \Theta$ to report:

$$a_i = Pr(\Theta_j) \ .$$

Mazumder (2010) developed a criterion based on the mutual information between the state of nature Θ_j and the test result E_k^{th} to quantify the information gained from the results produced by a particular test. The rational decision-maker chooses the test that maximizes the mutual information. The optimal test is therefore defined as the test that provides the most information on the state of nature.

Table 5.2: *Two-stage decision problems*

Reference	Decision	Action space $\mathcal{A}$	State space Θ	Test space $\mathcal{T}$	Evidence space $\mathcal{E}$ (test results)
Mazumder (2010)	How many markers to type?	report $Pr(\Theta_j \vert E_k^{t_h})$ for all $\Theta_j \in \Theta$	propositions in the case	number of markers to type	DNA typing result(s)
Mazumder (2010)	Which markers to type?	report $Pr(\Theta_j \vert E_k^{t_h})$ for all $\Theta_j \in \Theta$	propositions in the case	markers	DNA typing result(s)
Mazumder (2010)	Which individuals to type?	report $Pr(\Theta_j \vert E_k^{t_h})$ for all $\Theta_j \in \Theta$	propositions in the case	individuals to type	DNA typing result(s)
Taroni et al. (2010) Biedermann et al. (2012a)	How many items to sample?	a_1 = state that $\theta \leq \theta^*$ a_2 = state that $\theta > \theta^*$	Θ_1: $\theta \leq \theta^*$ Θ_2: $\theta > \theta^*$	number of items to sample	number of items found to be positive
Shen et al. (2006)	What investigative action to undertake?	report $Pr(\Theta_j \vert E_k^{t_h})$ for all $\Theta_j \in \Theta$	propositions in the case	investigative actions	results of investigative actions

From an information-theoretic point of view, the mutual information is defined as the difference between the unconditional entropy of Θ and the conditional entropy of Θ given $E_k^{t_h}$:

$$H(\Theta) - H(\Theta|E_k^{t_h}) \ ,$$

where $H(\cdot)$ denotes the entropy.[2]

From a decision-theoretic point of view, this value is equal to the $EVOI(t_h)$ of the test result when the utility function is defined as the logarithmic score (Mazumder, 2010):

$$\begin{aligned} u(a_i, \Theta_j) &= -ln\,[a_i] \\ &= -ln\,[Pr(\Theta_j)] \ . \end{aligned}$$

In this context, t_h denotes typing a number h of markers. Hence Mazumder (2010) evaluated the expected added value of typing an additional marker as the difference between the mutual information of typing $h+1$ markers and the mutual information of typing h markers. When this value is equal to 0, or is defined by the decision-maker as being "sufficiently small", it is not necessary to type an additional marker, because typing the additional $h+1$'th marker would lead to an expected utility that is equal to, or sufficiently close to, the expected utility obtained after typing h markers, producing an $EVOI$ equal to, or very close to, 0.

With this approach, Mazumder (2010) evaluated the optimal number of markers to type for disputed paternity, victim identification and kinship determination cases, with a "sufficiently small"-threshold set to 0.05%. Again, these results indicate that it is not always necessary to type all available markers in order to obtain "sufficient" information for discriminating between the propositions of interest.

The major contributions of this study is to show how to model this decision as a two-stage decision problem, and how using a logarithmic score as the utility function provides a means to evaluate the reduction in uncertainty over Θ (i.e., the true proposition) produced by typing each additional marker.

5.2.3 Sampling scenarios (Taroni et al., 2010; Biedermann et al., 2012a)

In sampling scenarios, the question of "How many tests to perform?" takes the meaning of "How many items of a consignment to examine or analyze?". The objective of sampling items in a consignment is to make a statement about the proportion of items that have a particular trait. In a forensic context, the trait of interest is usually something illegal, such as a drug (Aitken, 1999; Aitken and Taroni, 2004). In this scenario, the forensic scientist must make a statement on whether the proportion of items in the consignment containing this illegal substance is greater than a particular threshold (Taroni et al., 2010), for example to determine whether a case is a high-profile case requiring priority treatment by an examining magistrate or investigative unit (Biedermann et al., 2012a). However, a laboratory's limited resources in terms of money and time, or particular risks related to the testing of the items, usually do not allow the scientist to test every item, so that the scientist faces the task of choosing a sample size.

In forensic science, there have been several probabilistic approaches to this question (e.g., Frank et al., 1991; Tzidony and Ravreby, 1992; Aitken, 1999; Biedermann et al., 2008b) until Taroni et al. (2010) proposed a decision-theoretic approach (the fourth row in Table 5.2

[2]The entropy is an information-theoretic measure of the uncertainty on a random variable. The entropy of a random variable X, with generic elements X_z, is equal to:

$$H(X) = -\sum_{X_z \in X} Pr(X_z)\ log[Pr(X_z)] \ ,$$

where the *log* can be for a base of 2 (this is the most common and measures the information contents in bits), a base of e (measuring the information in nats), or a base of 10 (measuring the information in hartleys) (Lathi, 1998: pp. 682–683).

summarizes the decision-theoretic framework used). Taroni et al. (2010) and Biedermann et al. (2012a) modeled this problem as a two-stage decision problem according to the theory described in Section 3.6.1: the choice of the sample size is a preliminary decision for the terminal decision of stating whether the proportion of items in the consignment containing the illegal substance, denoted by θ, is greater than the threshold value, denoted by θ^*. Defining the possible states of nature as $\theta \leq \theta^*$ and $\theta > \theta^*$, the terminal decision's possible consequences consist of correct statements and incorrect statements. Taroni et al. (2010) and Biedermann et al. (2012a) defined a symetric loss function on a monetary scale, with the loss of a correct statement equal to 0€ and the loss of an incorrect statement equal to 100,000€. The monetary scale allows the formulae to take into account the sampling cost for making the decision on the optimal sample size. To describe the scientist's uncertainty on the true proportion, they used a Beta distribution, following Aitken (1999)'s suggestion for large consignments. Taroni et al. (2010) present equations for solving fixed sample size and sequential analysis decision problems, and Biedermann et al. (2012a) works through an example of a fixed sample size decision problem to show how to determine the optimal sample size.

5.3 What tests?

5.3.1 Which markers to type (Lauritzen and Mazumder, 2008) (Mazumder, 2010)

The results of Mazumder (2010)'s study do not only indicate how many markers to type (Section 5.2.2), but also which markers to type. For this, the mutual information between the state of nature and the test result provides a meaure of the informativeness of an individual marker (Lauritzen and Mazumder, 2008). Mazumder (2010) evaluated the relative informativeness,

$$\frac{H(\Theta) - H(\Theta|E_k^{t_h})}{H(\Theta)} \ , \tag{5.1}$$

of individual genetic markers currently used in forensic genetics. Here, t_h refers to typing marker h, and the higher this value, the more information the marker provides on the state of nature.

Mazumder (2010) evaluated the informativeness of genetic markers for disputed paternity and criminal case scenarios, to produce a rank ordering of the markers from the most informative to the least informative. Her study on how many markers to type (Section 5.2.2) minimized the number markers for reaching a sufficiency threshold by taking the markers in this order, that is, by always taking the most informative of the remaining available markers.

Sensitivity analyses studied the effects of taking into account the possibility of mutations and of using population data from different ethnic groups. The results indicate that taking into account mutations slightly decreases the informativeness of the markers (Lauritzen and Mazumder, 2008; Mazumder, 2010), and the rank ordering of the markers depends on the population's relative allele frequencies (Mazumder, 2010).

5.3.2 Which individuals to type (Mazumder, 2010)

A further extension of Mazumder (2010)'s study examined which individuals to type for disputed paternity, victim identification and kinship determination cases when the DNA profiles of the individuals of interest are unavailable. This study used the mutual information criterion (Section 5.2.2), with t_h representing the individual(s) to type. With the help of BNs (Section 4.1) for making inferences about the propagation of the genetic information in the pedigree, Mazumder (2010) determined how much information the DNA profiles of different subsets of individuals in the pedigree provide on the state of nature. The results of

this study indicate which subset of individuals contributes the most information, and also highlight that it is useless to type additional relatives that have the same genetic information as an already typed individual.

5.3.3 What investigative action to undertake? (Shen et al., 2006)

Shen et al. (2006) present a decision support system for optimizing the investigator's strategy for which additional examinations and enquiries to carry out during an investigation. Based on the initial set of observations and traces collected in the case, this system first uses Bayesian networks to model plausible scenarios and the traces and observations each scenario leads to. These possible scenarios form the space Θ of the propositions of interest in the case. In a second step, the system applies an information-theoretic approach to evaluate the expected posterior entropy of Θ given the evidence collected by each possible investigative action $t_h \in \mathcal{T}$:

$$H(\Theta|E_k^{t_h}) = -\sum_{\Theta_j \in \Theta} Pr(\Theta_j|E_k^{t_h})\ log_2[Pr(\Theta_j|E_k^{t_h})]\ .$$

It uses the Bayesian networks to assign $Pr(\Theta_j|E_k^{t_h})$. The goal is to find the investigative action that provides the most information for discriminating between the scenarios in Θ, which means that the optimal investigative action is the t_h that minimizes the expected posterior entropy $H(\Theta|E_k^{t_h})$.[3]

5.4 What is the conclusion?

5.4.1 Individualizations and exclusions (Biedermann et al., 2008a)

To evaluate the value of the evidence with regard to a pair of source level propositions (Cook et al., 1998b), the forensic scientist compares a trace's characteristics with those of a reference sample from a potential source. We saw in Section 2.3 that the forensic scientist should represent the value of these observations in the form of a Bayes factor. However, in many forensic science laboratories, it is still common practice to *individualize* or to *exclude* a potential source as the trace's origin. This is problematic in a strictly probabilistic framework, because probabilities only allow for making inferences on a common source, but never an **individualization** defined here as formally excluding all other sources. More than twenty years ago, Stoney (1991a) described this passage from comparing the trace's features to concluding an individualization as:

> somewhat analogous to a leap of faith. It is a jump, an extrapolation, based on the observation of highly variable traits among a few characteristics, and then considering the case of many characteristics. Duplication is inconceivable to the rational mind and we conclude that there is absolute identity. (p. 198)

Ten years later, Phillips et al. (2001) recognized that concluding an individualization or an exclusion in forensic science is a decision. They explained that this decision depends on the forensic scientist's probability distribution over the pair of source level propositions, and on her utilities associated with correct and false conclusions (i.e., a correct individualization, a correct exclusion, a false individualization and a false exclusion). Later, Whitman and Koppl (2010) presented similar discussions for offence level propositions.

[3] Another approach based on the expected posterior entropy uses the **empirical cross-entropy** (ECE) and ECE plots to represent the informational gain expected from a forensic analysis and the analysis of the data it produces (Ramos Castro, 2007; Ramos and Gonzalez-Rodriguez, 2008). The ECE is the prior-weighted average of the logarithmic scoring rule over different cases. Its plot provides a visual representation of a method's expected informational gain in the inferential process that allows the scientist to compare different methods with each other.

Biedermann et al. (2008a) formalized these ideas by presenting them in a decision-theoretic framework. They defined the scientist's three possible conclusions of *individualization*, *inconclusive* and *exclusion* as the possible actions, and whether or not the potential source is the trace's true source as the state of nature (the third row in Table 5.1 summarizes the decision-theoretic framework used). Combining the possible actions with the possible states of nature produces a total of six possible consequences: a correct individualization, a false individualization, an inconclusive conclusion when the trace and the reference sample share a common origin, an inconclusive conclusion when the trace and reference sample do not share a common origin, a false exclusion and a correct exclusion. Among these, Biedermann et al. (2008a) considered the accurate conclusions of a correct individualization and a correct exclusion as the best possible consequences, and assigned a utility value of 1 to each of these. They considered a false individualization as the worst possible consequence, and assigned a utility value of 0 to this consequence. They then considered both consequences of an inconclusive conclusion as equally desirable, with a utility value between the two extremes, which they denoted $\alpha \in (0,1)$. And finally, they considered the consequence of a false exclusion as less desirable than an inconclusive conclusion, and assigned a utility value of $\beta \in (0,1)$, $\beta < \alpha$ to this consequence. Based on these utilities, the conclusions' expected utilities are:

$$\bar{u}(a_1) = Pr(\Theta_1) \ ,$$

$$\bar{u}(a_2) = \alpha \ , \text{ and}$$

$$\bar{u}(a_3) = \beta \ Pr(\Theta_1) + Pr(\Theta_2) \ .$$

For a given α and β, the Bayes action depends on the probabilities $Pr(\Theta_1)$ and $Pr(\Theta_2)$. Since $Pr(\Theta_2) = 1 - Pr(\Theta_1)$, Biedermann et al. (2008a) plotted the three expected utilities in function of $Pr(\Theta_1)$. This graph shows the threshold value(s) of $Pr(\Theta_1)$ that define(s) the decision strategy as the intersection(s) of the functions that maximize the expected utility. For a small $Pr(\Theta_1)$, the Bayes action is to exclude; for a large $Pr(\Theta_1)$, the Bayes action is to individualize. For a $Pr(\Theta_1)$ close to 0.5, the Bayes action could be any of the conclusions, depending on the values assigned to α and β.

This study emphasizes the distinction between inferences and decisions, and illustrates how decision making is the logical step following probabilistic inferences.

5.4.2 Case linkage (Taroni et al., 2006b)

When a forensic scientist cannot associate a trace with a potential source, she may compare this trace with traces from other cases. This comparison may lead to an inference on two traces having a common source. Or, if both traces are considered to have been left by the perpetrator, then this inference is on the two cases having been committed by the same individual.

Taroni et al. (2006b) recognized that stating that two traces have a common source, or that two cases were committed by the same person is a decision. These statements make up the action space, and whether or not the two traces come from the same person, or whether or not the same person committed the two crimes, are the possible states of nature (the fourth row in Table 5.1 summarizes the decision-theoretic framework used). The consequences of this decision is either a correct or an incorrect linkage, or either a correct or an incorrect non-linkage. For these, Taroni et al. (2006b) defined a utility function with values of 10 for correct statements, and values of 0 for incorrect statements.

The major contribution of this study is its emphasis on the difference between the inference made on the state of nature and the decision of making a concluding statement about

it. It explains how making the decision is only possible by taking into account the utilities or losses associated with the possible consequences, and how making a decision about the uncertain event does not change the probability distribution over the state of nature.

5.4.3 Data analysis (Taroni et al., 2010)

How many red wool fibres were transferred onto the car seat? What is the proportion of a toner's resin group? How many GSR particles are on the hands of an individual who discharged a firearm 3 hours ago? What is an individual's alcohol concentration in the blood? Is it greater than the legal threshold? What is the width of the landmarks on a bullet recovered on a crime scene? What is the height of an individual figuring on images of a surveillance camera? What is an allele's mutation rate? Were the seized banknotes involved in drug dealing or do they come from the general circulation? Based on the number of CMS counted in the landmarks of two bullets, were these fired by the same gun or from different guns?

The answer to each of these questions is a statement: "54 red wool fibres were transferred onto the car seat", "the toner contains 2% polystyrene", etc. Taroni et al. (2010) model making these statements as one-stage decision problems (the fifth row in Table 5.1 summarizes the decision-theoretic framework used). A forensic scientist observes, counts and measures, to try to answer each of the above questions, yet making a final statement does not depend on the collected data alone. As a decision, it combines the inferences made from the collected data with the prior probability distributions and a loss function (Taroni et al., 2010). Each of these statements is therefore a conclusion, in the same sense as stating an individualization or an exclusion (Section 5.4.1). The action space consists of the possible statements. The state of nature is the true value of the parameter of interest, with the state space being either discrete or continuous, depending on the parameter. The loss function describes how undesirable the decision-maker considers the different possible false conclusions.

The major contribution of this book is to highlight how each conclusion, in every forensic domain, is a decision. It does not only highlight this point, but presents all of the underlying Bayesian tools necessary to implement this decision-theoretic framework in practice.

5.5 Graphical models

Aitken and Gammerman (1989) recognized that BNs could be useful in forensic science applications. Edwards (1991 - 1992), Schum (1994) and Kadane and Schum (1996) described the potential of probabilistic models for reasoning about evidence in real cases, and Dawid and Evett (1997) its potential for handling complicated dependence relationships among different items of evidence. Since then, numerous publications have described the application of BNs and OOBNs to the interpretation of scientific evidence (Taroni et al., 2004, 2006a; Biedermann, 2007), in particular for the evaluation of DNA evidence (Evett et al., 2002; Aitken et al., 2003; Dawid et al., 2002, 2007; Mortera et al., 2003; Cowell et al., 2007b, 2008; Hepler and Weir, 2008; Vicard et al., 2008; Green and Mortera, 2009; Biedermann et al., 2012b),[4] evidence collected in fire debris (Biedermann et al., 2005a,b), firearm evidence (Biedermann and Taroni, 2006; Biedermann et al., 2009a), fibre evidence (Garbolino and Taroni, 2002), black toner analyses (Biedermann et al., 2009b, 2011b) and the combination of multiple items of evidence (Hepler et al., 2007; Juchli et al., 2012).

The use of IDs in forensic science applications has remained rather scarce. Taroni et al. (2005) and Biedermann et al. (2008a) used the generic 3-node structure shown in Fig. 4.3(a) for their one-stage decision problems. Taroni et al. (2010) explained how to extend this ID to include forensic observations and analytical results in a structure of the type shown in

[4]See Biedermann and Taroni (2012) for a more complete list of references on the application of BNs and OOBNs to inference problems related to DNA evidence.

Fig. 4.3(b). Biedermann et al. (2012a) extended this generic structure to take into account the variables in a sampling problem. And finally, Taroni et al. (2007)'s decision-theoretic framework translates into the ID structure shown in Fig. 5.1(a), and Taroni et al. (2006b) presented the ID in Fig. 5.1(b) for their decision problem.

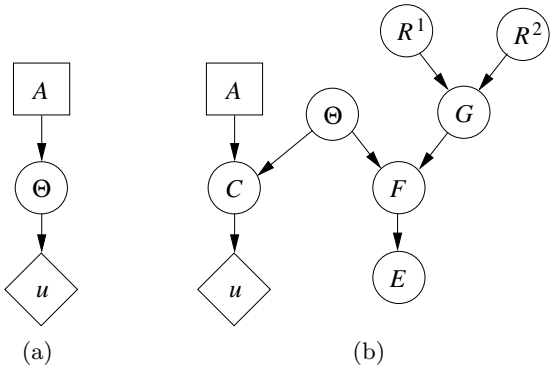


Figure 5.1: *Two IDs presented in forensic science literature.* **(a)** *ID presented by Taroni et al. (2007) to answer the question of how many markers to type (Section 5.2.1), where A represents the number of markers to type, Θ the intervals of the obtained likelihood ratio, and u the utility function defined over the likelihood ratio intervals.* **(b)** *ID presented by Taroni et al. (2006b) to answer the question of whether to link two cases (Section 5.4.2), where A is the decision of linking or not linking two cases, Θ the pair of propositions on whether or not the two cases are truly linked (i.e., offense level propositions), C whether or not the decision made in A is correct, u the utilites of a correct and an incorrect decision, E the observed evidence in the two cases, F a pair of propositions on whether the evidence in the two cases share a common source (i.e., source level propositions), R^1 whether the evidence in the first case is relevant, R^2 whether the evidence in the second case is relevant, and G whether the evidence in neither, one or both cases is relevant.*

5.6 Final remarks

This chapter presented the application of decision theory to several isolated decision problems encountered by forensic scientists. These applications demonstrate that it is possible to apply decision theory to forensic decision problems, yet these applications each focus on only a single, very particular problem. Compared with each other, these applications propose to apply decision theory in many different ways. That is, they structure the decision problem differently from one case to another: they have modeled the states of nature as being the posterior probability of the proposition of interest (Taroni et al., 2005), the likelihood ratio obtained from a particular analysis (Taroni et al., 2007), and the proposition of interest (Taroni et al., 2006b; Shen et al., 2006; Biedermann et al., 2008a; Mazumder, 2010; Taroni et al., 2010; Biedermann et al., 2012a). They have modeled the same decision as a one-stage decision problem (Taroni et al., 2007) and as a two-stage decision problem (Mazumder, 2010). Hence, they do not provide a uniform approach for handling forensic decision problems. As a result, there exists no uniform framework that can be applied to the broad range of decision problems encountered by forensic scientists.

Part II

Practical Aspects

Chapter 6

Thesis

Research studies have shown that there is a need for forensic scientists to make more consistent choices (Section 1.1). The serious consequences that the decisions made by forensic scientists can have in the administration of justice reinforce this need, and create a demand for forensic scientists to be able to justify their choices. **The aim of this thesis is to respond to this need and demand by presenting a normative framework for decision problems encountered by scientists in forensic science laboratories.**

Decision theory guarantees coherent decision making (Chapter 3). Statisticians and legal scholars have shown that Bayesian inference and decision theory provides an adequate normative model for the legal fact-finding process (Sections 2.3 and 2.4). Other domains such as medicine, where practitioners face similar inference and decision problems as in forensic science, use models implementing Bayesian inference and decision theory as a valuable tool for making and justifying their choices (Section 4.2.1). Forensic scientists have recognized the potential of Bayesian inference for organizing forensic inference problems (Section 2.3), yet have, for the most part, stopped there. **Logically, the next step is for forensic scientists to extend Bayesian inference models to decision theory for organizing forensic decision problems.**

Sofar, there have already been a few studies which applied a decision-theoretic approach to a forensic decision problem (Chapter 5). However, these have each focused on very particular decision problems. They do not provide a uniform framework for the broad range of decision problems that face forensic scientists, nor a framework that applies to all the different categories of trace evidence. **The objective of this thesis is to present a *uniform* normative framework that applies to decision problems encountered for different categories of trace evidence.**

However, a practical application of Bayesian inference and decision theory involves mathematical expressions that become increasingly complex and less transparent for non-statisticians when they take into account an increasing number of non-independent variables. Many decision problems in forensic science require numerous, non-independent variables for making the necessary inferences on the state of nature. For forensic inference problems, Bayesian networks (BNs) provide a tool for organizing the variables in a problem through a visual representation (Sections 4.1 and 5.5). In medicine, studies have extended BNs to influence diagrams (IDs) for modeling complex decision problems (Section 4.2.1). **If BNs provide a means for handling complex forensic inference problems, then this thesis proposes that IDs can provide a tool to help forensic scientists apply decision theory to forensic decision problems.**

To attain this objective, this thesis consists of three propositions on how to handle forensic inference and decision problems. These advocate the use of decision theory (Proposition 1), BNs (Proposition 2) and IDs (Proposition 3). Theoretical case studies applying decision theory, BNs and IDs to forensic inference and decision problems will address each of these propositions. The rest of this chapter describes each proposition, the objectives of the case

studies that will try to support each proposition, and the scenarios of the inference and decision problems chosen for these studies.

6.1 Proposition 1

Decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence.

This proposition puts forth that decision theory applies to:

(i) one-stage and two-stage decision problems encountered by scientists in forensic science laboratories; and to

(ii) decisions made in the context of different categories of trace evidence.

Decision theory organizes a decision problem by defining the action space and the state space for one-stage decision problems, and the action space, state space, test space and space of possible test results for two-stage decision problems. This organization of the decision problem produces a space of consequences over which it is possible to define a utility or loss function. Once these spaces have been defined, decision theory provides the decision strategy.

Addressing this proposition requires an application of decision theory to one-stage and two-stage decision problems in scenarios involving different categories of trace evidence in forensic science. Three case studies will devote themselves to this task, of which two will focus on one-stage decision problems, and one on a two-stage decision problem. The one-stage decision problems consist of the following scenarios:

- **What to conclude from a database search producing a single hit?**
  When a trace has no suspected source, it is common to search for this trace's intrinsic characteristics (i.e., the characteristics shared by the trace and its source) in a database containing a subset of the population of the trace's potential sources. This scenario considers a case where the results of the database search are binary: there is either a match or no match between the trace's profile and that of the reference samples in the database. This scenario also assumes that no errors are possible, neither in analyzing and observing the trace's characteristics, nor in producing the database search results of matches and no matches, and that the individuals registered in the database are all unrelated. The particular case of interest here is one where a database search among a number n_{db} of potential sources produces exactly 1 match and $n_{db} - 1$ non-matches with the trace's intrinsic characteristics. The decision question of interest is, what can the scientist conclude from these results? In particular, can the scientist individualize the source of the matching reference sample as the trace's source?

  Theoretically, this scenario applies to any category of trace evidence where database searches produce binary results. In forensic science and legal literature (e.g., Balding and Donnelly, 1996; Evett and Weir, 1998), though, this scenario is known as the **database search problem** and refers to DNA database searches. In this case, the trace's intrinsic characteristics is a DNA profile searched for in a database containing the reference profiles of n_{db} unrelated individuals in the population of potential sources. The above cited literature debated whether the strength of the evidence against an individual selected through a database search increased or decreased with regard to the probable cause case, where the trace's profile matches the profile of an individual the investigators associated with the case based on other information. The logical answer to this question is that the non-matches in the database provide additional information with regard to a probable cause case, therefore *increasing* the value of the evidence against the selected individual with regard to the probable cause case (e.g., Dawid, 2001). However, recent publications showed that this answer is still not understood by everyone today (Schneider et al., 2010; Fimmers et al., 2011): these publications not only insisted that the value of the evidence is smaller in the database search case than in the probable cause case, but also declared that the probability of a false individualization is greater in a database

search case. Biedermann et al. (2011a) responded to the first point. This case study will focus on the second. Following in the footsteps of the literature on the database search problem, it will therefore also discuss this problem in the context of a DNA database search.

- **What genotype to search for in a database of DNA profiles?**
  This scenario focuses specifically on DNA profiling results and DNA database searches, because DNA database searches have become a popular investigative tool for searching for possible sources of biological traces. A DNA database search requires the investigator to specify the genotype of the DNA profile she wants to search for in the database. For example, for single-donor traces where DNA profiling is not subject to stochastic effects, she will specify a homozygous designation for an alllele when she observes a single allelic peak at that locus, and a heterozygous designation for a pair of alleles when she observes two allelic peaks at that locus. However, for single-donor DNA traces producing **low-template-DNA** (**lt-DNA** for short) profiles (Gill and Buckleton, 2010), stochastic effects of **allele drop-out** and **allele drop-in** may cause the resulting electropherogram (EPG) to contain peaks for a different allelic configuration than that of its donor (e.g., Taberlet et al., 1996; Gill et al., 2000; Whitaker et al., 2001): if allele drop-out occurs, an allele in the donor's profile fails to appear in the trace's profile, and if allele drop-in occurs, a peak for an allele not present in the donor's profile appears in the trace's profile. As a result, observing a single allelic peak in an lt-DNA profile may be the result of either a homozygous donor or a heterozygous donor with one allele drop-out. Similarly, observing a pair of allelic peaks may be the result of either a heterozygous donor or a homozygous donor with one allele drop-in. The designation of the donor's genotype is therefore not straightforward, and this raises the question of what genotype to search for in the database for the configuration of allelic peaks observed on a trace's EPG.

  There have been several pragmatic approaches for answering this question. For a locus with only a single peak, one approach consists of comparing this peak's height (or area) with a predetermined threshold (Buckleton et al., 2005; Gill et al., 2009): if the peak height is greater than the threshold, the locus's genotype is designated as a homozygote, and if the peak height is smaller than the threshold, the genotype is designated as the allele of the observed peak paired with a wildcard F representing any allele at that locus (Gill et al., 2000). For a locus with one or several peaks, another approach consists of obtaining several EPGs for the trace and retaining a consensus profile formed by alleles observed a particular number of times over all of the replicates (Taberlet et al., 1996). This case study will focus on the decision of choosing among the possible designations, including the wildcard F, based on the results of one or more replicates.

The objectives of applying decision theory to these one-stage decision problems are:

- to determine the Bayes action in a particular case;
- to understand which parameters in the model affect the Bayes action, and how these parameters affect the Bayes action;
- to establish under what circumstances a particular action is the Bayes action; and
- to state the decision strategy.

Given that the scenarios for the one-stage decision problem involve DNA profiling results, even though the first applies to all domains where a database search produces binary results, the scenario for the two-stage decision problem is placed in the context of a different category of trace evidence. This scenario considers the following decision problem in the domain of fingerprints:

- **Process a fingermark?**
  For each detected fingermark, a fingerprint examiner must decide whether to submit the recovered trace to a fingermark examination process, which seeks to associate the mark with its source (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011). This fingermark examination process compares a submitted fingermark with the fingerprints of potential sources, which may produce data in the form of similarities or

dissimilarities with reference prints. This data may support a proposition that the mark and a print come from the same finger, with regard to an alternative proposition that the mark and the print come from two different fingers, and/or lead to exclusions of possible sources. The objective of processing the fingermark is therefore to produce data that will narrow down the pool of possible sources to a smaller population containing the mark's true source. However, the fingermark examination process uses the laboratory's resources, which are limited. This means that if a submitted fingermark does not produce any such discriminatory data, or leads to an erroneous conclusion, the laboratory will have spent its resources in vain. In addition, an erroneous conclusion may lead to an erroneous verdict. Furthermore, a fingermark expected to lead to a strong association with a print of its true source, does not necessarily represent *useful information* in the context of the fingermark's case (for example, if other traces and information in the case have already established an association between the crime scene or evidentiary item and the person who left the mark, or if the mark comes from a victim whose presence on the crime scene is not disputed). To examine the usefulness of fingermarks not processed by a forensic science laboratory, Neumann et al. (2011b) conducted a field study that compared the additional data obtained from these marks with their processing cost. Using this idea as a starting point, this case study will focus on the decision of processing or not processing a fingermark given the context of its case and the laboratory's cost for processing the mark.

The objectives of applying decision theory to this two-stage decision problem are:

- to establish the expected value of information $EVOI(t_1)$ of processing a fingermark;
- to understand which parameters in the model affect the $EVOI(t_1)$; and
- to state the decision strategy.

Supporting this first research proposition will provide us with a theoretical framework for forensic decision problems. However, when these decision problems contain an increased number of variables and probabilistic relationships (e.g., for making the necessary inferences on the state of nature), the application of a decision-theoretic framework involves complex computations that are difficult to implement in practice using an algebraic approach. To prevent the application of the proposed framework from being limited to simple problems, Propositions 2 and 3 advocate the use of graphical models for handling complex inference and decision problems in forensic science. First, Proposition 2 proposes BNs for modeling complex inference problems, and then Proposition 3 proposes to use their decision-theoretic extension, IDs, as a tool to help implement a decision-theoretic framework for complex forensic decision problems.

6.2 Proposition 2

Bayesian networks (BNs) provide a means for handling new complex inference problems encountered by forensic scientists.

This proposition puts forth the use of BNs as a tool to help forensic scientists reason for inference problems involving numerous variables and numerous dependence relationships among these variables. These graphical models should provide an approach that overcomes the hurdle of complexity that limits the application of a purely algebraic approach to, rather, simple inference problems. For this, we need to show that:

- BNs provide a mathematically robust and transparent approach for taking into account numerous variables and their dependence relationships in the computation of the values of interest;
- the results produced using BNs are in agreement with the equations currently proposed in forensic science literature;

- BNs provide a means for going beyond these algebraic equations by allowing the user to easily incorporate additional variables into the computation of the values of interest and relax the existing equations' assumptions; and
- BNs provide a transparent approach for extending less complex models for forensic inference problems to more complex models.

To address this proposition, a case study of the so-called **two-trace problem** will focus on these points. The two-trace problem is a forensic inference problem that has puzzled, and continues to puzzle, many forensic scientists. It consists of the following scenario:

- **The two-trace problem**
  Introduced by Evett (1987), this problem describes a scenario where investigators recover two items of the same category of trace evidence (e.g., two bloodstains) on a crime scene, and a laboratory analysis performed on these traces, reveals that they possess different intrinsic traits (e.g., the blood group or DNA profile). Later there is a single suspect whose sample matches one of the two traces. For a particular pair of propositions (where one proposition is advanced by the prosecution and the other by the defence), the question is: "How strong is the evidence resulting from these two comparisons in favor of the prosecution's or the defence's proposition?". The answer to this question depends on the formulation of the two propositions. Assuming that no analytical errors are possible, Evett (1987) and Meester and Sjerps (2003) algebraically derived the Bayes factor for source level propositions.[1] Evett (1987) derived a Bayes factor for a pair of propositions that takes into account the observation of both traces:

  proposition 1: *at least one of the crime stains comes from the suspect*,
  proposition 2: *neither of the crime stains comes from the suspect*.

  Meester and Sjerps (2003) showed that the Bayes factor is different when one or both of the propositions takes into account the observation of only one of the traces, that is:

  proposition 1: *crime stain 1 comes from the suspect*,
  proposition 2: *neither of the crime stains comes from the suspect*;

  or

  proposition 1: *crime stain 1 comes from the suspect*,
  proposition 2: *crime stain 1 does not come from the suspect*;

  and they thought that it is confusing to present the Bayes factor for a two-trace problem.[2] For the activity level propositions:

  proposition 1: *the suspect is one of the two men who were in contact with the victim during the offense*,
  proposition 2: *two other men were in contact with the victim during the offense*;

  Triggs and Buckleton (2003) proposed an algebraic approach to arrive at a mathematically more complex solution incorporating activity level parameters concerning the transfer and background presence of trace material (Evett, 1984). With the different

[1] Evett (1987) actually derived the Bayes factor for the offense level propositions:

proposition 1: *the suspect was one of the two men who committed the crime*,
proposition 2: *the suspect was not one of the two men who committed the crime*;

assuming both traces to be relevant. An offense level evaluation with maximal relevance produces the same Bayes factor as a source level evaluation (e.g., Aitken and Taroni, 2004).

[2] Actually, obtaining different Bayes factors for different pairs of propositions is a natural consequence of applying the laws of probability for evaluating the Bayes factor for different pairs of propositions. Rather than being a confusing inconvenience, this should be seen as an example of the flexibility of the value of the evidence to adapt to a particular formulation of a pair of propositions (see Chapter 9).

expressions for a source level Bayes factor, and the more extensive algebraic development to arrive at an activity level Bayes factor, the two-trace problem is a perplexing inference problem.

The objectives of this case study are:

- to construct a BN for the different pairs of source level propositions advanced in forensic science literature;
- to extend this BN to activity level propositions;
- to deduce the mathematical expressions for the values of interest from the BN, and compare them with the algebraically-derived equations in forensic science literature;
- to identify and relax the simplifying assumptions made by the algebraically-derived equations; and
- to extend the BN developed for two traces to a general situation involving more than two traces.

Supporting this proposition will provide us with a means for handling complex forensic inference problems by using graphical models. In view of applying decision theory to forensic decision problems, this result would provide a means for making inferences on the state of nature when these inferences involve numerous, non-independent variables. Thus, if graphical models in the form of Bayesian networks provide a useful tool for handling complex inference problems in forensic science, then Proposition 3 proposes that their extension to IDs can help forensic scientists in handling complex decision problems.

6.3 Proposition 3

Influence diagrams (IDs) provide forensic scientists with a practical tool for structuring and providing the elements required for making coherent choices in forensic decision problems.

This proposition puts forth the use of IDs as a tool to help forensic scientists make coherent choices in the face of uncertainty. If decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence (Proposition 1), and BNs provide a means for handling complex inference problems encountered by forensic scientists (Proposition 2), then IDs can provide a practical tool for coherent decision-making in forensic science by:

(i) incorporating the advantages offered by graphical models (such as BNs) for visually structuring and performing probabilistic calculations in complex inference problems; and

(ii) combining the inferences with utility (or loss) functions to compute the expected utilities (or losses) for each of the possible actions to reach a decision.

The implementation of decision theory in the software currently available for constructing IDs varies slightly from one program to another. The IDs in this thesis were constructed using the software *Hugin Researcher*, versions 7.0 to 7.6, by Hugin Expert A/S. From a computational point of view, this program computes the expected utilies (or losses) for each of the possible actions of a terminal decision, the maximum expected utility, and the $EVOI(t_h)$, $h = 0, 1, \ldots, p$, and contains algorithms for finding the $\underset{h}{arg\ max}\ ENG(t_h)$ and computing the maximum expected utility minus the cost for performing this test. However, it does not contain any algorithm for evaluating the $max_h\ ENG(t_h)$, which is, for example, available in the software *GeNIe 2.0* by the Decision Systems Laboratory of the University of Pittsburgh, another software currently available for constructing IDs. To distinguish between an ID's formal properties and the different levels of computational implementations in ID-constructing software, this proposition does not pretend that IDs will solve the decision problem for the scientist, but only that it provides the required elements for solving the decision problem.

Addressing this proposition requires the construction of IDs for forensic decision problems. For this, the three case studies presented for Proposition 1 (Section 6.1) will each be extended to include the construction of an ID. This will produce two IDs for one-stage decision problems and one ID for a two-stage decision problem. The objectives of constructing these IDs are:

- to provide a visual representation of each decision model's underlying assumptions; and
- to propose a probabilistic and decision-theoretic model that implements decision theory in each decision problem to provide the elements needed for making a coherent choice.

To clarify how to use IDs for one-stage and two-stage decision problems, Chapter 7 will provide additional explanations on how the IDs in the case studies provide the elements required for making coherent choices.

Chapter 7

Results and Discussion

The results take the form of individual research papers presenting theoretical case studies of the scenarios presented in Chapter 6. These individual research papers are in Part III. The present chapter summarizes the results of these studies in view of the three propositions described in the previous chapter, and explains how to confront one-stage and two-stage decision problems in forensic science. First, Section 7.1 discusses the use of graphical models (notably BNs) for handling complex forensic inference problems in forensic science. This section addresses Proposition 2. Then, Sections 7.2 and 7.3 discuss one-stage and two-stage decision problems, respectively, to show how decision theory organizes these decision problems, and how to use IDs as a tool for implementing this framework. These sections address Propositions 1 and 3.

7.1 Graphical models

Chapters 9 and 10 model the two-trace problem using BNs to address Proposition 2—*Bayesian networks provide a means for handling new complex inference problems encountered by forensic scientists.*

Chapter 9 presents a BN that combines all the different pairs of source level propositions proposed by Meester and Sjerps (2003) as separate nodes in a single model. Through its graphical representation, it highlights the logical relationships between the different propositions and the analytical results. Since there are different logical relationships between the different propositions and the analytical results, different pairs of propositions inevitably lead to different Bayes factors. The BN's visual representation of the problem provides transparency and clarity on why this is so. In addition, its logical decomposition of the problem brings to light the assumptions underlying the proposed Bayes factors, and provides an expression for the Bayes factor that relaxes these assumptions.

Chapter 10 extends this model to activity level propositions, and to more than two traces using OOBNs. To overcome the hurdle of complexity that limits the application of an algebraic approach to developing the Bayes factor for increasingly complex inference problems, this study proposes to use BNs for assigning the Bayes factor. For this, it shows how the Bayes factor resulting from a graphical approach modeling the two-trace problem at the activity level is in agreement with the previously published, algebraically-derived expressions. Again, this graphical approach provides transparency on the assumptions that underlie the algebraically-derived expression, here notably the activity level Bayes factor published by Triggs and Buckleton (2003). The extension to more than two traces using OOBNs illustrates the BNs' capacity to handle an increased number of variables and probabilistic relationships.

These studies lead to the following conclusions:

- BNs offer transparency to understand what assumptions lie behind algebraic expressions published in forensic science literature and allow the user to go beyond these

equations by relaxing these assumptions.

- A BN's properties of rigorously handling probabilistic calculations in a mathematically robust environment, decomposing complicated events into a set of distinct variables, and describing and visualizing the assumed dependencies among the variables provide a structured and logical means for modeling complex inference problems in forensic science, and for using these models to compute the probabilistic values of interest.
- A BN's flexible and transparent architecture allows the user to easily incorporate additional variables into existing models, and to coherently combine and structure different aspects of a problem as separate objects in distinct hierarchical levels of an OOBN.
- A BN's graphical architecture allows a model to be inserted as part of a larger network, allowing the scientist to address more complex inference problems.

According to these conclusions, a graphical approach using BNs can handle complex inference problems encountered by forensic scientists. If decision theory provides an adequate framework for organizing forensic decision problems, then IDs could prove to be practical tools for handling complex decision problems encountered by forensic scientists.

7.2 One-stage decision problems

A one-stage decision problem describes any decision problem where the decision-maker's satisfaction depends directly on the decision's consequence. That is, the decision-maker is capable of specifying utilities or losses for the decision's possible consequences. Modeling a one-stage decision problem requires the decision-maker to define the action space, the state space, and a utility or loss function over the actions' possible consequences (Sections 3.2 and 3.5).

In the context of forensic science, this model applies to any decision where the scientist formulates a conclusion. The word "conclusion" is used here in a very broad sense, so that it includes all intermediate conclusions, or statements, that the scientist makes during the entire process of analyzing, comparing and evaluating her observations. Because, as Savage (1951) explains:

> all problems of statistics, including those of inference, are problems of action, for to utter or publish any statement is, after all, to take certain action. (p. 55)

Many forensic procedures consist of making a sequence of such statements. For example, the interpretation of a DNA profile is a process that goes from a set of signals on an electropherogram (EPG) to the genotype of the contributor, by stating (Butler, 2013):

1. for each signal, whether this signal is a peak or instrument noise;
2. for each peak, whether this peak indicates the presence of an allele or is a stutter;
3. for all loci with single peaks, whether the single peak is a homozygous genotype or a heterozygous genotype with allele drop-out; and
4. for all loci with two peaks, whether the two peaks are a heterozygous genotype or do not come from the same source.

The interpretation of a mixture follows similar steps (Gill et al., 2006). From a decision-theoretic point of view, as explained by Savage (1951), each of these steps consists of making a decision. Each of these is a one-stage decision problem.

This idea is not new. There are currently a number of empirically-derived thresholds that tell the forensic scientist how to make these decisions. For example, for the above decisions

in DNA interpretation, these thresholds take the form of peak heights, peak areas,[1] or ratios of peak heights or areas.[2] These existing thresholds are based on empirical approaches (e.g., Gill et al., 1997, 2009). They use empirical data to determine what threshold leads to, for example, 95% of correct choices, or 99% of correct choices. However, there is nothing that justifies why a laboratory should use the 95%-level threshold rather than the 99%-level threshold, or vice-versa. This is because these thresholds do not take into account the utility of a correct choice, or the loss of an incorrect choice.

A decision-theoretic approach produces thresholds that take into account both the probabilistic aspects of the inference problem and the preferential aspects of the decision problem. Given a probability model and a utility or loss function, this approach produces a decision strategy for choosing the Bayes action. Sensitivity analyses on the probability model's parameters illustrates the parameters' impact on the Bayes action. Sensitivity analyses on the utility (or loss) function shows how the threshold(s) vary in function of the decision-maker's objectives and preferences.

Chapters 11 and 12 illustrate this idea for the database search problem and the genotype designation decision, respectively, to address Proposition 1—*Decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence.*

7.2.1 Organizing the decision problem

Decision theory provides a way of organizing one-stage decision problems by breaking them down into an action space and a state space. In the context of the forensic decision problems discussed above, the action space is the exhaustive list of possible conclusions that the scientist can draw (i.e., statements that the scientist can make), the state of nature is the true condition of whatever the decision is concluding about and which is unknown, therefore the state space is all of the possible conditions. For the database search problem (Chapter 11) and the genotype designation decision (Chapter 12), Table 7.1 presents the action and state spaces.

In the database search problem (first row in Table 7.1), the state of nature is the crime stain's origin. The state space is represented as a population of potential sources of N unrelated individuals, numbered from 1 to N. However, for the scenario in question, we are not actually interested in all of the 1 to N individuals: we are interested in only one of these individuals, the one whose sample matches the trace's profile to produce the hit during the database search. It suffices therefore to represent this space with regard to a single individual. For any individual J, $J \in \{1, 2, \ldots, N\}$, this space consists of:

Θ_J: *the crime stain comes from individual* J,
$\neg\Theta_J = \{\Theta_1, \ldots, \Theta_{J-1}, \Theta_{J+1}, \ldots, \Theta_N\}$: *the crime stain comes from someone else, unrelated to individual* J.

As for the action space, it contains the actions of individualizing each of the N individuals in the population of potential sources as the trace's source, and the action of not individualizing anybody as the trace's source. Again, for the scenario in question, it suffices to represent this space with regard to a single individual. For individual J, $J \in \{1, 2, \ldots, N\}$, this action space consists of:

a_J: *individualize individual* J *as the source of the trace recovered on the crime scene*,
$\neg a_J = \{a_1, \ldots, a_{J-1}, a_{J+1}, \ldots, a_N, a_{N+1}\}$: *not individualize individual* J *as the source of the trace recovered on the crime scene.*

Summarizing the state and the action spaces in this way produces a space of possible consequences of only four elements. These are: a correct individualization, an incorrect indi-

[1] For example, 50 rfu for the limit of detection threshold, and 150 rfu for the stochastic threshold (Gill et al., 2009).

[2] For example the heterozygote balance of 60% (Gill et al., 1997).

Table 7.1: *One-stage decision problems*

Reference	Decision	Action space $\mathcal{A}$	State space Θ
Gittelson et al. (2012b), Chapter 11	What to conclude from a database search producing a single hit?	for a population of potential sources of N unrelated individuals, numbered from 1 to N, the action space consists of $m = N + 1$ elements: a_1 = individualize individual 1 as the source of the trace recovered on the crime scene a_2 = individualize individual 2 as the source of the trace recovered on the crime scene $\vdots \quad = \quad \vdots$ a_N = individualize individual N as the source of the trace recovered on the crime scene a_{N+1} = not individualize anybody in the population of N individuals as the source of the trace recovered on the crime scene	for a population of potential sources of N unrelated individuals, numbered from 1 to N, the state space consists of $n = N$ elements: Θ_1 = the crime stain comes from individual 1 Θ_2 = the crime stain comes from individual 2 $\vdots \quad = \quad \vdots$ Θ_N = the crime stain comes from individual N
Gittelson et al. (2013b), Chapter 12	What genotype to search for in a database of DNA profiles?	possible genotype designations (for a locus $\mathcal{X}$'s alleles $x_1, x_2, \ldots, x_s$): $a_1 = \{x_1, x_1\}$ $a_2 = \{x_1, x_2\}$ $\vdots \quad = \quad \vdots$ $a_s = \{x_1, x_s\}$ $a_{s+1} = \{x_1, F\}$ $a_{s+2} = \{x_2, x_2\}$ $\vdots \quad = \quad \vdots$ $a_{m-1} = \{x_s, F\}$ $a_m = \{F, F\}$	the donor's possible genotypes (for a locus $\mathcal{X}$'s alleles $x_1, x_2, \ldots, x_s$): $\Theta_1 = \{x_1, x_1\}$ $\Theta_2 = \{x_1, x_2\}$ $\vdots \quad = \quad \vdots$ $\Theta_s = \{x_1, x_s\}$ $\Theta_{s+1} = \{x_2, x_2\}$ $\vdots \quad = \quad \vdots$ $\Theta_n = \{x_s, x_s\}$

vidualization, a correct non-individualization and an incorrect non-individualization (i.e., missing an individualization).

In the genotype designation decision (second row in Table 7.1), the state of nature is the genotype of the trace's donor. For a particular locus, the state space consists of all of the donor's possible genotypes for that locus. Here, to coherently represent this scenario, it is not possible to summarize the state space in a smaller number of subsets of Θ. The action space consists of all of the possible designations for this genotype. This includes the possibility of using a wildcard F for either one or both alleles: in the former case, the designation specifies only one of the two alleles, and in the latter case the designation does not specify any of the alleles, and is entirely uninformative. Combining these two spaces produces a very large space of possible consequences, yet it is here that it is possible to divide these consequences into four subsets: correct designations without using F (when the two designated alleles are the donor's alleles), correct designations using F (when the one designated allele is one of the donor's two alleles), incorrect designations (whenever at least one of the designated alleles is not one of the donor's alleles), and the uninformative designation of type $\{F, F\}$.

In both of these decision problems with large state and action spaces, it was therefore possible to organize the problem so as to have a small size of possible consequences over which the decision-maker defines a utility or loss function. In these decision problems, the decisions' possible consequences describe how correct the conclusion is. It is common to use loss functions to describe the decision-maker's preferences among these consequences, because a decision-maker can usually quantify the penalties associated with false conclusions more easily than the gain associated with correct conclusions. Hence, on a scale from 0 to 1, correct conclusions have minimum losses of 0, the most serious incorrect conclusion(s) has (have) a loss of 1, and less serious incorrect conclusions, or non-informative conclusions, have a loss with a value between 0 an 1, determined by the decison-maker's preferences among gambles (Section 3.3). In the two case studies presented here, the greek letters λ and η denote the intermediate loss values for the database search problem and the genotype designation decision, respectively. The two loss functions are:

Chapter 11:

What to conclude from a database search producing a single hit?

	loss
correct conclusion	0
missing an individualization (i.e., incorrect non-individualization)	λ
incorrect individualization	1

Chapter 12:

What genotype to search for in a database of DNA profiles?

	loss
correct designation	0
correct designation using F	η_1
uninformative designation (i.e., $\{F, F\}$)	η_2
incorrect designation	1

The objective of organizing the decision problem in this framework is to obtain a decision strategy. The decision strategy is the result of combining this organization with the aim of minimizing the expected loss.

When the expected losses of all the actions in the action space can be represented as a function of a single probability, the decision strategy reduces to comparing this probability with a threshold defined by the loss function. This is the case for the database search problem. For example, for a database search in a database containing individuals 1 to n_{db}, $n_{db} \leq N$, and producing a single hit with individual J^*, $J^* \in \{1, 2, \ldots, n_{db}\}$, the expected losses for actions a_{J^*} and $\neg a_{J^*}$ are a function of the probability that the crime stain comes from individual J^*, denoted by

$$Pr(\Theta_{J^*} | E^{t_{J^*}}, \neg(E^{t_1}, \ldots, E^{t_{J^*-1}}, E^{t_{J^*+1}}, \ldots, E^{t_{n_{db}}}), I),$$

with I representing the information associating individual J^* to the crime scene prior to the database search, and E^{t_h} a match between the crime stain's profile and individual h,

$h = 1, 2, \ldots, n_{db}$.[3] The decision strategy reduces to individualizing individual J^* if and only if:

$$Pr(\Theta_{J^*}|E^{t_{J^*}}, \neg(E^{t_1}, \ldots, E^{t_{J^*-1}}, E^{t_{J^*+1}}, \ldots, E^{t_{n_{db}}}), I) > \frac{1}{1+\lambda} \ .$$

This result is analogous to the decision strategy produced by the decision-theoretic analysis for convicting a defendant in the legal factfinding process (Eq. (2.5)), since the λ in this equation is by definition (Eq. (11.16) in Chapter 11):

$$\lambda = \frac{\textit{loss of missing an individualization}}{\textit{loss of an incorrect individualization}} \ .$$

This decision strategy highlights the following points:

- Changing λ produces a different threshold to which the decision-maker compares the probability of Θ_{J^*} to make the decision (Figures 11.1 and 11.7 in Chapter 11). The Bayes action therefore varies for decision-makers with different values for λ. However, this threshold is independent of the circumstances and probabilities in a case. It is the same for a probable cause case and for a database search case, providing a coherent approach to both scenarios. Since the probability of a false individualization depends on the decision threshold defined by the decision-maker's loss function, it is the decision-maker who sets the maximum probability of a false individualization, and (again) not whether the case is a probable cause or a database search case.

- Whether or not the probability of Θ_{J^*} is greater than this threshold depends on the set of parameters that the probability model uses to determine the probability of Θ_{J^*}. In this probability model, the probability of Θ_{J^*} is:

  $$Pr(\Theta_{J^*}|E^{t_{J^*}}, \neg(E^{t_1}, \ldots, E^{t_{J^*-1}}, E^{t_{J^*+1}}, \ldots, E^{t_{n_{db}}}), I) = \frac{1}{1+p_1-p_1p_2} \ ,$$

  with
  p_1 = the expected number of matches in the population of potential sources, and
  p_2 = the proportion of the population of potential sources registered in the database.

  Figure 11.6 in Chapter 11 plots $Pr(\Theta_{J^*}|E^{t_{J^*}}, \neg(E^{t_1}, \ldots, E^{t_{J^*-1}}, E^{t_{J^*+1}}, \ldots, E^{t_{n_{db}}}), I)$ in function of these two parameters, for $p_1 = N\gamma$ (where γ represents the match probability of the crime stain's intrinsic characteristics in the population of potential sources) and $p_2 = \frac{n_{db}}{N}$.

This case study concludes that a decision-theoretic approach is applicable to the decision of individualizing a trace's potential source in both the probable cause case and the database search case.

When the expected losses of all the actions in the action space cannot be represented as a function of a single probability, then the decision strategy does not reduce to comparing a single probability with a single threshold. This is the case for the genotype designation decision problem. In this case, one cannot restrain the action space to two possible actions as in the database search case. One must consider at least three possible designations (a homozygote, a designation using F, and an uninformative designation) for every case where the EPG shows at least one allelic peak for the locus in question. There are therefore three expected losses to be compared with each other for a given loss function and probability model. These expected losses are functions of several probabilities that are not mutually exclusive, so that one cannot represent these comparisons by a single equation describing the threshold in terms of the comparison of one probability with a value defined by the loss function. However, one can compute threshold values of a parameter of interest for a given

[3] Note that E^{t_h}, $h = 1, 2, \ldots, n_{db}$, is a Boolean variable here, so that E^{t_h} represents a match with individual h and $\neg E^{t_h}$ a non-match with individual h.

loss function and other model parameters. Hence, Tables 12.1 and 12.2 in Chapter 12 present threshold values in terms of the average peak height in the EPG. In addition, sensitivity analyses allow one to study the behavior of the expected losses in function of a parameter of interest. These analyses show how the threshold depends on the model's parameters and its underlying assumptions. Figures 12.2-12.7 in Chapter 12 plot the expected losses in function of the average peak height in the EPG. They show how the threshold depends on the observed alleles, the locus and, of course, the average peak height in each EPG. In this case study, the decision-theoretic approach shows that there is no single, universal threshold for deciding between any two designations. The Bayes action depends on parameters that vary from one case to another, from one scientist to another, and from one laboratory to another. The decision model provides a case-specific approach, that can be adapted to each laboratory's and scientist's needs and objectives.

According to these results, the framework provided by decision theory provides an adequate and useful structure for organizing one-stage forensic decision problems. However, forensic decisions such as the genotype designation decision contain numerous variables with complex probabilistic relationships. For these, the next section discusses the application of influence diagrams to these decision problems.

7.2.2 Influence diagram

To overcome the difficulty of evaluating the expected losses using an algebraic approach when there are numerous dependent variables, IDs offer a graphical approach for evaluating the expected losses. Chapters 11 and 12 present IDs for modeling the decision problem. In this sense, they address Proposition 3—*Influence diagrams provide forensic scientists with a practical tool for structuring and providing the elements required for making coherent choices in forensic decision problems.*

The IDs constructed in these studies illustrate the advantages provided by graphical models to visually represent a model's underlying assumptions, as already discussed in Section 7.1 for BNs. This section presents how to use an ID as a tool for solving a one-stage decision problem. For this, we will use the generic model shown in Fig. 4.3(b). The IDs in Chapters 11 and 12 have the same underlying structure as this model, they are just more complex because they involve more observed evidence nodes of type E, and more parameters describing the relationships between these observations and the state of nature.

Figure 7.1 applies this generic model to a fictive decision problem of choosing between actions a_1 and a_2 based on the observation $E_1^{t_1}$. In this example, the state of nature is one of

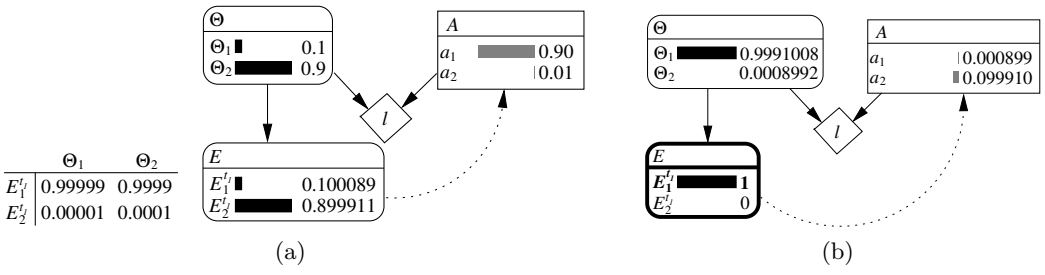


Figure 7.1: *Expanded representations of the generic ID shown in Fig. 4.3(b) for a one-stage decision problem. Node A contains the action space, node Θ the state space, node E the possible test results, and node l the loss function, which is here:* $l(C_{11}) = l(C_{22}) = 0$, $l(C_{12}) = 1$ *and* $l(C_{21}) = 0.1$. *Nodes Θ and E show probability distributions and node A the expected losses. Node l is not expanded here to focus the reader's attention on the values of interest in nodes A, E and Θ. The node with a thicker border represents an instantiated node, with the certain state in bold.* **(a)** *The uninstantiated ID, with the probability table for node E.* **(b)** *Instantiating node E to state* $E_1^{t_1}$ *produces the expected losses of* a_1 *and* a_2 *after observing* $E_1^{t_1}$.

two possible states, Θ_1 or Θ_2. Result $E_1^{t_1}$ is such that it supports state Θ_1: $Pr(E_1^{t_1}|\Theta_1, I) =$

0.99999 and $Pr(E_1^{t_1}|\Theta_2, I) = 10^{-4}$. A loss function of $l(C_{11}) = l(C_{22}) = 0$, $l(C_{12}) = 1$ and $l(C_{21}) = 0.1$ describes the decision-maker's preferences. In its expanded form, the ID indicates the probability distributions over Θ and E^{t_1}, and the expected losses of a_1 and a_2. In its uninstantiated form (Fig. 7.1(a)), these expected losses do not take into account the result $E_1^{t_1}$:

$$\begin{aligned}\bar{l}(a_1|I) &= l(C_{11})Pr(\Theta_1|I) + l(C_{12})Pr(\Theta_2|I) \\ &= 0 \times 0.1 + 1 \times 0.9 \\ &= 0.9 \ ,\end{aligned}$$

and

$$\begin{aligned}\bar{l}(a_2|I) &= l(C_{21})Pr(\Theta_1|I) + l(C_{22})Pr(\Theta_2|I) \\ &= 0.1 \times 0.1 + 0 \times 0.9 \\ &= 0.01 \ .\end{aligned}$$

To evaluate the expected losses for each of the possible actions given result $E_1^{t_1}$, the user instantiates node E^{t_1} to state $E_1^{t_1}$ (Fig. 7.1(b)). The expected losses indicated in the ID correspond to the following evaluations:

$$\begin{aligned}\bar{l}(a_1|E_1^{t_1}, I) &= l(C_{11})\frac{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)} \\ &\quad + l(C_{12})\frac{Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)} \\ &= 0 \times \frac{0.99999 \times 0.1}{0.99999 \times 0.1 + 10^{-4} \times 0.9} + 1 \times \frac{10^{-4} \times 0.9}{0.99999 \times 0.1 + 10^{-4} \times 0.9} \\ &= 0.000899 \ ,\end{aligned}$$

and

$$\begin{aligned}\bar{l}(a_2|E_1^{t_1}, I) &= l(C_{21})\frac{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)} \\ &\quad + l(C_{22})\frac{Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)} \\ &= 0.1 \times \frac{0.99999 \times 0.1}{0.99999 \times 0.1 + 10^{-4} \times 0.9} + 0 \times \frac{10^{-4} \times 0.9}{0.99999 \times 0.1 + 10^{-4} \times 0.9} \\ &= 0.099910 \ .\end{aligned}$$

A rational decision-maker then chooses the action with the minimum expected loss, in this case action a_1.

Note that both *Hugin Researcher 7.6* and *GeNIe 2.0*, two software systems currently available for constructing IDs, contain algorithms for computing the maximum expected utility.[4] To use this tool for degrees of satisfaction defined in terms of losses, it suffices to apply a linear transformation to represent the loss values on a scale of utilities (for example, by multiplying the losses by -1).

According to this illustrative example, IDs thus provide a practical tool for forensic scientists to structure one-stage decision problems and compute the expected losses (or utilities) required for finding the Bayes action.

[4] In *Hugin Researcher 7.6*, the user must select the decision node in the model's *Run Mode* and then open the *Value of Information* tool in the menu *Network* and submenu *Analysis*. In *GeNIe 2.0*, the decision node's value table automatically indicates the maximum expected utilities as bold values when the model is updated.

7.3 Two-stage decision problems

A two-stage decision problem describes any decision problem where the decision-maker's satisfaction does not depend on this decision's outcome, but on the consequence of a subsequent decision. Notably, two-stage decision problems are decisions on whether or not to invest in obtaining information that may increase the utility or reduce the loss incurred by making a subsequent decision. For the first decision, that is the preliminary decision, the decision-maker is incapable of specifying utilities or losses for the possible consequences, because her satisfaction depends on whether the information was *useful* for attaining another objective. Measuring this usefulness requires determining the $EVOI(t_h)$ or $ENG(t_h)$ of this preliminary decision's possible actions with regard to a terminal decision (Section 3.6). The terminal decision describes this other objective, for which the decision-maker is capable of specifying utilities or losses for the possible consequences. Modeling a two-stage decision problem therefore requires the decision-maker to define the preliminary and the terminal decisions. The preliminary decision consists of a test space and the space of the test's possible results, and the terminal decision of the action space, state space and utility or loss function over the actions' possible consequences.

In the context of forensic science, this model applies to any decision where the scientist invests money or time in obtaining information in view of making another decision. This terminal decision can be:

1. another laboratory decision of performing or not performing an analysis,
2. a laboratory "conclusion" decision discussed in Section 7.2, or
3. a subsequent decision made outside of the forensic science laboratory, such as the court of law's decision in the legal factfinding process for which the expertise in forensic science is destined.

An example of the first case would be a preliminary decision of whether to perform a presumptive test for blood where the terminal decision is to submit or not to submit the unknown substance for a DNA analysis. An example of the second case is a preliminary decision of whether to invest in an additional replicate in view of the terminal decision of designating the donor's genotype. The third case applies to all forensic analyses. The decision of processing any category of trace evidence in a forensic science laboratory is a preliminary decision in view of the case's judicial decision. Chapter 13 illustrates this idea for the decision of processing or not processing a fingermark. As a two-stage decision problem, and a decision problem in a category of trace evidence different from DNA, this case study addresses Proposition 1—*Decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence.*

7.3.1 Organizing the decision problem

Decision theory provides a way of organizing two-stage decision problems by defining an action space, state space, test space, and a space of possible test results. The test space consists of actions that provide some form of information that can be represented in the space of possible test results. The action and state space describe the terminal decision, for which the decision-maker defines the utility or loss function. Table 7.2 presents the definition of these spaces for the decision of processing a fingermark.

Chapter 13 models the decision of processing a fingermark as a preliminary decision for the terminal decision of convicting or acquitting a defendant. This case study presents the theoretical framework for modeling this decision problem. With an action space of convicting or acquitting the defendant, and a state space consisting of the defendant's guilt or innocence, there are four possible consequences: a correct conviction, a false conviction, a false acquittal and a correct acquittal. Defining the utility function over these four possible

Table 7.2: *Two-stage decision problem*

Reference	Decision	Action space $\mathcal{A}$	State space Θ	Test space $\mathcal{T}$	Evidence space $\mathcal{E}$[a]
Gittelson et al. (2013c), Chapter 13	Process a fingermark?	a_1 = convict the defendant a_2 = acquit the defendant	Θ_1 = the defendant is guilty of the alleged offense Θ_2 = the defendant is innocent of the alleged offense	t_0 = not process t_1 = process	$E_1^{t_0,1}$ = no result for the level 1 features $E_1^{t_1,1}$ = similarity of the level 1 features $E_2^{t_1,1}$ = dissimilarity of the level 1 features $E_1^{t_0,2}$ = no result for the level 2 features $E_1^{t_1,2}$ = similarity of the level 2 features $E_2^{t_1,2}$ = dissimilarity of the level 2 features

[a]Note that the evidence space $\mathcal{E}$ contains two levels of observations (indexed here with the superscripts 1 and 2). This means that the states $E_1^{t_0,1}$, $E_1^{t_1,1}$ and $E_2^{t_1,1}$ are not mutually exclusive with regard to the states $E_1^{t_0,2}$, $E_1^{t_1,2}$ and $E_2^{t_1,2}$. This is not a problem: an ID can model these two sets in two separate nodes (see Chapter 13).

consequences on a monetary scale to quantify the gains and losses of these consequences for the society, allows us to compare the $EVOI(t_h)$, $h = 0, 1$, with the cost of processing the fingermark: the decision strategy is to process the fingermark if $EVOI(t_1)$ is greater than or equal to the procesing cost. This case study presents the algebraic expression for the $EVOI(t_1)$. It illustrates how to model this two-stage decision problem in a decision-theoretic framework. According to these results, the decision-theoretic framework provides an adequate and useful structure for organizing this two-stage decision problem.

However, the developed expression for the $EVOI(t_1)$ contains numerous parameters and its use is not so transparent. For this reason, the next section explains how to use an ID modeling a two-stage decision problem to find the action that maximizes the $ENG(t_h)$. Note that, as in Section 7.2 for the one-stage decision problem, this explanation will use the generic model for the two-stage decision problem, and leave the more complex ID for the decision of processing a fingermark to be described in Chapter 13.

7.3.2 Influence diagram

In a two-stage decision problem, the rational decision-maker chooses the action in the test space $\mathcal{T}$ that maximizes the $ENG(t_h)$. IDs can serve as tools to help the decision-maker find this action in a complex decision problem.[5] Gittelson et al. (2013c) constructed an ID for a two-stage decision problem. This case study therefore also addresses Proposition 3—*Influence diagrams provide forensic scientists with a practical tool for structuring and providing the elements required for making coherent choices in forensic decision problems.* This section explains how to find the action that maximizes the $ENG(t_h)$ using an ID as a tool. For this, we will demonstrate how to use an ID with only the most basic computational algorithms implemented in the ID, that is, the propagation of information according to the laws of probability, and the computation of the expected utilities (losses) for the terminal decision's possible actions according to decision theory. In this demonstration, we will use the generic model presented in Fig. 4.3(c). The aim is to obtain the $ENG(t_h)$ for $h = 1, 2, \ldots, p$, and choose the action with the maximum $ENG(t_h)$.

Since the IDs presented in this thesis were constructed using *Hugin Researcher*, this section presents *Hugin* representations of the ID. The *Hugin* interface for an ID does not present the $ENG(t_h)$ values, but presents, instead, the values of its key components. Mathematically, the $ENG(t_h)$ is equal to:

$$ENG(t_h) = \sum_{k=1}^{q} max_i \underbrace{\left[\bar{u}(a_i|E_k^{t_h}, I) - c\right]}_{III.} \underbrace{Pr(E_k^{t_h}|I)}_{II.} - max_i \underbrace{\bar{u}(a_i|I)}_{I.} \tag{7.1}$$

for utilities, and to:

$$ENG(t_h) = min_i \underbrace{\bar{l}(a_i|I)}_{I.} - \sum_{k=1}^{q} min_i \underbrace{\left[\bar{l}(a_i|E_k^{t_h}, I) + c\right]}_{III.} \underbrace{Pr(E_k^{t_h}|I)}_{II.} \tag{7.2}$$

for losses. As the roman numerals in these equations indicate, these expressions are made up of three types of components:

I. the expected utilities or losses without the additional information;

II. the marginal probabilities of $E_k^{t_h}$, $k = 1, 2, \ldots, q$; and

III. the expected utilities or losses with an additional piece of information, substracting the cost of obtaining this information.

[5] Some software programs currently available for constructing IDs, such as *GeNIe 2.0*, include an algorithm for evaluating the $max_h\ ENG(t_h)$, but this is not always the case. *Hugin Researcher 7.6*, for example, will find the $\underset{h}{arg\,max}\ ENG(t_h)$, but not indicate the value of the $max_h\ ENG(t_h)$.

With the basic computational algorithms for computing probability distributions and expected utilities or losses, the ID can evaluate each of these.

Figure 7.2 illustrates this for a fictive example of deciding whether or not to perform a test t_1 to obtain additional information before choosing between actions a_1 and a_2. The actions

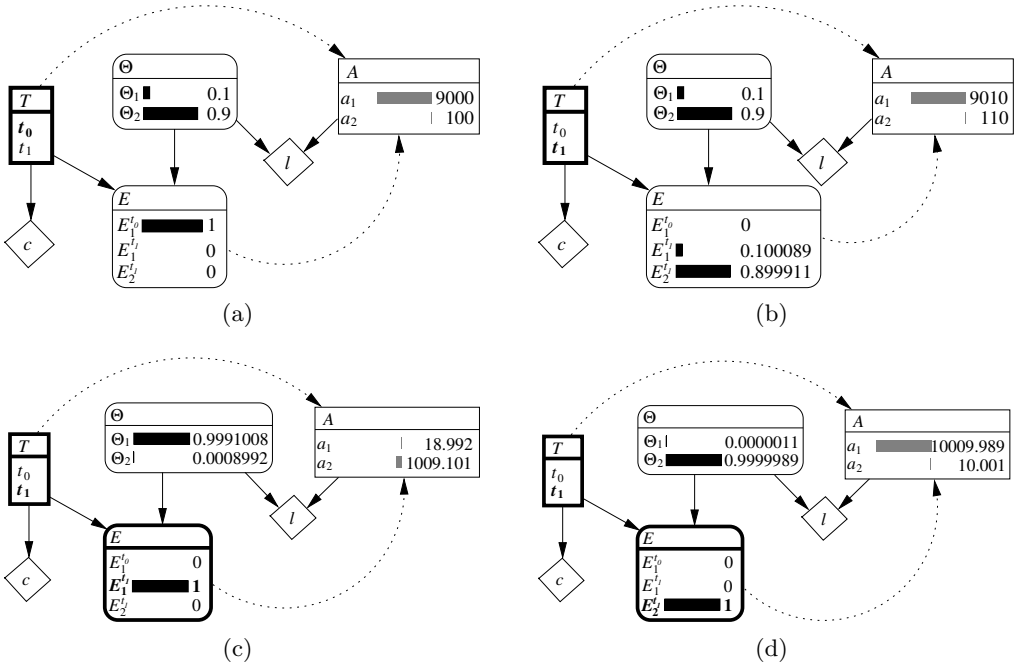


Figure 7.2: *Expanded representations of the generic ID shown in Fig. 4.3(c) for a two-stage decision problem. This model consists of the preliminary decision (node T), the terminal decision (node A), the space of possible test results (node E), the state space (node Θ), a loss function (node l) and the costs of the tests in T (node c). Here, these expanded representations illustrate how to use this ID to obtain the values needed for solving a two-stage decision problem for a fictive example. The preliminary decision is whether or not to perform a test t_1 (e.g., a forensic analysis) in view of the terminal decision of choosing between a_1 and a_2 (e.g., a factfinder's decision of convicting or acquitting a defendant). Performing this test costs 10 units and provides information on Θ (e.g., whether the defendant is guilty, Θ_1, or innocent, Θ_2, of the alleged offense): $E_1^{t_1}$ supports Θ_1 ($Pr(E_1^{t_1}|\Theta_1, I) = 0.99999$ and $Pr(E_1^{t_1}|\Theta_2, I) = 0.0001$), and $E_2^{t_1}$ supports Θ_2 ($Pr(E_2^{t_1}|\Theta_1, I) = 0.00001$ and $Pr(E_2^{t_1}|\Theta_2, I) = 0.9999$). The loss function over the terminal decision's four possible consequences is on the same scale as the cost of performing test t_1: $l(C_{11}) = l(C_{22}) = 0$ (e.g., a loss of 0 for convicting a guilty defendant and acquitting an innocent defendant), $l(C_{12}) = 10000$ (e.g., a penalty of 10000 for convicting an innocent person) and $l(C_{21}) = 1000$ (e.g., a penalty of 1000 for acquitting the offender). Nodes Θ and E show probability distributions and node A the expected losses. The loss values in nodes l, c and T have been omitted to focus the reader's attention on the values of interest in nodes A, E and Θ. The nodes with thicker borders represent instantiated nodes, with the certain state in bold.* **(a)** *Instantiating node T to t_0 produces the expected losses of a_1 and a_2 without any additional information.* **(b)** *Instantiating node T to t_1 gives the marginal probabilities of the test results $E_1^{t_1}$ and $E_2^{t_1}$ when the test is performed.* **(c)** *Instantiating node T to t_1 and node E to $E_1^{t_1}$ indicates the values for $\bar{l}(a_i|E_1^{t_1}, I) + c$, $i = 1, 2$, in node A.* **(d)** *Similarly, instantiating node T to t_1 and node E to $E_2^{t_1}$ indicates the values for $\bar{l}(a_i|E_2^{t_1}, I) + c$, $i = 1, 2$, in node A.*

a_1 and a_2 are, for example, the factfinder's decision of convicting or acquitting the defendant, and test t_1 is a forensic analysis whose results provide information on whether the defendant is guilty (this is state Θ_1) or innocent (state Θ_2) of the alleged offense. In this simplified example, the forensic analysis t_1 can produce either result $E_1^{t_1}$ or result $E_2^{t_1}$. Of the two possible states of nature, Θ_1 (guilty) and Θ_2 (innocent), $E_1^{t_1}$ supports the defendant's guilt, Θ_1, and $E_2^{t_1}$ supports the defendant's innocence, Θ_2, according to the following probability table for the conditional probabilities of $E_1^{t_1}$ and $E_2^{t_1}$:

	Θ_1	Θ_2
$E_1^{t_1}$	0.99999	0.0001
$E_2^{t_1}$	0.00001	0.9999

If no test is performed, then there is no test result (this is state $E_1^{t_0}$). Consider an example where the factfinder's prior probability (i.e., the probability without the information from the forensic analysis) of the defendant's guilt is 0.1, the cost of performing the forensic analysis is 10 units, and the loss function (in the same units as the cost of performing the test) is:

$$l(C_{11}) = l(C_{22}) = 0,$$
$$l(C_{12}) = 10000,$$
$$l(C_{21}) = 1000.$$

This loss function assigns no loss for convicting a guilty defendant and for acquitting an innocent defendant, and gives a penalty of 10000 units for convicting an innocent person, and a penalty of 1000 units for acquitting the offender.

The ID in Fig. 7.2 indicates the probability distributions over Θ and E, and the expected losses of convicting and acquitting the defendant in node A. Instantiating node T to t_0 produces the expected losses without the additional information from the forensic analysis (Fig. 7.2(a)), that is, the components of type *I.* in Eq. (7.1) and (7.2):

$$\begin{aligned} \bar{l}(a_1|I) &= l(C_{11})Pr(\Theta_1|I) + l(C_{12})Pr(\Theta_2|I) \\ &= 0 \times 0.1 + 10000 \times 0.9 \\ &= 9000\ , \end{aligned} \tag{7.3}$$

and

$$\begin{aligned} \bar{l}(a_2|I) &= l(C_{21})Pr(\Theta_1|I) + l(C_{22})Pr(\Theta_2|I) \\ &= 1000 \times 0.1 + 0 \times 0.9 \\ &= 100\ . \end{aligned} \tag{7.4}$$

Without the information from the forensic analysis, the expected loss of convicting the defendant is 9000 units, and the expected loss of acquitting the defendant is 100 units. These numerical values reflect the low probability (in this case) that the defendant is guilty, and the high penalty for convicting an innocent person.
In addition, Fig. 7.2(a) indicates:

$$Pr(E_1^{t_0}|I) = 1 \tag{7.5}$$

in node E. By definition, if the test is not performed, there is no result (state $E_1^{t_0}$).

Instantiating node T to t_1, posts the marginal probability distribution over $E_k^{t_1}$ in node E (Fig. 7.2(b)). Instantiating node T to t_h, $h \neq 0$, therefore produces the components of type *II.* in Eq. (7.1) and (7.2):

$$\begin{aligned} Pr(E_1^{t_1}|I) &= Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I) \\ &= 0.99999 \times 0.1 + 0.0001 \times 0.9 \\ &= 0.100089\ , \end{aligned} \tag{7.6}$$

and

$$\begin{aligned} Pr(E_2^{t_1}|I) &= Pr(E_2^{t_1}|\Theta_1, I)Pr(\Theta_1|I) + Pr(E_2^{t_1}|\Theta_2, I)Pr(\Theta_2|I) \\ &= 0.00001 \times 0.1 + 0.9999 \times 0.9 \\ &= 0.899911\ . \end{aligned} \tag{7.7}$$

The prior probability distribution over Θ, and the forensic analysis's high specificity make test result $E_2^{t_1}$ the most probable test result when the forensic analysis is performed.

And finally, instantiating node T to t_1, and node E to $E_k^{t_1}$, $k = 1, 2$, produces the expected losses that take into account the information provided by $E_k^{t_1}$ and the cost of obtaining this information (Figs. 7.2(c) and 7.2(d)). Instantiating node T to t_h and node E to $E_k^{t_h}$, $k = 1, 2, \ldots, q$, produces the components of type *III.* in Eq. (7.1) and (7.2):

$$\begin{aligned}
\bar{l}(a_1|E_1^{t_1}, I) + c &= l(C_{11})\frac{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_1^{t_1}|I)} + l(C_{12})\frac{Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_1^{t_1}|I)} + c \\
&= 0 \times \frac{0.99999 \times 0.1}{0.100089} + 10000 \times \frac{0.0001 \times 0.9}{0.100089} + 10 \\
&= 18.992 \ , \qquad (7.8)
\end{aligned}$$

$$\begin{aligned}
\bar{l}(a_2|E_1^{t_1}, I) + c &= l(C_{21})\frac{Pr(E_1^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_1^{t_1}|I)} + l(C_{22})\frac{Pr(E_1^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_1^{t_1}|I)} + c \\
&= 1000 \times \frac{0.99999 \times 0.1}{0.100089} + 0 \times \frac{0.0001 \times 0.9}{0.100089} + 10 \\
&= 1009.101 \ , \qquad (7.9)
\end{aligned}$$

$$\begin{aligned}
\bar{l}(a_1|E_2^{t_1}, I) + c &= l(C_{11})\frac{Pr(E_2^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_2^{t_1}|I)} + l(C_{12})\frac{Pr(E_2^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_2^{t_1}|I)} + c \\
&= 0 \times \frac{0.00001 \times 0.1}{0.899911} + 10000 \times \frac{0.9999 \times 0.9}{0.899911} + 10 \\
&= 10009.989 \ , \qquad (7.10)
\end{aligned}$$

$$\begin{aligned}
\bar{l}(a_2|E_2^{t_1}, I) + c &= l(C_{21})\frac{Pr(E_2^{t_1}|\Theta_1, I)Pr(\Theta_1|I)}{Pr(E_2^{t_1}|I)} + l(C_{22})\frac{Pr(E_2^{t_1}|\Theta_2, I)Pr(\Theta_2|I)}{Pr(E_2^{t_1}|I)} + c \\
&= 1000 \times \frac{0.00001 \times 0.1}{0.899911} + 0 \times \frac{0.9999 \times 0.9}{0.899911} + 10 \\
&= 10.001 \ , \qquad (7.11)
\end{aligned}$$

and

$$\bar{l}(a_i|E_1^{t_0}, I) + c = \bar{l}(a_i|I) \ , \ i = 1, 2 \ . \qquad (7.12)$$

If performing the forensic analysis leads to result $E_1^{t_1}$, then the action of convicting the defendant minimizes the total expected loss (i.e., $18.992 < 1009.101$), and if performing the forensic analysis leads to result $E_2^{t_1}$, then the action of acquitting the defendant minimizes the total expected loss (i.e., $10009.989 > 10.001$). Not performing the forensic analysis does not produce any additional information and has a cost of 0 units, so that the expected losses are the same as in Eqs. (7.3) and (7.4).

Inserting the numerical values obtained from the ID for components *I.* (Eq. (7.3)-(7.4)), *II.* (Eq. (7.5)-(7.7)) and *III.* (Eq. (7.8)-(7.12)) into Eq. (7.2) produces:

$$\begin{aligned}
ENG(t_0) &= min(9000, 100) - min(9000, 100) \times 1 \\
&= 100 - 100 \\
&= 0 \ ,
\end{aligned}$$

and

$$\begin{aligned}
ENG(t_1) &= min(9000, 100) - \\
&\quad [min(18.992, 1009.101) \times 0.100089 + min(10009.989, 10.001) \times 0.899911] \\
&= 100 - [18.992 \times 0.100089 + 10.001 \times 0.899911] \\
&= 100 - 10.901 \\
&= 89.099
\end{aligned}$$

A rational decision-maker chooses to perform the test that maximizes the $ENG(t_h)$. Since $ENG(t_1) > ENG(t_0)$, the rational decision-maker chooses to perform the forensic analysis t_1.

Note that both *Hugin Researcher 7.6* and *GeNIe 2.0* include algorithms for computing the $EVOI(t_h)$ (in this example, $EVOI(t_0) = 0$ and $EVOI(t_1) = 99.099$).[6] In addition, *Hugin Researcher 7.6* has an algorithm for finding the $\underset{h}{arg\ max}\ ENG(t_h)$ and computing the maximum expected utility minus the cost for performing this test (for a translation of the loss function to a utility function by multiplying the losses by -1 in this example, this value would be -10.901).[7] *GeNIe 2.0* has algorithms for both the maximum expected utilities minus the cost for performing t_h, $h = 0, 1, \ldots, p$, (for a translation of the loss function to a utility function by multiplying the losses by -1 in this example, these values would be -100 for t_0 and -10.901 for t_1),[8] and for computing the $max_h\ ENG(t_h)$ (in this example, $max_h\ ENG(t_h) = 89.099$).[9]

According to this illustrative example, IDs provide a means for forensic scientists to structure two-stage decision problems and compute the expected losses (or utilities) and probabilities required to evaluate the $ENG(t_h)$ which the scientist must maximize for choosing the most rational course of action.

[6] In *Hugin Researcher 7.6*, the user must instantiate node T to t_h, select node A, open the *Value of Information* tool in the menu *Network* and submenu *Analysis*, select *Custom group* and node E under the tab *Information Variables*, and press *Calculate* under the tab *Analysis*. In *GeNIe 2.0*, the user must select the *Value of Information* tool in the menu *Network*, select node E, specify that the analysis is for decision node A and from the point of view of node A, and press *Update*. In both of these programs, the structure of the ID must not include the precedence link between nodes E and A for this evaluation to be correct.

[7] For this, the user selects the δ *spu* (i.e., *Decision: single policy update*) function in the tool bar of the model's *Run Mode*. For this evaluation, the structure of the ID must include the precedence link between nodes E and A.

[8] The test node automatically indicates these values when the model is updated. This computation requires the structure of the ID to include the precedence link between nodes E and A.

[9] For this, the user must select the *Value of Information* tool in the menu *Network*, select node E, specify that the analysis is for decision node A and from the point of view of node T, and press *Update*. This evaluation requires the structure of the ID to leave out the precedence link between nodes E and A.

Chapter 8

Conclusions

This thesis studied a normative approach to decision problems encountered by scientists in forensic science laboratories by performing theoretical case studies to address the following three propositions:

Proposition 1: *Decision theory provides a framework for organizing a broad range of decision problems encountered by scientists in forensic science laboratories dealing with different categories of trace evidence.*

Proposition 2: *Bayesian networks (BNs) provide a means for handling new complex inference problems encountered by forensic scientists.*

Proposition 3: *Influence diagrams (IDs) provide forensic scientists with a practical tool for structuring and providing the elements required for making coherent choices in forensic decision problems.*

The results of the case studies have provided elements in support of each of these propositions.

The uniform framework presented in this thesis for confronting forensic decision problems distinguishes between one-stage and two-stage decision problems. It shows how to organize each type of problem by breaking it down into its different elements so as to form the spaces necessary for applying Bayesian inference and decision theory. The intention of this framework is not to describe a scientist's degrees of belief, preferences and choices as they currently are, but to suggest what they ought to be if the scientist wishes to be a coherent decision-maker. The application of this framework therefore provides the scientist with a tool for thinking coherently about a decision problem. In particular, it:

- offers a transparent structure in which the scientists can express their views and communicate with each other,
- shows what parts of the decision problem can be analyzed independently of other parts, and
- helps the scientists better understand their situation, degrees of belief and preferences, as well as the problem's key parameters that should have an impact on their choices.

Hence, the proposed framework is an interactive and exploratory tool for better understanding a decision problem so that this understanding may lead to better informed choices.

In view of the serious consequences that the decisions made by forensic scientists can lead to in the administration of justice, this framework provides a logical foundation for justifying the scientist's course of actions. It is the natural extension of the Bayesian inference models already advocated for the interpretation of scientific evidence to handling decision problems under uncertainty. Its major advantage over current decision protocols is its case-specific approach to decision problems: unlike fixed standards prescribing threshold values

and laboratory protocols applicable to all cases alike, a decision-theoretic approach takes into account the probability distributions and utility (or loss) functions for the particular case in question to study the most rational course of action for that particular case. Hence, future perspectives would be to develop this normative approach into a prescriptive approach that applies these ideas in forensic practice.

Furthermore, some of the ideas presented in this thesis for scientific evidence describe a general reasoning process for decision problems concerning evidential testimony. These are, of course, not restricted to forensic science. They apply just as well to other categories of evidential testimony, such as eyewitnesses and legal medicine. Hence, another possible development of these ideas is their application to other types of evidential testimony.

Part III

Published Papers

Chapter 9

Modeling the Forensic Two-trace Problem with Bayesian Networks

Abstract

The forensic two-trace problem is a perplexing inference problem introduced by Evett (1987). Different possible ways of wording the competing pair of propositions (i.e., one proposition advanced by the prosecution and one proposition advanced by the defence) led to different quantifications of the value of the evidence (Meester and Sjerps, 2003). Here, we re-examine this scenario with the aim of clarifying the interrelationships that exist between the different solutions, and in this way, produce a global vision of the problem. We propose to investigate the different expressions for evaluating the value of the evidence by using a graphical approach, i.e. Bayesian networks, to model the rationale behind each of the proposed solutions and the assumptions made on the unknown parameters in this problem.

9.1 Introduction

The two-trace problem, introduced by Evett (1987), is a perplexing inference problem that continues to puzzle many forensic scientists. It considers a scenario where forensic investigators recover two items of a particular category of trace evidence on a crime scene, e.g. two bloodstains, and compare both of these to the sample taken from a suspect. The question of interest to the court is, 'How strong is the evidence resulting from these two comparisons in favor of the prosecution or the defence?'

The objective of the forensic scientist's testimony is to answer this question. The answer to this question takes the form of the *value of the evidence* (e.g., Aitken and Taroni, 2004):

$$V = \frac{Pr(\text{evidence}|\text{proposition 1}, I)}{Pr(\text{evidence}|\text{proposition 2}, I)} \,, \tag{9.1}$$

Reprinted from Artificial Intelligence and Law (2013) 21:221-252, Modeling the forensic two-trace problem with Bayesian networks, S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni, Copyright Springer Science+Business Media Dordrecht 2012, with the kind permission from Springer Science+Business Media B.V.

where proposition 1 is the proposition advanced by the prosecution, proposition 2 the proposition advanced by the defence, and I the background information consisting of the forensic scientist's knowledge on the case circumstances prior to observing the evidence. The evidence is an intrinsic trait (e.g., the blood group or DNA profile) of the two traces and the suspect's sample, observed as a result of the test or analysis performed in the forensic laboratory. Prior to hearing the forensic scientist's testimony, the prosecution and the defence each take position on the origin of the traces. These views are formalized into the two propositions, that is, into two statements that are each either true or false. As a pair, these propositions must be mutually exclusive[1], yet there is no requirement for them to be exhaustive[2] (e.g., Robertson and Vignaux, 1995a; Aitken and Taroni, 2004). In this case, the first proposition (advanced by the prosecution) links the suspect to the crime stains, and the second (advanced by the defence) rejects such a link. The fact-finder (a judge or jury member) has a particular degree of belief in the truth of each of these propositions before hearing the forensic testimony. By presenting the value of the evidence V, the forensic scientist's testimony conveys by how much more or less the evidence supports the first proposition with regard to the second proposition: if $V > 1$, the evidence supports the first proposition; if $V < 1$, the evidence supports the second proposition; and if $V = 1$, the evidence does not provide support for either of the two propositions, meaning that it is irrelevant for discriminating between them. Hence, the value of the evidence allows the fact-finder to update his or her belief in the truth of these propositions, and construct an informed opinion about each party's account of the events.

9.1.1 Aim and outline of this paper

With two traces making up the recovered evidence, there are several possibilities for formulating a pair of propositions: they can focus on one of the two traces, or on both, and in the latter case, either specify or not specify which of the two traces originates (or does not originate) from the suspect. What is disturbing for a fact-finder hearing a forensic scientist's testimony in the context of a two-trace problem, is that the value of the evidence, as given by Eq. (9.1), is *different* for different pairs of propositions (Meester and Sjerps, 2003).

The aim of this paper is to investigate the value of the evidence in a two-trace problem with regard to different pairs of propositions, by unifying three different pairs in a single framework. To accomplish this, we will construct a Bayesian network, i.e., a graphical probability model. In forensic contexts, Bayesian networks help examine the reasonableness of the formal derivation of a formula, that is, the assumptions that have been made (Taroni et al., 2006a). This allows us to compare the derived values of the evidence for different pairs of propositions. In addition, these models allow the user to perform complex probabilistic calculations that take into account the probability assignments over all of the unknown parameters. In this way, we hope to provide a global model which offers a complete and realistic approach to the valuation of scientific evidence in a two-trace problem. With this model, we hope to draw the reader's attention to the importance of the formulation of a pair of propositions, and increase his/her awareness of the impact that subtle differences in these formulations can have on the value of the evidence.

Besides a brief description of what Bayesian networks are and how they work (Section 9.2), we do not give a detailed explanation on Bayesian networks, and refer the interested reader to one of the many publications on the subject (e.g., Jensen, 2001; Kjaerulff and Madsen, 2008). Section 9.3 gives an overview of the two-trace problem as we will treat it in this paper, and Section 9.4 describes the notation we will use. Section 9.5 explains how we construct the Bayesian network, and Sections 9.6 and 9.7 illustrate the use of this model and the influence of the different parameters through a numerical example. Concluding remarks are in Section 9.8.

[1] Two propositions are *mutually exclusive* if they cannot both be true at the same time.

[2] A set of propositions is *exhaustive* if it covers all scenarios, so that at least one of its propositions is always true.

9.2 Bayesian networks

Bayesian networks are graphical probability models, also known as probabilistic expert systems. The key advantage of these models is their capacity of splitting up a complex inference problem into its different parts. They represent random variables as nodes, and dependence relationships between the random variables as arrows connecting the nodes to form a directed acyclic graph. The random variables can be either discrete or continuous, but for the sake of simplicity we will use discrete nodes in this paper. Thus, each random variable will consist of a finite and exhaustive list of mutually exclusive states. The arrows model the probabilistic relationships between the variables by connecting a 'parent' node to a 'child' node. They condition the probability distribution of the child node upon each of its parents with probability tables that allow the user to quantify the probabilistic relationships.

In this way, the Bayesian network decomposes the joint probability distribution of a set of random variables $X_1, ..., X_n$ into the product of each of their probabilities conditioned on their parents. This is known as the Markov property:

$$Pr(X_1, ..., X_n) = \prod_{i=1}^{n} Pr(X_i | parents(X_i)) \ . \tag{9.2}$$

It is important to stress that there is no true model, because a model is personal and reflects the constructor's view of the problem and the information available at the time of the construction (Lindley, 2000). As our understanding of the issue progresses, the constructed network may evolve to model a situation more accurately, so that several different Bayesian network structures may be accepted as a description of the same scenario (Garbolino, 2001).

The relevance of Bayesian networks for applications in forensic science was first recognized - in print - by Aitken and Gammerman (1989). Key publications in legal literature followed (Edwards, 1991 - 1992; Schum, 1994; Kadane and Schum, 1996), presenting thorough descriptions of the potential of probabilistic models for reasoning about evidence in real cases. Since then, the application of Bayesian networks has covered different aspects of evidential assessment (Taroni et al., 2006a), in particular the evaluation of DNA evidence (Dawid et al., 2002, 2007; Mortera et al., 2003), evidence collected in fire debris (Biedermann et al., 2005a,b), firearm evidence (Biedermann and Taroni, 2006; Biedermann et al., 2009a), and fibre evidence (Garbolino and Taroni, 2002), as well as practical considerations on how to present their results in court (Fenton and Neil, 2011).

In this study, we constructed the Bayesian networks using the software *Hugin Researcher 7.3*, by Hugin Expert A/S. This program allows the user to construct and use Bayesian networks that contain numerical probability values. It can only carry out numerical propagations between the nodes. To derive the algebraic expressions corresponding to the calculations performed by the model, the user applies Eq. (9.2). The probabilistic relationships defined by the structure and the probability tables tell the user how the probability of a compound event is broken down into separate conditional probabilities.

9.3 The two-trace problem

We denote the following pair of propositions 'pair H':

proposition 1: *at least one of the crime stains comes from the suspect*;
proposition 2: *neither of the crime stains comes from the suspect*.

These propositions are called source level propositions according to the hierarchy of propositions defined by Cook et al. (1998b), because they describe whether a particular object or person is the source, or origin, of the traces recovered on the crime scene. Source level propositions are different from activity level propositions (describing the activity that led to the transfer of the traces from their source to the crime scene) and crime level propositions (concerned with whether the suspect actually committed the crime under investigation). In

this paper we treat only source level propositions. For activity or crime level evaluations of the evidence in a two-trace problem, see Triggs and Buckleton (2003) and Gittelson et al. (2012a), and Dawid (2004), respectively.

The value of the evidence with regard to the above propositions depends on the evidence observed. There are three possibilities:

1. if neither of the two crime stains matches the suspect's sample, then the likelihood of the first proposition is 0, and consequently Eq. (9.1) becomes

$$V = 0 \ ; \tag{9.3}$$

2. if one of the crime stains matches the suspect's sample, Evett (1987) showed that this leads to

$$V = \frac{1}{2\gamma} \ , \tag{9.4}$$

where γ represents the *match probability* (Weir, 2000) of the matching trait in the relevant population of possible crime stain donors [Note that originally, Evett (1987) did not deduce this expression for the source level propositions as described, but for their equivalents at the crime level, assuming the relevance of both traces to be maximal. A crime level evaluation with maximal relevance produces the same value of the evidence as the source level evaluation presented here (see, e.g., Aitken and Taroni (2004)).];

3. and if both of the crime stains match the suspect's sample, most forensic scientists would assume that the two traces come from a single contributor so that Eq. (9.1) reduces to

$$V = \frac{1}{\gamma} \ . \tag{9.5}$$

These assessments are based on the assumption that no laboratory errors are possible, an assumption we maintain throughout this paper. Note, however, that relaxing this assumption may have a considerable effect on the value of the evidence (Thompson et al., 2003).

Among these three ratios, Eqs. (9.3) and (9.5) are the same as for a scenario involving a single crime stain. This is because the differentiation between the two traces is not necessary in these cases in order to describe the observed evidence. In these two cases, one can combine the two crime stains into a single group, which we see as either matching (Eq. (9.5)), or not matching (Eq. (9.3)) the suspect's sample. In both of these cases, the reasoning that leads to Eq. (9.3) and (9.5) is the same as that applied to the evaluation of the value of a single crime stain.

This is different for the case involving one matching stain and one non-matching stain (item 2 in the list). This case requires the forensic scientist to distinguish between the two traces by multiplying the traditional value of $\frac{1}{\gamma}$ by a factor of 0.5 (We will discuss the meaning of this additional factor in Section 9.7). This is the case which interests us in this paper.

Evett (1987) was not the only author to treat this problem. After Evett (1987), the case of one matching stain and one non-matching stain gave rise to the formulation of other propositions, which led to evidential values that were not equal to Eq. (9.4). According to Meester and Sjerps (2003), a pair of propositions worded slightly differently (note that the wording of these propositions has been modified here with regard to their original formulation in Meester and Sjerps (2003), yet their logical meaning remains unchanged), that is,

proposition 1: *crime stain 1 comes from the suspect*;
proposition 2: *neither of the crime stains comes from the suspect*;

(we denote this pair 'pair H_1'') produced a value of

$$\frac{1}{\gamma}\ , \tag{9.6}$$

and the pair

proposition 1: *crime stain 1 comes from the suspect*;
proposition 2: *crime stain 1 does not come from the suspect*;

(denoted 'pair H_1''') produced a value of[3]

$$\frac{2-\delta}{2\gamma(1-\delta)}\ , \tag{9.7}$$

for evidence consisting of a match between crime stain 1 and the suspect's sample, and a non-match between crime stain 2 and the suspect's sample. In Eq. (9.7), the probability denoted δ represents the prior probability that the suspect was one of two crime stain donors. This probability had to be introduced to correctly evaluate the probability of the evidence given proposition 2 and I. See Section 9.7.3 for further explanations.

The pair of propositions H (on page 69) is related to the above two pairs (pairs H_1' and H_1'') when the forensic scientist observes a match between the suspect's sample and stain 1, and a non-match between the suspect's sample and stain 2: in this case, all three pairs of propositions have identical posterior odds of

$$\frac{\delta}{2(1-\delta)\gamma}\ , \tag{9.8}$$

for a prior probability of δ that the suspect was a crime stain donor. (Note that we call the odds of a pair of propositions 'prior odds' before observing the evidence, and 'posterior odds' after observing the evidence. We use the terms 'prior probability' and 'posterior probability' in the same way.)

Here, the observation of a matching trait in the suspect's sample and stain 1, and a non-matching trait between the suspect's sample and stain 2 has made the three pairs of propositions logically equivalent, since it has become impossible for the suspect to be the donor of stain 2. Algebraically, this comes down to multiplying the value of the evidence by the corresponding pair of propositions' prior odds according to the odds' form of Bayes' theorem (Meester and Sjerps, 2003):

$$\underbrace{\frac{\delta}{1-\delta}}_{\textit{prior odds for pair } H} \times \underbrace{\frac{1}{2\gamma}}_{\textit{value of the evidence for pair } H}\ , \tag{9.9}$$

$$\underbrace{\frac{\delta/2}{1-\delta}}_{\textit{prior odds for pair } H_1'} \times \underbrace{\frac{1}{\gamma}}_{\textit{value of the evidence for pair } H_1'}\ , \tag{9.10}$$

$$\underbrace{\frac{\delta/2}{1-\delta/2}}_{\textit{prior odds for pair } H_1''} \times \underbrace{\frac{2-\delta}{2\gamma(1-\delta)}}_{\textit{value of the evidence for pair } H_1''}\ . \tag{9.11}$$

Meester and Sjerps (2003) and Meester and Sjerps (2004a) conclude from this that forensic scientists should use posterior odds in the place of the value of the evidence to communicate the strength of forensic evidence, a recommendation which makes no attempt at

[3]Note that Eq. (9.7) gives the simplified form of the value of the evidence, so that the numerator and denominator of this ratio do not represent the probabilities forming the numerator and denominator in Eq. (9.1).

clarifying the logical relationships between the three pairs of propositions and their different values for the same evidence (a point criticized by Dawid (2004)).

The Bayesian network we present in Section 9.5 will illustrate the interrelationships between these three pairs of propositions by modeling them in separate nodes. Before explaining the rationale behind this model, the next section presents the notation we will use in the rest of this paper.

9.4 Notation

We distinguish between the background information (Section 9.4.1), the propositions (Section 9.4.2), the unknown parameters δ, λ and τ (Section 9.4.3), and the evidence (Section 9.4.4).

9.4.1 Background information

The background information I is all of the knowledge available *prior* to observing the evidence. This information includes the case circumstances (e.g., the location of the crime scene), the facts surrounding the recovery of the two traces on the crime scene (e.g., their exact location on the scene), the fact that one suspect has been found from whom a sample has been obtained for comparison with the recovered traces, and the non-scientific information associating this suspect to the crime scene (e.g., witness statements asserting the suspect's presence near the scene). All of the probabilities assessed in a case are conditional probabilities given I. However, for the sake of brevity in the mathematical expressions that follow, we shall hereafter omit I from their notation.

9.4.2 Propositions

The propositions reflect the viewpoints of the prosecution and the defence. At the time they are formulated, the evidence has not yet been observed, so that these formulations are independent of the evidence, and based solely on the background information. Each proposition depicts the most plausible situation(s) given the party's point of view and the background information (Robertson and Vignaux, 1995a). Since the background information is case-specific, one pair of propositions may be reasonable in one case, yet unreasonable in another case.

Section 9.1 introduced four propositions, which we denote with capital letters as follows:

D - *at least one of the crime stains comes from the suspect;*
$\bar{D}$ - *neither of the crime stains comes from the suspect;*

C_1 - *crime stain 1 comes from the suspect;*
$\bar{C}_1$ - *crime stain 1 does not come from the suspect.*

The horizontal bar over a capital letter means that the proposition described is the negation of the proposition denoted by that letter (i.e., its complement). The number figuring as a subscript to propositions C and $\bar{C}$ indicates which crime stain the proposition refers to. Analogous to C_1 and $\bar{C}_1$, we also formulate:

C_2 - *crime stain 2 comes from the suspect;*
$\bar{C}_2$ - *crime stain 2 does not come from the suspect.*

Meester and Sjerps (2003) considered 3 pairs of propositions, denoted here as pairs H, H_1', and H_1'' (the subscript '1' indicates that the pair contains at least one proposition referring only to crime stain 1). These combine in different ways the four propositions D, $\bar{D}$, C_1 and $\bar{C}_1$ as follows:

Table 9.1: *Definition of the parameters δ, λ and τ.*

δ	Probability that the suspect is a crime stain donor
λ	Probability that there were two distinct donors
τ	Probability that crime stain 1 comes from the suspect, given that the suspect is one of two donors

pair H: D - *at least one of the crime stains comes from the suspect;*
$\bar{D}$ - *neither of the crime stains comes from the suspect;*

pair H_1': C_1 - *crime stain 1 comes from the suspect;*
$\bar{D}$ - *neither of the crime stains comes from the suspect;*

pair H_1'': C_1 - *crime stain 1 comes from the suspect;*
$\bar{C}_1$ - *crime stain 1 does not come from the suspect.*

To model a pair of propositions as a node in a Bayesian network, the node must have an exhaustive list of states (see Section 9.2). The propositions in pairs H and H_1'' already form an exhaustive set of possibilities, and can therefore be modeled as nodes with two states. Yet pair H_1' is not exhaustive because it does not consider the possibility that crime stain 2 comes from the suspect. An exhaustive list would need to include all of the possible combinations between C_1, $\bar{C}_1$, C_2 and $\bar{C}_2$, i.e.,

$C_1 \cap C_2$ - *both crime stains come from the suspect;*
$C_1 \cap \bar{C}_2$ - *crime stain 1 comes from the suspect, and crime stain 2 does not come from the suspect;*
$\bar{C}_1 \cap C_2$ - *crime stain 1 does not come from the suspect, but crime stain 2 comes from the suspect;*
$\bar{C}_1 \cap \bar{C}_2$ - *neither of the crime stains comes from the suspect.*

In this list, proposition $\bar{D}$ is equivalent to $\bar{C}_1 \cap \bar{C}_2$, and proposition C_1 to $\{C_1 \cap C_2\} \cup \{C_1 \cap \bar{C}_2\}$. Modeling pair H_1' as a node with exhaustive states in a Bayesian network will therefore require the additional state $\bar{C}_1 \cap C_2$ in this node.

Analogous to pair H_1', we define pair H_2' for the combination of C_2 and $\bar{D}$:

pair H_2': C_2 - *crime stain 2 comes from the suspect;*
$\bar{D}$ - *neither of the crime stains comes from the suspect;*

and analogous to pair H_1'', we define pair H_2'' for the combination of C_2 and $\bar{C}_2$:

pair H_2'': C_2 - *crime stain 2 comes from the suspect;*
$\bar{C}_2$ - *crime stain 2 does not come from the suspect.*

Modeling pair H_2' as a node in a Bayesian network will follow the same reasoning as for pair H_1' by requiring the additional state $C_1 \cap \bar{C}_2$ to make the node's states exhaustive.

9.4.3 Unknown parameters

The two-trace problem involves three unknown parameters (Table 9.1):

- **δ:** The first parameter, δ, we encountered in Eq. 9.7. This is the prior probability that the suspect is a crime stain donor, i.e.,

  $$Pr(D) = \delta \ ,$$

  as defined in Meester and Sjerps (2003). δ describes the probability that a trace recovered on the crime scene comes from the suspect based on the information available prior to the laboratory analyses of the crime stains. This parameter takes into account

the background information regarding the suspect's presence on or near the crime scene during the lapse of time when the traces were deposited (for example, witness statements, data from mobile phone providers, and images from surveillance cameras), as well as background information regarding the suspect's ability to transfer the type of trace evidence in question (for example in the case of recovered bloodstains, the fact that the suspect had a scratch, cut or other injury with blood loss after the time when the traces were deposited would increase δ). In this model, the value of δ is based on this background information alone, independent of whether the recovered traces come from a single source or from two different sources. Note however that in some cases this assumption of δ being independent of the total number of crime stain donors may not be reasonable. Notably when the background information described above is very poor or not available, it may be reasonable to assume that δ is greater in the case of two donors than in the case of a single donor (Meester and Sjerps, 2004a,b). This situation is not treated in this paper, but it would require an additional dependence relationship in the Bayesian network presented in Section 9.5 (Fig. 9.2).

- λ**:** The second parameter describes the uncertainty on the number of donors (Dawid, 2004). Defined by Dawid (2004), λ represents the probability that there are two distinct donors. Before observing the evidence, all we know is that there are two traces. A priori, these may come from the same source with a probability of $1-\lambda$, and from two different sources with a probability of λ.

- τ**:** The third parameter, τ, considers the conditional probability that crime stain 1 comes from the suspect given that the suspect is one of two crime stain donors, i.e.,

  $$Pr(C_1|2 \text{ donors}, D) = \tau \ .$$

  From this definition, it follows that $1-\tau$ is the probability of crime stain 2 coming from the suspect given that the suspect is the source of one of the two traces, i.e.,

  $$Pr(C_2|2 \text{ donors}, D) = 1-\tau \ .$$

All of these parameters are assessed on the basis of the background information alone, that is, before observing the evidence: the value of δ will depend on the prior information regarding the suspect's connection to the crime scene; and the values of λ and τ are based on the circumstancial information of the case, including the location of the traces on the scene, witness reports, and images from surveillance cameras.

These prior assessments determine the probabilities of the propositions (see column 3 of Table 9.2). The probabilities of D and $\bar{D}$ are determined by δ, as described above. The probabilities of C_1, $\bar{C}_1$, C_2 and $\bar{C}_2$ are made up of the probabilities of $C_1 \cap C_2$, $C_1 \cap \bar{C}_2$, $\bar{C}_1 \cap C_2$ and $\bar{C}_1 \cap \bar{C}_2$, which are

$$\begin{aligned} Pr(C_1 \cap C_2) &= \delta(1-\lambda) \\ Pr(C_1 \cap \bar{C}_2) &= \delta\lambda\tau \\ Pr(\bar{C}_1 \cap C_2) &= \delta\lambda(1-\tau) \\ Pr(\bar{C}_1 \cap \bar{C}_2) &= 1-\delta \ , \end{aligned}$$

so that the probabilities of C_1, $\bar{C}_1$, C_2 and $\bar{C}_2$ are

$$\begin{aligned} Pr(C_1) &= Pr(C_1 \cap C_2) + Pr(C_1 \cap \bar{C}_2) \\ &= \delta(1-\lambda) + \delta\lambda\tau \ , \end{aligned}$$

$$\begin{aligned} Pr(\bar{C}_1) &= Pr(\bar{C}_1 \cap C_2) + Pr(\bar{C}_1 \cap \bar{C}_2) \\ &= \delta\lambda(1-\tau) + 1 - \delta \ , \end{aligned}$$

$$\begin{aligned} Pr(C_2) &= Pr(C_1 \cap C_2) + Pr(\bar{C}_1 \cap C_2) \\ &= \delta(1-\lambda) + \delta\lambda(1-\tau) \ , \end{aligned}$$

and

$$\begin{aligned} Pr(\bar{C}_2) &= Pr(C_1 \cap \bar{C}_2) + Pr(\bar{C}_1 \cap \bar{C}_2) \\ &= \delta\lambda\tau + 1 - \delta \ . \end{aligned}$$

The examples in Sections 9.6 and 9.7 will illustrate the impact of parameters δ, λ and τ on the value of the evidence and on the posterior odds of the different pairs of propositions.

9.4.4 Evidence

The evidence is the new piece of information we observe. It is the compound event of observing the states of the three variables X, Y_1 and Y_2. X denotes the profile of the suspect's sample, Y_1 the profile of the first of the recovered traces, which we call 'crime stain 1', and Y_2 the profile of the second of the recovered traces, which we call 'crime stain 2'.

We assume that the analysis performed is capable of distinguishing between k different profiles, which we label $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$. Profile Γ_i, $i = 1, 2, \ldots, k$, has a match probability of γ_i in the relevant population of possible crime stain donors. Note that the relevant population is defined on the basis of the background information. Before observing the evidence, X, Y_1, and Y_2 each have a probability of γ_i, $i = 1, 2, \ldots, k$, $\sum \gamma_i = 1$, to have profile Γ_i. After observing the evidence, the states of X, Y_1, and Y_2 are known with certainty. They are each equal to one of the k profiles, $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$.

In the next section, we combine the above evidence, propositions, and parameters in a Bayesian network for the two-trace problem.

9.5 Constructing a Bayesian network

The aim of this section is to construct a Bayesian network containing the propositions, the parameters, and the evidence, defined in the previous section. This section contains several technical details of the constructed Bayesian network, and may be skipped by readers interested more in the application of the model than in its construction. For the propositions we create nodes H, H_1', H_1'', H_2' and H_2'', and for the evidence, nodes X, Y_1 and Y_2. Table 9.2 provides the exhaustive list of the states and probabilities associated to each of these nodes.

Taroni et al. (2006a) proposed a model containing some of these nodes for a very specific scenario of a two-trace problem (Fig. 9.1). In this model, node F contains the inexhaustive list of states $C_1 \cap \bar{C}_2$, $\bar{C}_1 \cap C_2$ and $\bar{C}_1 \cap \bar{C}_2$. This model sets the profile of Y_1 equal to the profile

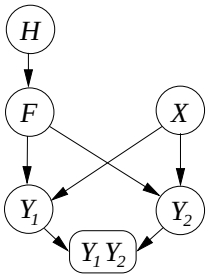


Figure 9.1: *The Bayesian network presented in Taroni et al. (2006a) for a very specific scenario of the two-trace problem. Nodes H, X, Y_1 and Y_2 consist of the states presented in Table 9.2, and node F of the states $C_1 \cap \bar{C}_2$, $\bar{C}_1 \cap C_2$ and $\bar{C}_1 \cap \bar{C}_2 \equiv \bar{D}$.*

of X if C_1 is true, and the profile of Y_2 equal to the profile of X if C_2 is true. Concerning the

Table 9.2: *Description of the states of the non-parametric nodes in the Bayesian network in Fig. 9.2. The parameters δ, λ, and τ are defined in Table 9.1.*

Nodes	States	Probabilities	Definitions of the states
H	D	δ	At least one of the crime stains comes from the suspect
	$\bar{D}$	$1-\delta$	Neither of the crime stains comes from the suspect
L	1 donor	$1-\lambda$	The crime stains come from the same source
	2 donors	λ	The crime stains come from two different sources
H_1''	C_1	$\delta(1-\lambda+\lambda\tau)$	Crime stain 1 comes from the suspect
	$\bar{C_1}$	$\delta\lambda(1-\tau)+1-\delta$	Crime stain 1 does not come from the suspect
H_2''	C_2	$\delta[1-\lambda+\lambda(1-\tau)]$	Crime stain 2 comes from the suspect
	$\bar{C_2}$	$\delta\lambda\tau+1-\delta$	Crime stain 2 does not come from the suspect
H_1'	C_1	$\delta(1-\lambda+\lambda\tau)$	Crime stain 1 comes from the suspect
	$\bar{C_1}\cap C_2$	$\delta\lambda(1-\tau)$	Only crime stain 2 comes from the suspect
	$\bar{D}$	$1-\delta$	Neither of the crime stains comes from the suspect
H_2'	C_2	$\delta[1-\lambda+\lambda(1-\tau)]$	Crime stain 2 comes from the suspect
	$C_1\cap\bar{C_2}$	$\delta\lambda\tau$	Only crime stain 1 comes from the suspect
	$\bar{D}$	$1-\delta$	Neither of the crime stains comes from the suspect
X	Γ_1	γ_1	Profile of the suspect's sample
	Γ_2	γ_2	
	$\vdots$	$\vdots$	
	Γ_k	γ_k	
Y_1	Γ_1	γ_1	Profile of crime stain 1
	Γ_2	γ_2	
	$\vdots$	$\vdots$	
	Γ_k	γ_k	
Y_2	Γ_1	γ_1	Profile of crime stain 2
	Γ_2	γ_2	
	$\vdots$	$\vdots$	
	Γ_k	γ_k	
Y_1Y_2	$\Gamma_1\Gamma_1$	$\gamma_1\gamma_1$	Profiles of crime stains 1 and 2 (as ordered pairs)
	$\Gamma_1\Gamma_2$	$\gamma_1\gamma_2$	
	$\Gamma_2\Gamma_1$	$\gamma_2\gamma_1$	
	$\vdots$	$\vdots$	
	$\Gamma_k\Gamma_k$	$\gamma_k\gamma_k$	

Table 9.3: *Probability table for node H_1' in Fig. 9.2. The states of H_1' are defined by the combinations of the states in nodes H_1'' and H_2''.*

	H_1'':	C_1		$\bar{C}_1$	
	H_2'':	C_2	$\bar{C}_2$	C_2	$\bar{C}_2$
H_1':	C_1	1	1	0	0
	$\bar{C}_1 \cap C_2$	0	0	1	0
	$\bar{D}$	0	0	0	1

Table 9.4: *Probability table for node H_2' in Fig. 9.2. The states of H_2' are defined by the combinations of the states in nodes H_1'' and H_2''.*

	H_1'':	C_1		$\bar{C}_1$	
	H_2'':	C_2	$\bar{C}_2$	C_2	$\bar{C}_2$
H_2':	C_2	1	0	1	0
	$C_1 \cap \bar{C}_2$	0	1	0	0
	$\bar{D}$	0	0	0	1

propositions, it specifies that C_1 and C_2 can only be true if D (in node H) is true. However, this model makes the assumption that C_1 and C_2 are equally likely under D, and it does not consider the possibility of C_1 and C_2 being true at the same time (i.e., node F contains an inexhaustive list of states). Node Y_1Y_2 combines the states of Y_1 and Y_2 as ordered pairs, so that the model computes the compound probability of the two crime stain profiles. This node is necessary for evaluating the value of the evidence (see Section 9.7).

We use this model as a starting point to extend and improve it to a more general Bayesian network for evaluating the value of the evidence in a two-trace problem. For this, we examine the following points: the relationship between the propositional nodes (Section 9.5.1), the uncertainty on the number of donors (Section 9.5.2), and the relationship between the propositional and the evidential nodes (Section 9.5.3).

9.5.1 Relationships between the propositional nodes H, H_1', H_1'', H_2' and H_2''

The postdata equivalence presented in Section 9.3 indicates a relationship between the nodes containing the pairs of propositions H, H_1' and H_1''. To expose the links that exist between these nodes, we analyse the logical relationships between the propositions that form the nodes' states.

The difference between proposition D and propositions C_1 and C_2 is that the former does not specify which trace, or traces, come(s) from the suspect, whereas the latter do. Logically, this means that propositions C_1 and C_2 are two subsets of proposition D, i.e., $C_1 \subset D$ and $C_2 \subset D$. In a Bayesian network, this relationship may be modeled by conditioning the probabilities of C_1 and C_2 on D (Taroni et al., 2006a). In other words, we model node H (containing proposition D) as a parent of nodes H_1'' (containing proposition C_1) and H_2'' (containing proposition C_2).

As for nodes H_1' and H_2', their states $\bar{C}_1 \cap C_2$, $C_1 \cap \bar{C}_2$ and $\bar{D}$ ($\equiv \bar{C}_1 \cap \bar{C}_2$) are combinations of C_1, $\bar{C}_1$, C_2, and $\bar{C}_2$. Each of these combinations is a subset of its single components: $\{C_1 \cap \bar{C}_2\} \subset C_1$, $\{C_1 \cap \bar{C}_2\} \subset \bar{C}_2$, $\{\bar{C}_1 \cap C_2\} \subset \bar{C}_1$, $\{\bar{C}_1 \cap C_2\} \subset C_2$, $\bar{D} \subset \bar{C}_1$ and $\bar{D} \subset \bar{C}_2$. Again, we find it convenient to model a subset as a child of its superset. Therefore, we model nodes H_1' and H_2' as children of nodes H_1'' and H_2'', with the conditional probability distributions given in Tables 9.3 and 9.4.

Table 9.5: *Probability table for node H_1'' in Fig. 9.2. This probability table contains the parameter $\tau = Pr(C_1|2\ donors, D)$.*

	H:	D		$\bar{D}$	
	L:	1 donor	2 donors	1 donor	2 donors
H_1'':	C_1	1	τ	0	0
	$\bar{C}_1$	0	$1-\tau$	1	1

Table 9.6: *Probability table for node H_2'' in Fig. 9.2. Note that the second, fifth and seventh columns describe impossible combination of states (i.e., in the second column, the suspect is a crime stain donor, and there is only a single donor for both crime stains, yet the suspect is not the donor of crime stain 1; and in the fifth and seventh columns the suspect is not a crime stain donor, yet crime stain 1 comes from the suspect), so that the probability distribution over states C_2 and $\bar{C}_2$ is not defined for these events ('n/a' = not applicable). For an alternative way of modeling this conditional probability distribution over the states of node H_2'' that avoids having these impossible combinations in the conditional probability table, we refer the reader to the work by Fenton et al. (2011).*

	H:	D				$\bar{D}$			
	L:	1 donor		2 donors		1 donor		2 donors	
	H_1'':	C_1	$\bar{C}_1$	C_1	$\bar{C}_1$	C_1	$\bar{C}_1$	C_1	$\bar{C}_1$
H_2'':	C_2	1	n/a	0	1	n/a	0	n/a	0
	$\bar{C}_2$	0	n/a	1	0	n/a	1	n/a	1

The resulting hierarchical ordering, from the parent node to the child node, is therefore:

$$H \rightarrow \{H_1'', H_2''\} \rightarrow \{H_1', H_2'\} \ .$$

Our Bayesian network will reflect this hierarchy.

9.5.2 Uncertainty on the number of donors

To take into account the possibility that there was only one donor, we add an additional node L made up of the states *'1 donor'* and *'2 donors'*. We use the parameter λ, denoting the prior probability of *'2 donors'*, to introduce the uncertainty on the number of donors into this node.

The states of node L add a constraint on the probability distribution over $C_1, \bar{C}_1, C_2$, and $\bar{C}_2$, and on the observed profile of crime stain 2 (Y_2) given the profile of crime stain 1 (Y_1). That is, if there is only 1 donor, then Y_2 must be equal to Y_1, and both C_1 and C_2 must be true or false, together, according to whether D is true or false. If there are 2 donors, then either C_1 or C_2 will be true when D is true, but never both C_1 and C_2. In the case of two donors, the parameter τ (denoting $Pr(C_1|2\ \text{donors}, D)$) determines the probability distribution over C_1 and C_2 under proposition D. Tables 9.5 and 9.6 describe the logical relationships between the propositions $C_1, \bar{C}_1, C_2$, and $\bar{C}_2$ and the propositions D and $\bar{D}$ given the number of donors specified in node L.

9.5.3 Relationship between the propositional and evidential nodes

As proposed by Taroni et al. (2006a), the profile of each crime stain depends on whether that particular crime stain comes from the suspect, i.e., on propositions C_1 and C_2. This means that node Y_1 should be connected with a node containing state C_1, and node Y_2 connected with a node containing state C_2. The most straightforward way of achieving this in the model is for Y_1 to be a child of H_1'', and Y_2 a child of H_2''. Thus, Y_1 copies the state of X when C_1 is true, and is independent of X when $\bar{C}_1$ is true (see Table 9.7). The same

Table 9.7: *Probability table for node Y_1 in Fig. 9.2. If C_1 is true, the state of Y_1 is equal to the state of X. If $\bar{C}_1$ is true, the probability of observing each profile is equal to that profile's match probability in the relevant population of possible crime stain donors.*

	H_1'':	C_1			$\bar{C}_1$		
	X:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}
Y_1:	Γ_1	1	0	0	γ_1	γ_1	γ_1
	Γ_2	0	1	0	γ_2	γ_2	γ_2
	Γ_{other}	0	0	1	$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$

principle holds for Y_2 (Table 9.8), with the additional constraint that Y_2 copies the state of Y_1 in every case where both traces come from the same source (defined by node L). Finally, node Y_1Y_2 combines the states of Y_1 and Y_2 as ordered pairs, as proposed by Taroni et al. (2006a) (see Table 9.9). Putting all of these considerations together produces the Bayesian network shown in Fig. 9.2.

There are two ways to use the Bayesian network, which we will illustrate in the next two sections: the user can either update the prior probability distributions over the propositions to posterior probability distributions given the evidence (see Section 9.6), or the user can use the Bayesian network to evaluate the probabilities forming the ratio of the value of the evidence (Eq. 9.1) for a given pair of propositions (see Section 9.7). Both of these are useful means for a forensic scientist to convey the value of the evidence to a fact-finder.

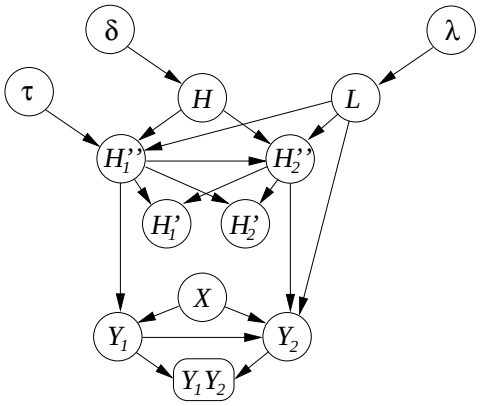


Figure 9.2: *The extended Bayesian network for the two-trace problem. This model is more flexible and realistic than the Bayesian network shown in Fig. 9.1, because it models the uncertainty on the number of crime stain donors, and the uncertainty on which trace comes from the suspect if the suspect is one of two donors. It also includes a node for each of the unknown parameters, allowing the user to define a probability distribution for each. Table 9.1 gives the definitions of the parameters, and Table 9.2 lists the definitions and probabilities of the states in each of the non-parametric nodes.*

9.6 Using the Bayesian network to update the prior probability distribution to a posterior probability distribution

Fact-finders and lawyers are interested in the probability distribution over the propositions given the forensic scientist's evidence. The Bayesian network presented in Fig. 9.2 can compute this posterior probability distribution for a given prior probability distribution over the propositions. There are two applications where a forensic scientist testifying in

Table 9.8: *Probability table for node Y_2 in Fig. 9.2. If L = 1 donor, the state of Y_2 is equal to the state of Y_1. If L = 2 donors, the probability table for node Y_2 is identical to the probability table for node Y_1 (Table 9.7).*

L:	*1 donor*																	
Y_1:	Γ_1						Γ_2											
H_2''	C_2			$\bar{C}_2$			C_2			$\bar{C}_2$								
X:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}						
Y_2: Γ_1	1	1	1	1	1	1	0	0	0	0	0	0						
Γ_2	0	0	0	0	0	0	1	1	1	1	1	1						
Γ_{other}	0	0	0	0	0	0	0	0	0	0	0	0						

1 donor									*2 donors*									
Γ_{other}									Γ_1									
C_2			$\bar{C}_2$						C_2			$\bar{C}_2$						
Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}				
0	0	0	0	0	0	1	0	0	γ_1		γ_1	γ_1		γ_1				
0	0	0	0	0	0	0	1	0	γ_2		γ_2	γ_2		γ_2				
1	1	1	1	1	1	0	0	1	$1-\gamma_1-\gamma_2$		$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$		$1-\gamma_1-\gamma_2$				

2 donors																		
Γ_2									Γ_{other}									
C_2			$\bar{C}_2$						C_2			$\bar{C}_2$						
Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}				
1	0	0	γ_1		γ_1	γ_1		γ_1	1	0	0	γ_1		γ_1				
0	1	0	γ_2		γ_2	γ_2		γ_2	0	1	0	γ_2		γ_2				
0	0	1	$1-\gamma_1-\gamma_2$		$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$		$1-\gamma_1-\gamma_2$	0	0	1	$1-\gamma_1-\gamma_2$		$1-\gamma_1-\gamma_2$				

Table 9.9: *Probability table for node Y_1Y_2. This node combines the states of Y_1 and Y_2 as ordered pairs.*

	Y_1:	Γ_1			Γ_2			Γ_{other}		
	Y_2:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}
Y_1Y_2:	$\Gamma_1\Gamma_2$	0	1	0	0	0	0	0	0	0
	other	1	0	1	1	1	1	1	1	1

court would use the model in this way: (i) when a fact-finder or lawyer interested in the probabilities of the propositions communicates the information required to define the prior distribution to the forensic scientist, (ii) when the forensic scientist wants to illustrate the evidence's effect on several prior probability distributions of different orders of magnitude to show what the posterior probability distribution would be for each.

To specify the prior probability of each proposition, the user must assess the values of δ, λ and τ (see the definitions given in Table 9.1 and in Section 9.4.3). Practically speaking, the user of the model must enter these values into the Bayesian network, an action called 'instantiating' the corresponding nodes. The Bayesian network then propagates these values to the rest of the network.

To find the posterior probability distribution given the evidence, the user instantiates the observed traits for the suspect's sample and the two traces in nodes X, Y_1 and Y_2, respectively. After entering this evidence, the Bayesian network updates the probability distributions in the remaining nodes according to the laws of probability and the probabilistic relationships specified by the model. The probability distributions indicated in the propositional nodes now correspond to the posterior distributions given the evidence. Mathematically, this updating corresponds to the application of the laws of probability, in particular Bayes' theorem. The following numerical example illustrates this concept.

> *Example: Consider a case where crime scene investigators recover two contact stains on a wall, at a given height above the floor: say, crime stain 1 at 1.5 meters, and crime stain 2 at 1.8 meters from the floor. There are no witness statements asserting whether these two traces come from a single source or from two different sources. We assume that it is, a priori, equally probable for the two traces to come from a single source as it is for them to come from two different sources, and thus set* $\lambda = 0.5$. *A suspect, with a prior probability assessed at* $\delta = 0.1$ *of being the source of at least one of the two traces recovered on the crime scene, comes to the attention of the police. This suspect is particular in that he is very short, measuring only 1.6 meters. This information makes a contact between the suspect and the location of crime stain 1 more probable than a contact between the suspect and the location of crime stain 2. In other words, if only one of the two traces comes from this suspect, it is more probable for this stain to be crime stain 1 than crime stain 2. For this reason, we set* $\tau = 0.75$.

The following analysis compares the probability distributions for the three different types of pairs of propositions by considering pairs H, H_1' and H_1''. (Note that the Bayesian network presented in Fig. 9.2 allows for the same analysis with regard to pairs H, H_2' and H_2'', focusing on crime stain 2 instead of on crime stain 1. Here, however, we will focus on crime stain 1.) For this, node H_2' is superfluous in the Bayesian network (Fig. 9.2). In this section, Fig. 9.3 has omitted this node to avoid cluttering the Bayesian network's expanded representations.

Fig. 9.3(a) gives the prior probability distribution over the nodes of the Bayesian network for the above described example. The ratio of the probabilities of the propositions of each

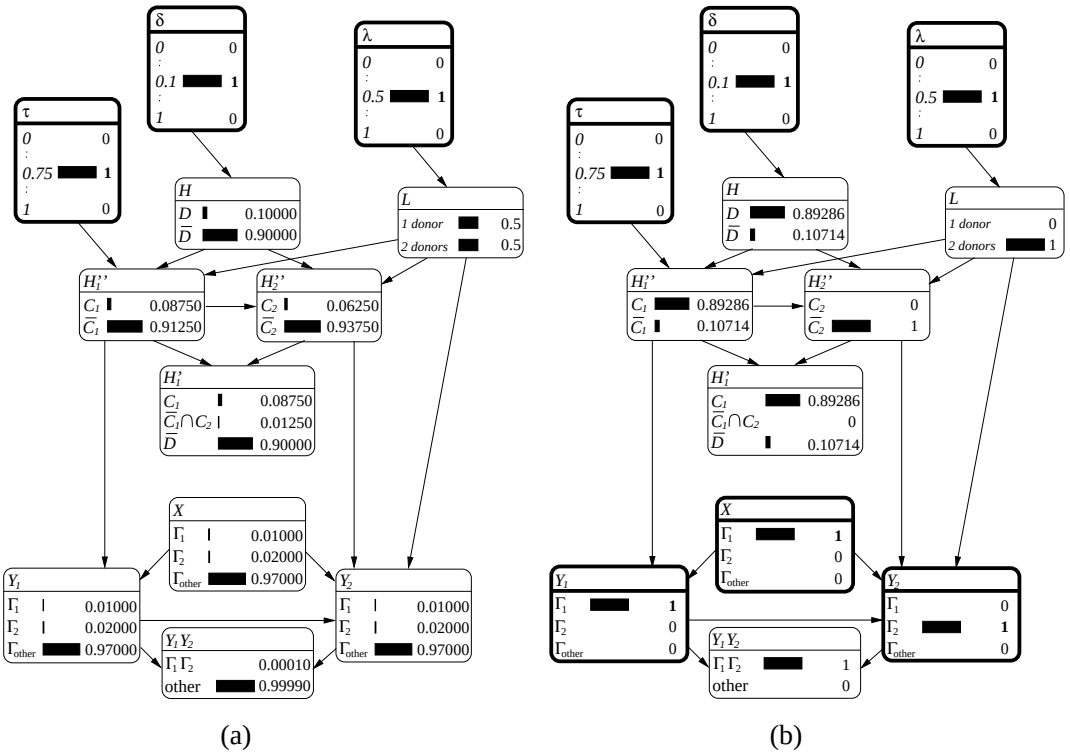


Figure 9.3: *Expanded representation of the Bayesian network presented in Fig. 9.2, without node H_2', which has been omitted to avoid cluttering the figure, and to focus the reader's attention on the probability distributions in nodes H, H_1' and H_1''. This Bayesian network updates* **(a)** *the prior probability distribution over the propositions, to* **(b)** *the posterior probability distribution obtained after observing the traits of X, Y_1 and Y_2. Here, the model is applied to the example described on pages 81 and 83, with $\delta = 0.1$, $\lambda = 0.5$, $\tau = 0.75$, $\gamma_1 = 0.01$ and $\gamma_2 = 0.02$. The observed evidence consists of $X = \Gamma_1$, $Y_1 = \Gamma_1$ and $Y_2 = \Gamma_2$. This information is communicated to the Bayesian network by instantiating the evidential nodes to these observed states. Here, the instantiated nodes are indicated by a thicker border, and the instantiated state with a probability of 1 in bold. Instantiating the evidential nodes produces identical posterior odds for the pairs of propositions H, H_1' and H_1''. Note that in (b), we could also have instantiated node Y_1Y_2 instead of nodes Y_1 and Y_2 and obtained the same outcome.*

pair forms the following prior odds for the three pairs of propositions defined in Section 9.4.2:[4]

$$\frac{\delta}{1-\delta} = \frac{0.1000}{0.9000} \tag{9.12}$$

$$= 0.1111 \quad \text{for pair } H,$$

$$\frac{\delta(1-\lambda+\lambda\tau)}{1-\delta} = \frac{0.0875}{0.9000} \tag{9.13}$$

$$= 0.0972 \quad \text{for pair } H_1', \text{ and}$$

$$\frac{\delta(1-\lambda+\lambda\tau)}{\delta\lambda(1-\tau)+1-\delta} = \frac{0.0875}{0.9125} \tag{9.14}$$

$$= 0.0959 \quad \text{for pair } H_1''.$$

With these numerical calculations, we do not imply that it is possible to attain this level of precision in practice. The precision of the numerical calculations in Eq. (9.12)-(9.14), and in the equations of the rest of this and the next section, is only for the purpose of showing the level of agreement between the Bayesian network's computations and the algebraic equations.

The comparison of the above results with the prior odds given in Meester and Sjerps (2003) (see Eqs. (9.9) - (9.11)) shows that the latter describe a case where $\lambda = 1$ and $\tau = 0.5$. The above expressions relax these assumptions by allowing the user to specify parameters λ and τ so that they reflect the circumstances of the case as accurately as possible.

Comparing the prior odds for each of the pairs of propositions with each other reveals that the most general pair of propositions (pair H) has the greatest odds, and the most specific pair of propositions (pair H_1'') has the smallest odds. This is logical since the prior odds for a specific crime stain cannot be greater than the general prior odds for the suspect being a donor of any one of the crime stains.

> *Example (continued). We now analyze the evidence, and observe* $Y_1 = \Gamma_1$, $Y_2 = \Gamma_2$ *and* $X = \Gamma_1$, *i.e., the suspect's sample matches crime stain 1. In the population of potential sources of the two traces, we assume* $\gamma_1 = 0.01$ *and* $\gamma_2 = 0.02$.

Instantiating the evidential nodes X, Y_1 and Y_2 to their observed traits (Fig. 9.3(b)), produces identical posterior odds of

$$\frac{0.89286}{0.10714} \approx 8.3333$$

for all three pairs of propositions. Algebraically, these posterior odds are given by

$$\frac{\delta\tau}{(1-\delta)\gamma_1} \ . \tag{9.15}$$

The comparison of this ratio with the posterior odds presented in Meester and Sjerps (2003) (see Eq. (9.8)) shows that Eq. (9.15) relaxes the assumption of $\tau = 0.5$, assumed in Eq. (9.8). Eq. (9.15) therefore gives the generalized expression for the posterior odds for any value of τ.

This approach of instantiating the evidential nodes in the Bayesian network is useful whenever one wants to find a posterior probability distribution for a given prior probability distribution. This application is limited to situations where the forensic scientist receives information about the prior probability distribution from an actor in the legal system, or

[4]Note that, by definition, pair H_1' consists of two nonexhaustive propositions. This is not problematic in this situation, because the evidence introduced later on will render the remaining proposition impossible.

Table 9.10: *The mathematical expressions used by the Bayesian network in Fig. 9.2 to compute the prior odds, value of the evidence (Bayes factor) and posterior odds for each of the three pairs of propositions, H, H_1', and H_1''. For the definitions of δ, λ and τ, see Table 9.1.*

for pair H:	$\underbrace{\frac{\delta}{1-\delta}}_{prior\ odds}$	$\times$	$\underbrace{\frac{\tau}{\gamma_1}}_{V}$	$=$	$\underbrace{\frac{\delta\tau}{(1-\delta)\gamma_1}}_{posterior\ odds}$
for pair H_1':	$\underbrace{\frac{\delta(1-\lambda+\lambda\tau)}{1-\delta}}_{prior\ odds}$	$\times$	$\underbrace{\frac{\frac{\tau}{1-\lambda+\lambda\tau}}{\gamma_1}}_{V}$	$=$	$\underbrace{\frac{\delta\tau}{(1-\delta)\gamma_1}}_{posterior\ odds}$
for pair H_1'':	$\underbrace{\frac{\delta(1-\lambda+\lambda\tau)}{\delta\lambda(1-\tau)+1-\delta}}_{prior\ odds}$	$\times$	$\underbrace{\frac{\frac{\tau}{1-\lambda+\lambda\tau}}{\gamma_1\frac{1-\delta}{(1-\tau)\lambda\delta+1-\delta}}}_{V}$	$=$	$\underbrace{\frac{\delta\tau}{(1-\delta)\gamma_1}}_{posterior\ odds}$

situations where the forensic scientist assigns hypothetical prior distributions to illustrate the evidence's effect on the probabilities of the propositions. However, the forensic scientist's role is not to determine the probability distribution over the propositions. The role of the forensic scientist is to evaluate the value of the evidence (e.g., Lindley, 1977a; Aitken and Taroni, 2004). This means that he/she wants to find out to what extent the observed evidence will affect the probability distribution over the propositions, without knowing what this probability distribution is.

In addition to computing the posterior probabilities seen in this section, the Bayesian network allows its user to evaluate the probabilities forming the value of the evidence for any of the three pairs of propositions. We discuss this use of the Bayesian network in the next section.

9.7 Using the Bayesian network to evaluate the value of the evidence

The objective of the forensic scientist's testimony is to present the value of the evidence. That is, he/she should present how much more or less probable the evidence is if the first proposition is true than if the second proposition is true. This value depends on the formulation of the two propositions. The value of the evidence for each pair of propositions corresponds to the Bayes factor obtained by dividing the posterior odds by the prior odds (Table 9.10).

Mathematically, this value is given by Eq. (9.1). Applying the third law of probability for dependent events according to a suspect-anchored perspective (e.g., Aitken and Taroni, 2004) makes this equation equal to

$$
\begin{aligned}
V &= \frac{Pr(X, Y_1, Y_2|\text{proposition 1})}{Pr(X, Y_1, Y_2|\text{proposition 2})} \\
&= \frac{Pr(Y_1, Y_2|X, \text{proposition 1})}{Pr(Y_1, Y_2|X, \text{proposition 2})} \times \frac{Pr(X|\text{proposition 1})}{Pr(X|\text{proposition 2})} ,
\end{aligned}
$$

which reduces to

$$
\begin{aligned}
&= \frac{Pr(Y_1, Y_2|X, \text{proposition 1})}{Pr(Y_1, Y_2|X, \text{proposition 2})} , \qquad (9.16) \\
& \qquad (9.17)
\end{aligned}
$$

given that the profile of the suspect's sample does not change under the competing propositions, i.e., $Pr(X|\text{proposition 1}) = Pr(X|\text{proposition 2})$.

So, to find the value of the evidence, the Bayesian network calculates the probabilities forming the numerator and the denominator of Eq. (9.16). The Bayesian network computes the compound probability of Y_1 and Y_2 in node Y_1Y_2. This node indicates the numerator of V for the observed traits of Y_1 and Y_2 when X and 'proposition 1' are instantiated, and the denominator of V when X and 'proposition 2' are instantiated.[5] Fig. 9.4, 9.5 and 9.7 illustrate the results obtained in this way for each of the three pairs of propositions, H, H_1' and H_1'', for the numerical example presented in Section 9.6. Again, the expanded representations of the Bayesian network omit node H_2' to avoid cluttering these figures, and to focus the reader's attention on the propositional nodes H, H_1' and H_1''. In the following sections, we discuss each value in turn, and examine how each is affected by the parameters τ, λ and δ.

9.7.1 The value of the evidence for pair H

According to Fig. 9.4, the value of the evidence for pair H is equal to

$$
\begin{aligned}
V &= \frac{0.0075}{0.0001} \\
&= 75 .
\end{aligned}
$$

Algebraically, this value is given by

$$
\begin{aligned}
V &= \frac{\lambda\tau\gamma_2}{\lambda\gamma_1\gamma_2} \\
&= \frac{\tau}{\gamma_1} . \qquad (9.18)
\end{aligned}
$$

The numerator describes the probability of observing the evidence given that at least one of the crime stains comes from the suspect (proposition D). In this case, the observation of the evidence is only possible when the two traces come from two different donors (for which the probability is λ), of which the suspect is the donor of the first trace (probability τ), and someone with trait Γ_2 the donor of the second trace (probability γ_2). The denominator describes the probability of observing the evidence given that neither of the crime stains comes from the suspect (proposition $\bar{D}$). In this case, the observation of the evidence corresponds to the event that the two traces come from two different donors (probability λ),

[5] Note that the Bayesian network presented here models the probability of Y_1 and Y_2 as independent of the suspect's sample given 'proposition 2'. This makes the probability of Y_1 and Y_2 when 'proposition 2' and X are instantiated identical to the probability of Y_1 and Y_2 when only 'proposition 2' is instantiated, so that the instantiation of X is not absolutely necessary in this case.

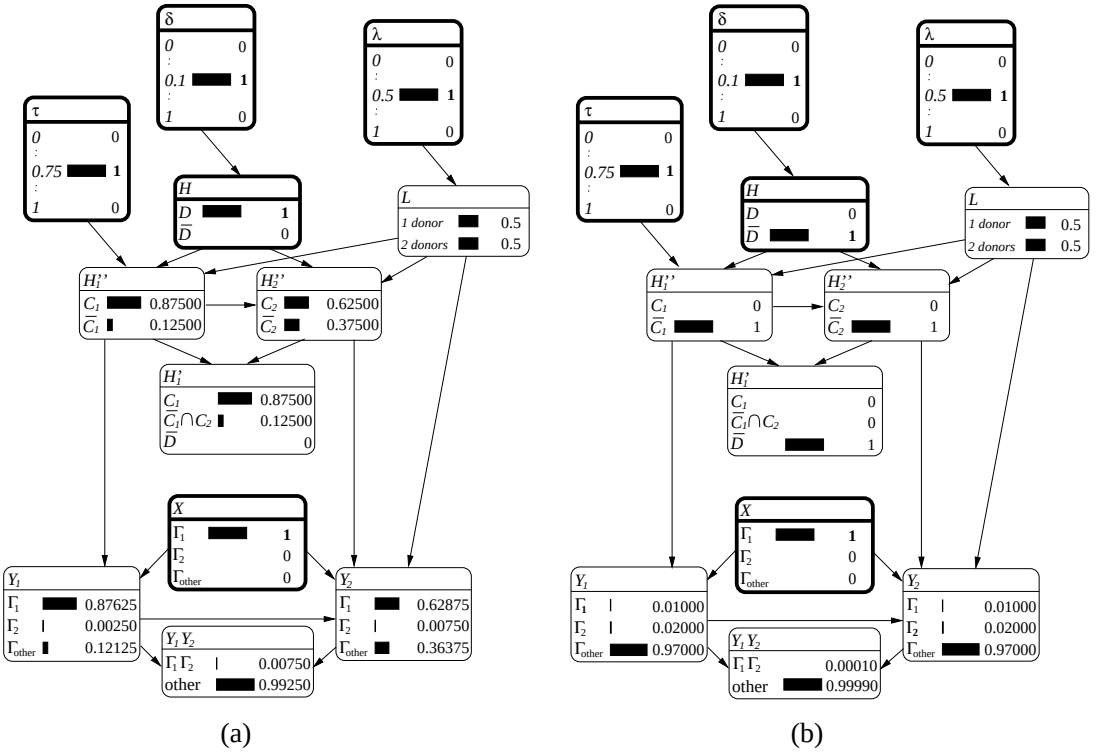


Figure 9.4: *The Bayesian network computes (***(a)** *the numerator, and* **(b)** *the denominator of the value of the evidence (Eq. (9.16)) for pair of propositions H, for evidence consisting of $X = \Gamma_1$, $Y_1 = \Gamma_1$ and $Y_2 = \Gamma_2$. The instantiated nodes are indicated by a thicker border, and the instantiated state with a probability of 1 in bold. The numerator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and D is instantiated in node H. The denominator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and $\bar{D}$ is instantiated in node H. The calculations are for the example described in Section 9.6 ($\delta = 0.1$, $\lambda = 0.5$ and $\tau = 0.75$).*

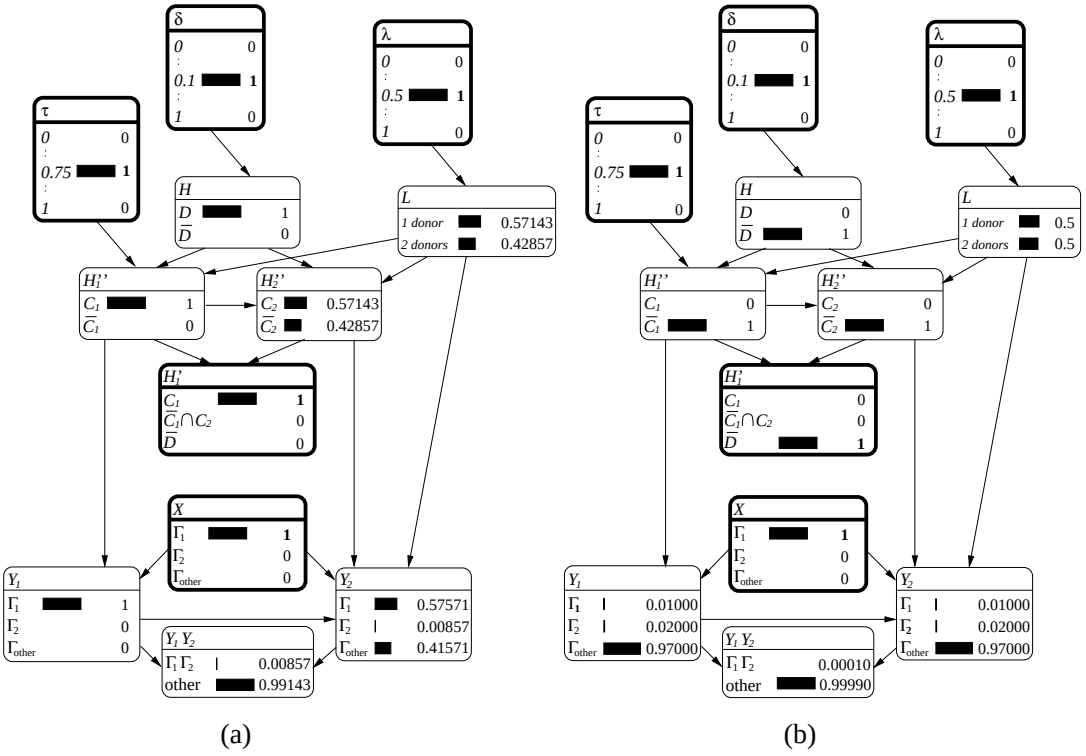


Figure 9.5: *The Bayesian network computes* **(a)** *the numerator, and* **(b)** *the denominator of the value of the evidence (Eq. (9.16)) for pair of propositions H_1', for evidence consisting of $X = \Gamma_1$, $Y_1 = \Gamma_1$ and $Y_2 = \Gamma_2$. The nodes with the thicker borders are the instantiated nodes, with the instantiated state indicated with a probability of 1 in bold. The numerator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and C_1 is instantiated in node H_1', and the denominator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and $\bar{D}$ is instantiated in node H_1'. These calculations are for the scenario described in Section 9.6 ($\delta = 0.1$, $\lambda = 0.5$ and $\tau = 0.75$).*

of which one donor has trait Γ_1 (probability γ_1) and the other donor trait Γ_2 (probability γ_2).

For this pair of propositions, V reduces to a linear function of τ, ranging from a minimum of 0 when $\tau = 0$ (i.e., when it is a priori impossible for the suspect to be the source of trace 1 in a case where the suspect is the source of one of the two traces), to a maximum of $\frac{1}{\gamma_1}$ for $\tau = 1$ (i.e., when it is a priori certain that the suspect is the source of trace 1 in the case that the suspect is the source of one of the two traces). In the latter case, the value of the evidence is the same as in a one-trace problem, because, just as in a one-trace problem, it becomes certain to observe a match between crime stain 1 and the suspect's sample if proposition D is true.

When $\tau = 0.5$, this means that it is equally likely for either of the two traces to come from the suspect in a case where the suspect is one of two crime stain donors. This is the additional factor multiplied by $\frac{1}{\gamma_1}$ to produce Eq. (9.4) derived by Evett (1987) for the value of one matching stain and one non-matching stain. Underlying Eq. (9.4) is therefore the assumption that each of the two traces is equally likely to come from the suspect if the suspect is the source of one of the two traces. Yet, as seen in the example on page 81, the two crime stains may not have the same prior probability of coming from the suspect if they were recovered at different locations on the crime scene. In this case, it is necessary to replace Eq. (9.4) with Eq. (9.18), and assign a more adequate value for τ based on the circumstances of the case.

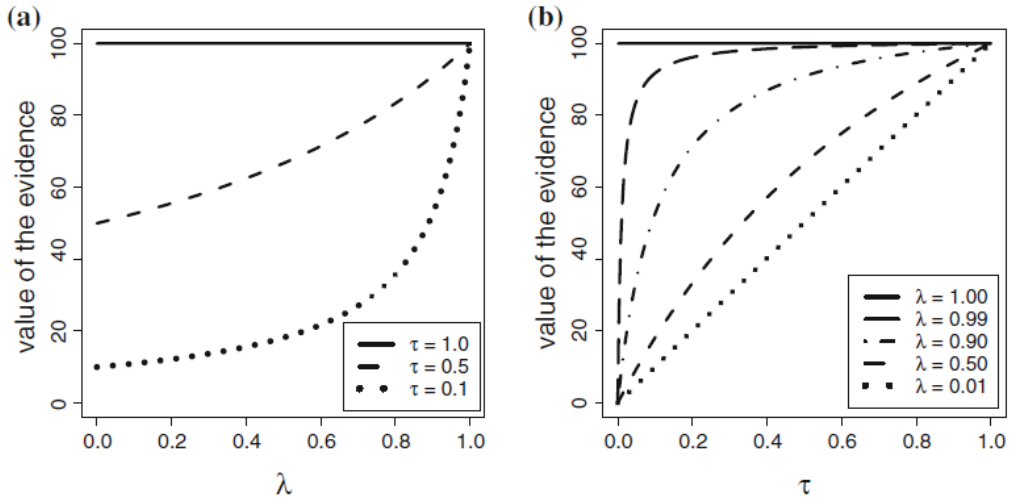


Figure 9.6: *The value of the evidence V for pair H_1' as a function of* **(a)** λ *and* **(b)** τ*. Here,* $\gamma_1 = 0.01$*. V is an increasing function of λ and τ, attaining a maximal value of $\frac{1}{\gamma_1}$ when $\lambda = 1$ or $\tau = 1$, for $\lambda \neq 0$, $\tau \neq 0$.*

9.7.2 The value of the evidence for pair H_1'

For pair H_1', Fig. 9.5 shows that the value of the evidence is equal to

$$V = \frac{0.00857}{0.0001}$$

$$= 85.7 \ ,$$

which, algebraically, corresponds to

$$V = \frac{\frac{\lambda\tau}{1-\lambda+\lambda\tau}\gamma_2}{\lambda\gamma_1\gamma_2} \tag{9.19}$$

$$= \frac{\frac{\tau}{1-\lambda+\lambda\tau}}{\gamma_1} \ . \tag{9.20}$$

Here, the probability in the numerator is the probability of observing the evidence given that crime stain 1 comes from the suspect (proposition C_1). A priori, there are two possibilities if proposition C_1 is true: either both traces come from the suspect (for which the probability is $1-\lambda$), or only crime stain 1 comes from the suspect (for which the probability is $\lambda\tau$). The observation of one matching and one non-matching trace is impossible if both traces come from the suspect. The probability of observing the evidence is therefore the normalized probability for the latter case times the probability that the donor of the second trace has trait Γ_2 (probability γ_2), as shown in the numerator of Eq. (9.19). The denominator remains the same as for pair H.

For this pair of propositions, V is an increasing function of both τ and λ for all $\tau < 1$ and $\lambda < 1$ (Fig. 9.6). V attains the maximum value of $\frac{1}{\gamma_1}$ when at least one of these parameters is equal to 1:

- When $\lambda = 1$, it is certain that the two traces come from two different sources. In this case, it is, a priori, certain that crime stain 2 does not come from the suspect given proposition C_1, and the only possibility left under this proposition is that the suspect is the source of only crime stain 1. The normalized probability in the numerator of Eq. (9.19) therefore reduces to 1, so that the probability of observing the evidence given C_1 and $X = \Gamma_1$ is equal to $1 \times \gamma_2$. With the denominator, this reduces the numerator of V to 1.

- When $\tau = 1$, it is a priori certain that trace 1 comes from the suspect if exactly one of the traces comes from the suspect. The normalized probability in the numerator of Eq. (9.19) thus reduces to λ. The probability of observing the evidence is therefore equal to the probability that the two traces come from two different sources times the probability that the other donor has trait Γ_2, i.e., $\lambda \times \gamma_2$. With the denominator, this reduces the numerator of V to 1.

According to the reasoning in Meester and Sjerps (2003), Eq. (9.6) was obtained for $\lambda = 1$ and $\tau = 0.5$. In this case, V is maximum for this pair of propositions because of $\lambda = 1$. However, the assumption of $\lambda = 1$ can only be made in very specific cases. By definition, this assumption must be made *before* observing the evidence, so it can only be based on other information in the case. For example, one could imagine a case with a crime scene in a location under high surveillance and cleaned on a regular and scheduled basis: here, a surveillance camera showing two unidentifiable individuals on the scene on the day the traces were deposited, where individual 1 was only present in the location of the recovery of crime stain 1, and individual 2 only in the location of the recovery of crime stain 2, might justify an assumption of $\lambda = 1$. Other than these very particular circumstances, it is difficult to imagine a scenario where such an unmitigated assumption could be made.

To justify an assumption of $\tau = 1$, the circumstances must be just as particular. In this case, they must be such that they make it impossible for the suspect to be the source of only crime stain 2. This could be the case when it is physically impossible for the suspect to have been in contact with the surface of crime stain 2. However, even in these cases it is difficult to justify $\tau = 1$ for DNA traces in situations where secondary transfer is possible (e.g., Goray et al., 2010).

In most cases, V will therefore be less than $\frac{1}{\gamma_1}$. For $\tau = 0.5$, the exact value will lie somewhere on the dashed curve of Fig. 9.6(a) below the maximum point at $\lambda = 1$.

According to Fig. 9.6(a), the range of values obtained for V for different values of λ is smaller for high values of τ. This is because a large value of τ leads to a high prior probability that trace 1 comes from the suspect, regardless of whether there was one donor or two donors. That is, if the suspect was the only donor, then it is certain that trace 1 comes from the suspect, and if the suspect was one of two donors, then the prior probability that trace 1 comes from the suspect ($= \tau$) is also high. Therefore a large value for τ leads to a high probability in the numerator of V (Eq. (9.20)), regardless of the value of λ.

This is no longer the case for small values of τ. If τ is small, the prior probability that trace 1 comes from the suspect will be determined mostly by the probability that both traces come from the suspect, i.e., $1-\lambda$. The numerator of V (Eq. (9.20)) will therefore vary greatly according to the value of λ. The greater λ, the smaller the prior probability of a single donor. Since the evidence is such that it rejects the hypothesis of a single donor, the probability of observing the evidence given proposition C_1 (i.e., the numerator of V) is greater when the prior probability of a single donor is small. That is, a small prior probability for a single donor increases the normalized probability of the event that only crime stain 1 comes from the suspect, figuring in the numerator of V. Thus, the overall value of the evidence is an increasing function of λ.

Fig. 9.6(b) shows that the range of values obtained for V for different values of τ remains 0 to $\frac{1}{\gamma_1}$, regardless of the value of λ. This is because the evidence (a match with crime stain 1 and a non-match with crime stain 2) is such that its value will always be 0 in a case where it is impossible for crime stain 1 to come from the suspect, given that there were two different donors (i.e., when $\tau = 0$), and equal to $\frac{1}{\gamma_1}$ whenever it is certain that crime stain 1 comes from the suspect, given that there were two different donors (i.e., when $\tau = 1$). Thus, the value of the evidence is an increasing function of τ.

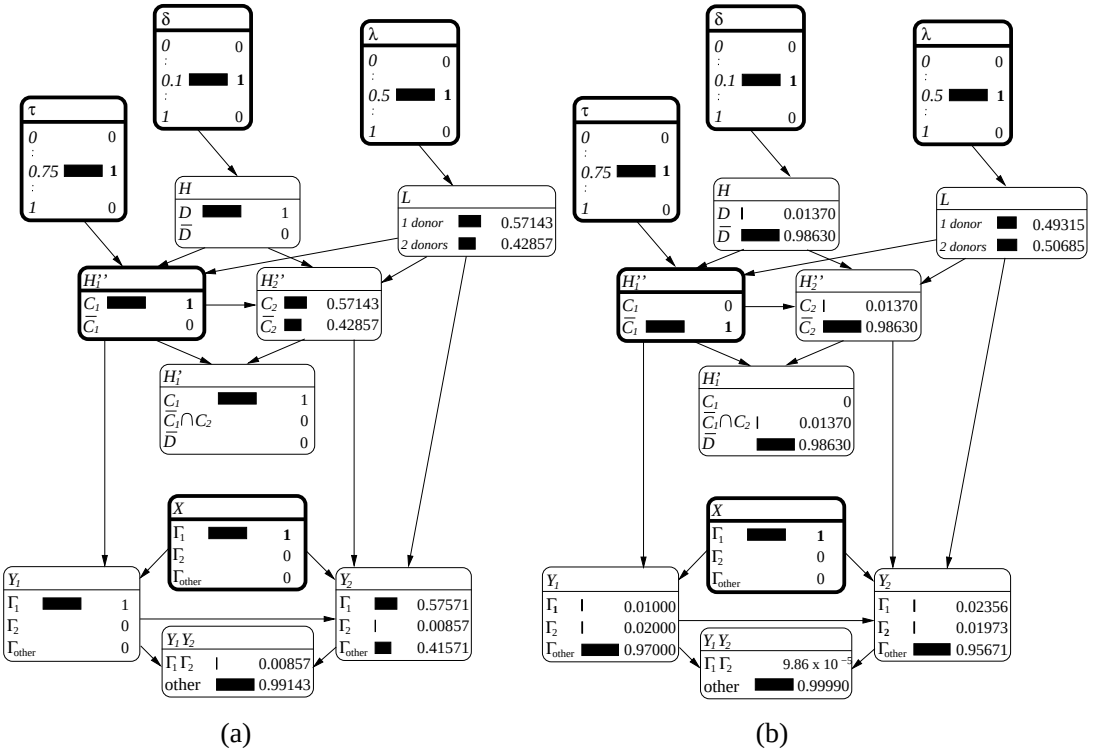


Figure 9.7: *The Bayesian network computes* **(a)** *the numerator and* **(b)** *the denominator of the value of the evidence (Eq. (9.16)) for pair of propositions H_1'', for evidence consisting of $X = \Gamma_1$, $Y_1 = \Gamma_1$ and $Y_2 = \Gamma_2$. The nodes with the thicker borders are the instantiated nodes, with the instantiated state indicated with a probability of 1 in bold. The numerator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and C_1 is instantiated in node H_1'', and the denominator is the probability of $\Gamma_1\Gamma_2$ in node Y_1Y_2 when Γ_1 is instantiated in node X and $\bar{C}_1$ is instantiated in node H_1''. These calculations are for the scenario described in Section 9.6 ($\delta = 0.1$, $\lambda = 0.5$ and $\tau = 0.75$).*

9.7.3 The value of the evidence for pair H_1''

According to Fig. 9.7, the value of the evidence for pair H_1'' is equal to

$$\begin{aligned} V &= \frac{0.00857}{9.86 \times 10^{-5}} \\ &= 86.9 \ , \end{aligned}$$

which is computed by

$$V = \frac{\frac{\lambda\tau}{1-\lambda+\lambda\tau}\gamma_2}{\lambda\gamma_1 \frac{1-\delta}{(1-\tau)\lambda\delta+1-\delta}\gamma_2} \tag{9.21}$$

$$= \frac{\frac{\tau}{1-\lambda+\lambda\tau}}{\gamma_1 \frac{1-\delta}{(1-\tau)\lambda\delta+1-\delta}} \tag{9.22}$$

Here, the probability in the numerator is the same as for pair H_1'. The probability in the denominator is the probability of observing the evidence given that crime stain 1 does not come from the suspect (proposition $\bar{C}_1$). A priori, there are two possibilities if proposition $\bar{C}_1$ is true: either the suspect is only the source of crime stain 2 [for which the probability is $(1-\tau)\lambda\delta$], or neither of the two traces comes from the suspect (for which the probability is $1-\delta$). The observation of the evidence is only possible in the latter case. Therefore, the probability of the evidence is the probability that the two traces come from two different donors, of which one has trait Γ_1 and the other trait Γ_2, i.e., $\lambda\gamma_1\gamma_2$, times the normalized probability that neither of the traces comes from the suspect (Eq. (9.21)).

For this pair of propositions, V is a function of τ, λ and δ (Fig. 9.8). Just like for pair H_1', V is equal to $\frac{1}{\gamma_1}$ whenever $\tau = 1$. In this case, the numerator of V reduces to $\lambda\gamma_2$ as explained above for pair H_1', and the denominator of V becomes equal to $\lambda\gamma_1\gamma_2$, because the possibility of the suspect being the source of crime stain 2 when there are two different crimes stain donors becomes impossible. With the numerator, the denominator of V therefore reduces to γ_1.

Yet, unlike for pair H_1', $\lambda = 1$ no longer produces $V = \frac{1}{\gamma_1}$ (e.g., Fig. 9.8(a), 9.8(c) and 9.8(e)). This is because $\lambda = 1$ (i.e., there were two different donors) does not, a priori, exclude the possibility that the suspect is the source of the second trace given that crime stain 1 does not come from the suspect (proposition $\bar{C}_1$). For $\lambda = 1$, V is actually a decreasing function of τ, attaining a minimum of $\frac{1}{\gamma_1}$ when $\tau = 1$ (Fig. 9.8(b), 9.8(d) and 9.8(f)). This is because the possibility of the suspect being only the source of crime stain 2 becomes less probable as τ increases, thus increasing the normalized probability of the event that neither of the traces comes from the suspect. This increases the denominator of V, and decreases the whole value of the evidence. However, when $\tau \to 0$, $\tau \neq 0$, the probability of the suspect being only the source of crime stain 2 increases, which decreases the normalized probability of neither trace coming from the suspect. This decreases the denominator of V, and increases the whole value of the evidence. Thus the maximum of V for this pair of propositions is *greater* than $\frac{1}{\gamma_1}$ (which is the maximum value for the other two pairs of propositions):

$$\text{when} \quad \tau \to 0 \quad \text{and} \quad \lambda = 1, \qquad V \to \frac{1}{\gamma_1(1-\delta)} \ . \tag{9.23}$$

In other words, for $\tau < 1$, the possibility that the suspect is the source of crime stain 2 is not excluded. Yet, if the suspect is not the source of crime stain 1 (proposition $\bar{C}_1$), the evidence is only possible when neither stain comes from the suspect (probability of $1-\delta$), so that the factor $1-\delta$ has an increasing influence in the denominator of V for $\tau \to 0$, $\tau \neq 0$.

This effect becomes more pronounced as δ increases (Fig. 9.8(b), 9.8(d) and 9.8(f)). A larger value of δ produces a smaller probability in the denominator of V, and therefore a greater value of V.

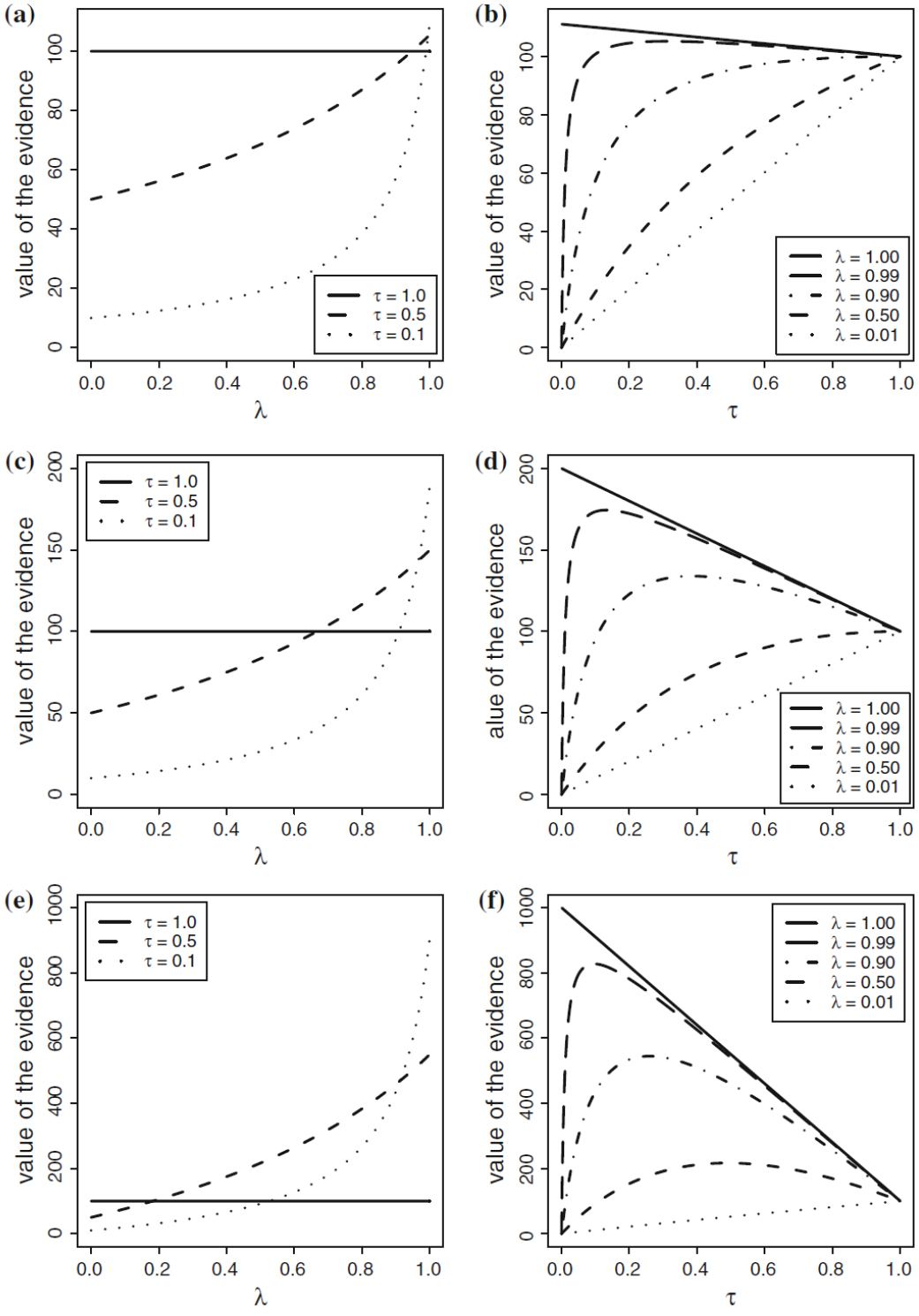


Figure 9.8: *The value of the evidence V for pair H_1'' in function of:* **(a)** λ *for* $\delta = 0.1$, **(b)** τ *for* $\delta = 0.1$, **(c)** λ *for* $\delta = 0.5$, **(d)** τ *for* $\delta = 0.5$, **(e)** λ *for* $\delta = 0.9$, *and* **(f)** τ *for* $\delta = 0.9$. *Here,* $\gamma_1 = 0.01$. *V is an increasing function of λ and δ, and equal to $\frac{1}{\gamma_1}$ for $\tau = 1$. It tends towards a maximum of $\frac{1}{\gamma_1(1-\delta)}$ for $\lambda = 1$ and $\tau \to 0$, $\tau \neq 0$.*

The value of the evidence proposed by Meester and Sjerps (2003) for this pair of propositions (Eq. (9.7)) is equal to

$$\frac{1}{\gamma_1 \frac{1-\delta}{\frac{1}{2}\delta+1-\delta}} \ .$$

Again, this value assumes $\lambda = 1$ and $\tau = 0.5$. Its application is therefore just as limited by the assumption $\lambda = 1$ as the value of the evidence they propose for pair H_1' (see page 89). This value is the point at $\tau = 0.5$ on the solid lines in Fig. 9.8(b), 9.8(d) and 9.8(f). With $\lambda < 1$, V would be smaller, lying on one of the other lines in these graphs.

9.7.4 Comparison of the values of the evidence

As Meester and Sjerps (2003) concluded, pairs of differently formulated propositions for the two-trace problem lead to different values of the evidence. For the example presented, the value of the evidence is greatest for pair H_1'', and smallest for pair H. This is because the probability of observing a match with stain 1 and a non-match with stain 2 is greatest given proposition C_1 and smallest given proposition $\bar{C}_1$.

The derived formulae for calculating the value of the evidence show that this value is a function of τ for all three pairs of propositions, a function of λ for two of the three pairs (pairs H_1' and H_1''), and a function of δ for one pair (pair H_1''). In the two-trace problem, the value of the evidence is therefore not based solely on the analytical results provided by the laboratory analyses of the collected evidence, that is, on the match probabilities of these results in the relevant population of possible sources. In addition, the value depends on parameters assessed on the basis of the case circumstances prior to observing the evidence. The more specific the competing pair of propositions are, the more parameters will determine the value of the evidence for these propositions. That is, propositions focusing only on one of the two traces require additional information regarding the total number of donors on the crime scene and/or the prior assumption on the suspect's implication as a donor of any of the traces on the scene. To accurately evaluate the value of the evidence in a two-trace problem, an evaluator's knowledge must therefore extend beyond the observations made on the evidence, to the facts regarding the case circumstances.

9.8 Discussion and conclusions

The role of the forensic scientist is to evaluate the value of the evidence (e.g., Evett, 1998). In the forensic two-trace problem, this has been somewhat perplexing since three different formulations of the competing pair of propositions lead to three different quantifications of this value (Meester and Sjerps, 2003).

In this paper, we have provided a more general vision of the entire two-trace problem by constructing a Bayesian network that includes each of the three pairs of propositions as a separate node in the model. Through an illustrative example, we demonstrate how to use the network to evaluate the value of the evidence for each pair of propositions. The different structural relationships between each of the pairs and the evidence inevitably leads to different values of the evidence, each addressing the two-trace problem from a different angle.

The flexibility of the value of the evidence to adapt to each pair of propositions is an advantage, not an inconvenience. A forensic scientist's task is to evaluate the relative support provided by the evidence for one proposition with respect to an alternative proposition (i.e., the value of the evidence) (e.g., Evett, 1998). And this, he/she must do with regard to the very particular framework of circumstances that reflects the case, and for the precise propositions of interest to the court. Therefore, it is important that the propositions be chosen and formulated with care, and that these be based on the particular circumstances related

to the case. The different values of the evidence then complement each other, providing the scientist with a range of formulae from which he/she can select the most appropriate in view of the pair of propositions of interest to the court.

The crucial issue is to understand what assumptions lie behind each formula, in order to correctly use it in the context of the case. In this respect, the Bayesian network offers transparency through its graphical representation of the dependence relationships among the variables. In particular, it models the dependency of each of the random variables on three unknown parameters:

- δ, the probability that at least one trace comes from the suspect,
- λ, the probability that the two traces come from two different donors (Dawid, 2004), and
- τ, the probability that trace 1 comes from the suspect in a case where the suspect is one of two different donors.

The value of the evidence is a function of τ for all three pairs of propositions, a function of λ for the two pairs where the prosecutor's proposition relates only to one trace, and (as presented in Meester and Sjerps (2003)) a function of δ for the pair where the defence's proposition relates only to one trace. To accurately evaluate the value of the evidence, an evaluator is therefore obliged to have information on the case circumstances. If it is difficult to obtain precise assessments for the unknown parameters, the Bayesian network environment allows the user to specify subjective probability distributions over each parameter space.

Note that the model presented in this paper is still based on several assumptions, notably on the independence between the three unknown parameters. The validity of this assumption will depend on the circumstantial information available in a case, and on the evaluator's personal assessments of the parameters. The results of this work justify a careful examination and further study on the dependence relationships between these parameters in cases where the assumption of independence no longer holds.

Notwithstanding, the major advantage of using the Bayesian network is when the evidence of the two traces must be combined with other types of evidence. The fundamental structure of this Bayesian network allows for an extension to more than two traces, as well as an extension to address activity level propositions (Gittelson et al., 2012a). Thanks to its graphical architecture, this model can be inserted as a component part in a larger network for a more complex inference problem. Given that most forensic cases involve numerous traces of different types of evidence, this possibility is an indispensable property for all practical applications. The generic Bayesian network presented in this paper therefore offers a transparent and practical tool for tackling two-trace problems in forensic casework.

Chapter 10

Bayesian Networks and the Value of the Evidence for the Forensic Two-trace Transfer Problem

Abstract

This paper presents a graphical approach for developing a likelihood ratio for complex forensic inference problems: an approach based on the construction of Bayesian networks (BNs). Here, we used this approach to frame the two-trace transfer problem for activity level propositions. In this problem, a suspect's sample matches one of two traces recovered on a crime scene, and the forensic scientist evaluates the value of this evidence with respect to the propositions of whether or not the suspect was engaged in a struggle with the victim. The evaluation of this scenario with BNs provides a transparent approach to the problem, and suggests a means for coherently addressing more complex challenges, such as generalizing this scenario to n traces.

A dead body is found in a public park. The medical examination of the body reveals signs of a physical struggle. On the crime scene, forensic investigators recover two items of a certain category of trace evidence, say, for example, bloodstains. Trace 1 was found in what we will call location 1, and trace 2 in location 2. These are both inside the perimeter of the crime scene.

The investigation is able to acquire images from two surveillance cameras which filmed different parts of the crime scene during the time lapse the crime was committed. One camera filmed location 1, and the other location 2. Each camera has an image showing the victim struggling with an assailant: the image from the first camera shows the victim and an assailant in location 1, the image from the second camera shows the victim and the assailant in location 2. Unfortunately, the images are of poor quality, and a comparison does not allow one to conclude with certainty whether it is the same assailant in both locations, or whether there were two assailants, one in location 1 and the other in location 2.

Reprinted from the Journal of Forensic Sciences, 57, S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni, Bayesian networks and the value of the evidence for the forensic two-trace transfer problem, pages 1199-1216, Copyright 2012 American Academy of Forensic Sciences, with permission from John Wiley and Sons.

A forensic laboratory analyzes an intrinsic characteristic of the two traces (e.g., the blood group or DNA profile). A comparison of these analytical results with the result obtained from the victim's sample allows the forensic scientists to exclude the victim as the source of both of these traces. The results show that the traces are of different types, say, for example, that trace 1 is of type Γ_1 and trace 2 of type Γ_2.

Later in the investigation, a suspect's sample matches one of these two traces. The evidence against this suspect thus consists of a combination of a matching item and a non-matching item. The question is, "How strong is this evidence against the suspect?"

In forensic science, this scenario is known as the two-trace transfer problem. The answer to the above question takes the form of a likelihood ratio (LR). This LR opposes two mutually exclusive propositions (where the word *proposition* designates a formal statement about an event that the forensic scientist formulates based on the circumstantial information of the case, as described by Evett et al. (2000b)):

$$LR = \frac{Pr(\text{evidence}|\text{proposition 1}, I)}{Pr(\text{evidence}|\text{proposition 2}, I)} \ .$$

The LR is the ratio of the probabilities of observing the evidence given each of the competing propositions and I. The evidence consists of the characteristics of the two traces and the suspect's sample, and the letter I denotes the background information. The background information consists of all the knowledge and circumstances that influence the numerical evaluations of the probabilities forming the numerator and the denominator. Note that all probabilities in this study are conditional on the background information I, yet, for the sake of brevity, we shall hereafter omit I from their notation.

For this scenario, the forensic scientists can formulate the following pair of source level propositions (we use the superscript s to indicate that these are source level propositions):

H^s—*one of the traces on the crime scene comes from the suspect;*
$\bar{H}^s$—*neither of the traces on the crime scene comes from the suspect;*

or the following pair of activity level propositions (we use the superscript a to indicate that these are activity level propositions):

H^a—*the suspect was engaged in a struggle with the victim;*
$\bar{H}^a$—*the suspect was not engaged in a struggle with the victim.*

The first pair of propositions is at the source level in the hierarchy of propositions (Cook et al., 1998b), because it is only concerned with the origin of the traces, that is, the item or person from which the traces come from. This pair ignores *how* and *when* the traces were transferred from their origin to the crime scene. A match between the profile of the suspect's sample and one of the traces, say for example trace 1, produces an LR supporting proposition H^s, regardless of the extrinsic characteristics of the traces (e.g., the quantity of material on the crime scene, or how fresh the trace was on the crime scene). If the trace is an old bloodstain, that was already present on the crime scene before the victim was assaulted, then this evidence will support H^s with an LR equal to $\frac{1}{2\gamma_1}$ (Evett, 1987), just as in the case where the bloodstain was transferred to the crime scene during the assault. This LR only takes into account γ_1, which denotes the probability of obtaining a match with trace 1's analytical characteristic Γ_1 (the matching characteristic, in this case) in the relevant population (Weir, 2000). For this reason, the LR for this pair of propositions only provides us with a very limited amount of information, which does not tell us whether the suspect was an assailant who struggled with the victim. To address the question of whether the suspect was an assailant, we must use the second pair of propositions.

The second pair of propositions is formulated at the activity level in the hierarchy of propositions (Cook et al., 1998b), because it describes the activity or action of interest to the case, that is, the activity or action that may have caused the transfer of the traces from the assailant to the crime scene. In addition to taking into account the origin of the traces,

these propositions also consider *how* and *when* the traces came onto the crime scene. An activity level LR thus consists of transfer, background and match probabilities (Evett, 1984). To differentiate the probabilities referring to trace 1 from those referring to trace 2, we use the subscript $i \in \{1, 2\}$ for all of the probabilities referring to trace i and location i:

- *Transfer probabilities* t_i, $i = 1, 2$, describe how probable it is for trace i to have been transferred during the alleged action, have persisted on the crime scene, and then to have been recovered by the investigators. The complement of t_i is $\bar{t}_i = 1 - t_i$.
- *Background probabilities* b_i, $i = 1, 2$, represent the probability for a trace to be present on the crime scene at location i as a consequence of another transfer event, unrelated to the alleged action. The probability of the absence of such a trace is $\bar{b}_i = 1 - b_i$.
- *Match probabilities* γ_i, $i = 1, 2$, are the probabilities for obtaining a match with characteristic Γ_i (in our scenario, we call trace 1's characteristic Γ_1 and trace 2's characteristic Γ_2, such that the subscript i here corresponds with the notation for trace i, $i \in \{1, 2\}$) in the relevant population. For a trace transferred during the struggle, the relevant population is that of the possible assailants. For a background trace, however, the relevant population is the population of background traces. To distinguish these two from each other, the match probability of a characteristic in the population of possible assailants is γ_i, and the match probability of a characteristic in the population of background traces is denoted γ_i'.

Thus, we have, for example, $\bar{b}_1$ denoting the probability that there was no background trace present at location 1, and t_2 denoting the probability that a trace was transferred to location 2 during the struggle between the assailant and the victim at that location, persisted there, and was recovered during the investigation. For the sake of simplicity, we will assume in the rest of the paper that the traces transferred during the alleged action all persisted on the crime scene and were all recovered by the investigators. Therefore, we will simply refer to these traces as traces that were transferred during the struggle.

Using an algebraic approach, previous authors (Triggs and Buckleton, 2003) came up with the following formula for evaluating this activity level LR:

$$LR = \frac{\frac{1}{2}\bar{b}_1\bar{b}_2t_1t_2(1-2q)\gamma_2 + \frac{1}{2}\bar{b}_1b_2t_1\bar{t}_2(1+\gamma_1)\gamma_2' + \frac{1}{2}b_1\bar{b}_2\bar{t}_1t_2\gamma_1'\gamma_2 + b_1b_2\bar{t}_1\bar{t}_2\gamma_1'\gamma_2'}{\bar{b}_1\bar{b}_2t_1t_2(1-2q)\gamma_1\gamma_2 + \bar{b}_1b_2t_1\bar{t}_2\gamma_1\gamma_2' + b_1\bar{b}_2\bar{t}_1t_2\gamma_1'\gamma_2 + b_1b_2\bar{t}_1\bar{t}_2\gamma_1'\gamma_2'}. \quad (10.1)$$

This equation combines the above described transfer, background and match probabilities. In addition, it contains the expression $1 - 2q$ to describe the probability that the transferred traces come from two different assailants (Triggs and Buckleton, 2003). That is, these authors assumed that there were two assailants and defined the probability that two transferred traces both come from assailant 1 as q, and the probability that two transferred traces both come from assailant 2 as q. This led to a probability of $2q$ that two transferred traces come from the same assailant, and to a probability of $1 - 2q$ that two transferred traces come from different assailants.

As for the background probabilities, note that in this previous study (Triggs and Buckleton, 2003), a single variable p was used to describe the background probabilities of both traces together, such that $\bar{b}_1 \times \bar{b}_2 = p_0$, $\bar{b}_1 \times b_2 = p_1^2$, $b_1 \times \bar{b}_2 = p_1^1$, and $b_1 \times b_2 = p_2^{1,2}$. Here a b is used instead of p to be able to compare the formula with other activity level formulae figuring in later sections.

The authors of (Triggs and Buckleton, 2003) developed Eq. (10.1) using an algebraic approach. That is, they considered four mutually exclusive transfer events to explain the evidence: the product of an appropriate combination of the transfer, background, and match probabilities depicts each of these transfer events, and the sum of these four products (one for each possible event) forms the numerator and the denominator of this LR.

10.1 Aim and outline of this study

Despite the valuable formal rigor of such an approach, it reaches its limits when applied to increasingly complex inference problems: either it will make simplifying assumptions that ignore probabilistic dependencies between the variables, or the mathematical development of the formula becomes so intricate that it is no longer transparent to non-statisticians (such as lawyers, prosecutors and judges). With regard to this issue, the aim of this study is to investigate a new approach for developing an LR formula for complex forensic inference problems: a graphical approach based on the construction of Bayesian networks (BNs). BNs have already proven to be practical tools for portraying inference problems in forensic science (e.g., (Aitken and Gammerman, 1989; Dawid et al., 2002; Mortera et al., 2003; Biedermann et al., 2005a,b; Taroni et al., 2006a; Biedermann and Taroni, 2006; Biedermann, 2007; Dawid et al., 2007; Hepler et al., 2007; Biedermann et al., 2008b) to name a few). Yet, up to now, forensic statisticians have used BNs to reproduce existing LR formulae, formulae developed through algebraic calculations. Instead of coming up with an algebraic formula, and then translating it into a BN, our approach inverses this process: first, we will construct a new BN that captures the problem by combining existing BNs in a logical way, and then, in a second step, verify the logic behind this BN by analyzing the mathematical expression for computing the LR produced by this network.

We will demonstrate the potential of this approach by applying it to the two-trace transfer problem described above. Up to now, only (Triggs and Buckleton, 2003) have proposed a formula for evaluating the corresponding LR (Eq. (10.1)) at the activity level. Note that our approach will not follow the reasoning that led to the development of Eq. (10.1). In this study, we relax the prior assumption of there being two assailants, as well as the prior assumption of it being equally probable for the suspect to have been the assailant of the victim in each of the two locations in the case of two different assailants.

This paper is organized as follows. First, we explain what BNs are and describe the BN for evaluating an activity level LR for a scenario involving the recovery of only a single trace. Second, we extend this reasoning process to the recovery of two traces in the two-trace transfer problem by constructing a new BN. Following this result, we deduce from the constructed network the algebraic expression corresponding to the model's computed LR, a formula, which we then compare with Eq. (10.1) and discuss in different situations, including an extension to n traces.

10.2 Bayesian networks

A BN (also known as a probabilistic expert system) is a directed acyclic graph composed of nodes and arrows (Jensen, 2001; Kjaerulff and Madsen, 2008). Nodes stand for random variables that can be either discrete or continuous—for the sake of simplicity, the examples in this study will all use discrete nodes. Hence, each variable will consist of a finite number of exhaustive and mutually exclusive states. The arrows represent probabilistic relationships between the variables. Each arrow connects a *parent* node to a *child* node and conditions the probability distribution of the child node upon its parent. Probability tables allow the user to quantify these probabilistic relationships. For an explanation of the different categories of relationships between variables that may be modeled by a BN, see for example Neil et al. (2000).

The key advantage offered by BNs is their capacity of splitting up a complex inference problem into its different variables. In this way, a BN decomposes the joint probability distribution of a set of random variables $X_1, \ldots, X_n$ into the product of their probabilities conditioned on their parents, which is nothing else than the Markov property:

$$Pr(X_1, ..., X_n) = \prod_{i=1}^{n} Pr(X_i | parents(X_i)).$$

Note that several different BN structures may be accepted as a description of the same scenario. There is no true model; a model is personal and reflects the constructor's view of the problem and the information available at the time of its construction (Lindley, 2000). Thus, as our understanding of the issue progresses, a constructed network may evolve to model a situation more accurately.

For constructing the BNs in this study, we used the software *Hugin Researcher 6.7*, by Hugin Expert A/S (DK-9000 Aalborg, Denmark).

10.2.1 Example of a source level Bayesian network for a single trace

To illustrate the use of BNs, Fig. 10.1 shows a BN for evaluating the source level LR for a single trace. This network consists of three variables:

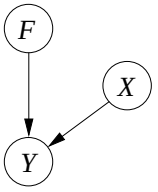


Figure 10.1: *BN for evaluating a source level LR for a single trace. Node F contains the pair of propositions F^s and $\bar{F}^s$, and nodes X and Y contain an exhaustive list of the possible analytical results of the analyses of the suspect's sample (node X) and of the trace recovered on the crime scene (node Y). Table 10.1 gives the conditional probability table for node Y.*

- F for the pair of propositions:

  F^s—*the trace on the crime scene comes from the suspect;*
  $\bar{F}^s$—*the trace on the crime scene does not come from the suspect;*
  (we use the capital letter F to distinguish the propositions in a one-trace problem from the propositions in a multiple trace problem denoted with the capital letter H),

- X for the characteristic of the suspect's sample, and

- Y for the characteristic of the trace recovered on the crime scene.

The states of nodes X and Y are an exhaustive list of the possible analytical results of the laboratory analysis (e.g., the possible blood groups or genotypes of a DNA marker). For the sake of illustration, consider that these possible results are limited to three: Γ_1, Γ_2, and Γ_{other}, where Γ_{other} groups together all the possible analytical results that are neither Γ_1, nor Γ_2.

The relationship between the three variables is the following: if F^s is true, then the characteristic of the trace must be the same as the characteristic of the suspect's sample; and if $\bar{F}^s$ is true, the characteristic of the trace is assumed to be independent of the characteristic of the suspect's sample. Note that the BNs in this paper do not include the possibility of laboratory errors (to introduce this possibility into a BN see e.g., Taroni et al. (2004, 2006a)). The characteristic of the trace, therefore, depends on which proposition is true and on the characteristic of the suspect's sample. This makes node Y a child of nodes F and X (Fig. 10.1). Table 10.1 presents the conditional probability table associated with node Y.

A match between the recovered trace and the suspect's sample—say, for example, $Y = \Gamma_1$ and $X = \Gamma_1$—produces the following LR for propositions F^s and $\bar{F}^s$:

$$LR = \frac{Pr(Y = \Gamma_1, X = \Gamma_1 | F^s)}{Pr(Y = \Gamma_1, X = \Gamma_1 | \bar{F}^s)} \ .$$

Table 10.1: *Probability table for node Y in Fig. 10.1. For simplicity, we use only three categories to describe the analytical results: Γ_1, Γ_2, and Γ_{other} (where Γ_{other} groups together all of the possible analytical results that are neither Γ_1, nor Γ_2).*

	F:	F^s			$\bar{F}^s$		
	X:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}
Y:	Γ_1	1	0	0	γ_1	γ_1	γ_1
	Γ_2	0	1	0	γ_2	γ_2	γ_2
	Γ_{other}	0	0	1	$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$	$1-\gamma_1-\gamma_2$

Applying the third law of probability for dependent events (Evett and Weir, 1998; Aitken and Taroni, 2004) produces:

$$LR = \frac{Pr(Y = \Gamma_1 | X = \Gamma_1, F^s)}{Pr(Y = \Gamma_1 | X = \Gamma_1, \bar{F}^s)} \times \frac{Pr(X = \Gamma_1 | F^s)}{Pr(X = \Gamma_1 | \bar{F}^s)} \ .$$

The characteristic of the suspect's sample ($X = \Gamma_1$) is independent of the propositions, such that $Pr(X = \Gamma_1 | F^s) = Pr(X = \Gamma_1 | \bar{F}^s)$. This reduces the second ratio to 1, and leaves us with:

$$LR = \frac{Pr(Y = \Gamma_1 | X = \Gamma_1, F^s)}{Pr(Y = \Gamma_1, | X = \Gamma_1, \bar{F}^s)} \ . \tag{10.2}$$

The BN calculates these probabilities by instantiating the states figuring to the right of the vertical bar (i.e., setting their probabilities to 1). The BN then updates the probabilities of the states in non-instantiated nodes of the model according to the laws of probability. The numerator of the LR is, therefore, given by the probability of $Y = \Gamma_1$ after instantiating states $X = \Gamma_1$ and F^s (Fig. 10.2(a)). This probability is equal to 1 (as defined in Table 10.1,

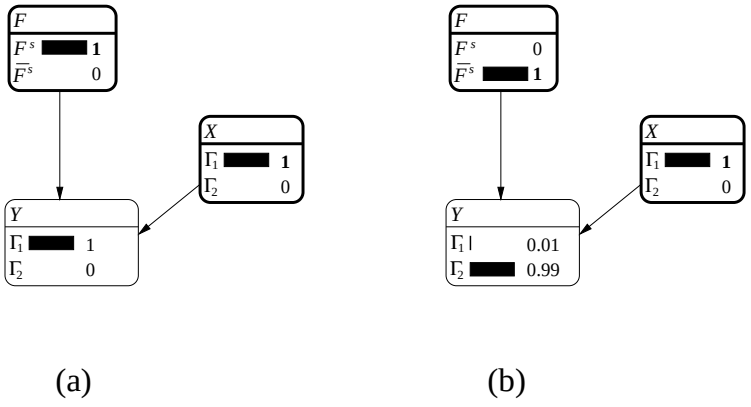


Figure 10.2: *The BN in Fig. 10.1 computes the probabilities forming the LR (Eq. (10.2)). The bold contour indicates that the node is instantiated. Here $\gamma_1 = 0.01$ and $\gamma_2 = 0.02$.* **(a)** *The numerator of the LR is the probability of $Y = \Gamma_1$ when states $X = \Gamma_1$ and F^s are instantiated;* **(b)** *the denominator the probability of $Y = \Gamma_1$ when states $X = \Gamma_1$ and $\bar{F}^s$ are instantiated.*

row 1, column 1). The denominator of the LR is given by the probability of $Y = \Gamma_1$ after instantiating states $X = \Gamma_1$ and $\bar{F}^s$ (Fig. 10.2(b)). This probability is equal to γ_1 (as defined in Table 10.1, row 1, column 4). Thus the LR is

$$LR = \frac{1}{\gamma_1} \ ,$$

which is the LR presented in forensic literature for a source level evaluation of a single trace (Evett and Weir, 1998).

10.3 Activity level Bayesian network for a single trace

Before addressing the two-trace transfer problem, this section presents the evaluation of an LR for activity level propositions in the case of a single trace. This explanation will be helpful in understanding the development of the LR for two traces. So, consider here the same scenario as described at the beginning of this study, but instead of recovering two traces on the crime scene, the investigators recover only a single trace. In this case, we consider the following pair of propositions (labeled F, because we are in a one-trace problem, with a superscript a, because they are activity level propositions):

F^a—the suspect was engaged in a struggle with the victim at the location where the trace was recovered;
$\bar{F}^a$—the suspect was not engaged in a struggle with the victim at the location where the trace was recovered.

To evaluate the LR for this pair of propositions, one must extend the BN in Fig. 10.1 to include transfer and background probabilities (Taroni et al., 2006a). For this, we must add a node B containing states:

B—presence of a background trace in the location where the trace was recovered; $\bar{B}$—absence of a background trace in the location where the trace was recovered;

and a node T containing the states:

T—there was a transfer from the assailant during the struggle at the location where the trace was recovered;
$\bar{T}$—there was no transfer from the assailant during the struggle at the location where the trace was recovered.

Both of these will determine the characteristic of the trace we observe on the crime scene. If the trace is a background trace, the probability distribution over the states of Y will be equal to the match probabilities of the characteristic in the population of background traces. If the trace is a transferred trace, Y will have the characteristic of the assailant. Nodes B and T are therefore parents of node Y (Fig. 10.3(a)).

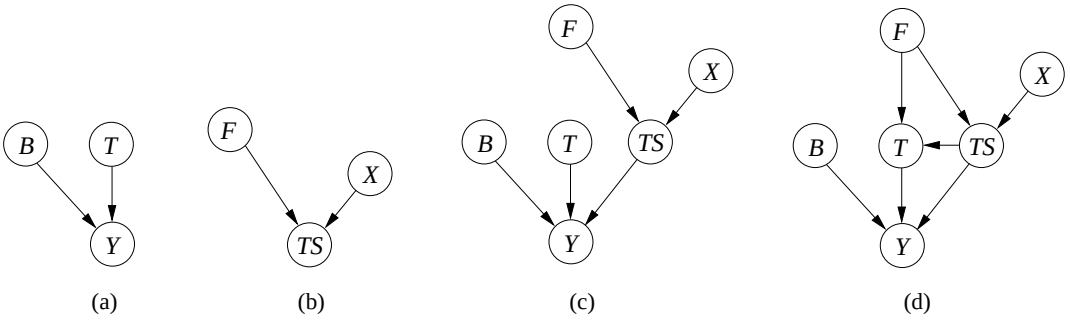


Figure 10.3: *The construction of a BN for evaluating an activity level LR for a single trace (Taroni et al., 2006a).* **(a)** *The characteristic of the trace (node Y) depends on whether the trace is a background trace (node B) or a transferred trace (node T).* **(b)** *A transferred trace's true source (node TS) will only be equal to the suspect's characteristic (node X) if the suspect was the assailant (node F).* **(c)** *If the trace was a transferred trace, its characteristic will be equal to the characteristic in node TS.* **(d)** *The transfer probabilities (in node T) may differ according to the proposition in node F and the characteristic of the transferred trace's true source in node TS. This is the complete model for evaluating the activity level LR for a single trace (Taroni et al., 2006a). The nodes figuring in this BN are defined in Table 10.2.*

The characteristic of the assailant depends on whether the suspect is this assailant (node F). If F^a is true, the characteristic of the assailant is equal to the characteristic of the suspect's sample (node X), and if $\bar{F}^a$ is true, the probability distribution over the possible

Table 10.2: *Description of the states and prior marginal probabilities of the nodes in Fig. 10.3. t and b denote the transfer and background probabilities, respectively. We differentiate between t, the transfer probability under F^a, and t', the transfer probability under $\bar{F}^a$. For node Y (the trace's characteristic), we differentiate between γ_1 and γ_2, denoting the match probabilities in the population of potential assailants, used in the case the trace was transferred during the alleged activity, and γ_1' and γ_2', denoting the match probabilities in the population of background traces, used in the case that the trace is a background trace. For simplicity, we use only three categories to describe the analytical results: Γ_1, Γ_2, and Γ_{other} (where Γ_{other} groups together all of the possible analytical results that are neither Γ_1, nor Γ_2).*

Nodes	States	Prior Marginal Probabilities	Definitions of the States
F	F^a	$Pr(F^a)$	The suspect was engaged in a struggle with the victim at the location where the trace was recovered
	$\bar{F}^a$	$1 - Pr(F^a)$	The suspect was not engaged in a struggle with the victim at the location where the trace was recovered
B	B	b	Presence of a background trace
	$\bar{B}$	$\bar{b}$	Absence of a background trace
T	T	t or t'	There was a transfer from the assailant
	$\bar{T}$	$\bar{t}$ or $\bar{t}'$	There was no transfer from the assailant
X	Γ_1	γ_1	Characteristic of the suspect's sample
	Γ_2	γ_2	
	Γ_{other}	$1 - \gamma_1 - \gamma_2$	
TS	Γ_1	γ_1	Characteristic of the trace's true source if it was transferred during the struggle with the victim
	Γ_2	γ_2	
	Γ_{other}	$1 - \gamma_1 - \gamma_2$	
Y	Γ_1	γ_1 or γ_1'	Characteristic of the trace
	Γ_2	γ_2 or γ_2'	
	Γ_{other}	$1 - \gamma_1 - \gamma_2$ or $1 - \gamma_1' - \gamma_2'$	

characteristics is given by the match probabilities in the population of possible assailants. To represent the characteristic of the assailant, we must create a new node containing the list of possible characteristics as its states, a node called TS for "true source" (Taroni et al., 2006a). This node is a child of F and X (Fig. 10.3(b)). Table 10.3 gives the conditional probability distribution over its states.

Table 10.3: *Probability table for node TS (true source) in Fig. 10.3. This node describes the characteristic of the transferred trace's true source. For simplicity, we use only three categories to describe the analytical results: Γ_1, Γ_2, and Γ_{other} (where Γ_{other} groups together all of the possible analytical results that are neither Γ_1, nor Γ_2).*

	F:	F^a			$\bar{F}^a$		
	X:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}
TS:	Γ_1	1	0	0	γ_1	γ_1	γ_1
	Γ_2	0	1	0	γ_2	γ_2	γ_2
	Γ_{other}	0	0	1	$1 - \gamma_1 - \gamma_2$	$1 - \gamma_1 - \gamma_2$	$1 - \gamma_1 - \gamma_2$

If the trace on the crime scene is a transferred trace, then its characteristic will be equal to the characteristic of the assailant given in node TS. Node TS is, therefore, a parent of node Y (Fig. 10.3(c)). Table 10.4 gives the conditional probability distribution over the states of Y given the states of its parents T, B, and TS.

The transfer probabilities defined in node T depend on the activity specified in node F. Sometimes, the activity described by proposition F^a will not be the same as the activity described by proposition $\bar{F}^a$. For example, the alternative proposition could describe a

Table 10.4: *Probability table for node Y (the trace's characteristic) in Fig. 10.3. For simplicity, we use only three categories to describe the analytical results: Γ_1, Γ_2, and Γ_{other} (where Γ_{other} groups together all of the possible analytical results that are neither Γ_1, nor Γ_2). To complete this probability table, it is necessary to add the state "not 1 trace" to represent the events $T \cap B$ and $\bar{T} \cap \bar{B}$. However, as we observe Y as being only a single trace, this state does not enter into our calculations.*

	T:	T						$\bar{T}$					
	B:	B			$\bar{B}$			B			$\bar{B}$		
	TS:	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}	Γ_1	Γ_2	Γ_{other}
Y:	Γ_1	0	0	0	1	0	0	γ_1'	γ_1'	γ_1'	0	0	0
	Γ_2	0	0	0	0	1	0	γ_2'	γ_2'	γ_2'	0	0	0
	Γ_{other}	0	0	0	0	0	1	$1-\gamma_1'-\gamma_2'$	$1-\gamma_1'-\gamma_2'$	$1-\gamma_1'-\gamma_2'$	0	0	0
	not 1 trace	1	1	1	0	0	0	0	0	0	1	1	1

legitimate activity between the suspect and the victim (such as *the suspect was trying to rescue the victim*). In this case, the transfer probabilities will be different under each of the two propositions, making node T a child of node F. We therefore use t' (and its complement $\bar{t'}$) to denote the transfer probability given proposition $\bar{F}^a$ to distinguish this probability from t (and its complement $\bar{t}$) denoting the transfer probability given proposition F^a.

The occurrence of a transfer may also depend on the attributes of the transferred material. The extent of such an influence depends on the type of trace evidence considered. For example, for fiber evidence, a wool fiber may be transferred more easily than a silk fiber. If the analytical results in the BN are the types of fiber, then node TS must also be a parent of node T to specify different transfer probabilities for different types of fibers. Fig. 10.3(d) shows the complete BN for the transfer of a single trace, modeling all of the possible dependence relationships (Taroni et al., 2006a).

If $X = \Gamma_1$ and $Y = \Gamma_1$, the LR computed by this BN is equal to

$$LR = \frac{Pr(Y = \Gamma_1 | X = \Gamma_1, F^a)}{Pr(Y = \Gamma_1 | X = \Gamma_1, \bar{F}^a)} \tag{10.3}$$

$$= \frac{\bar{b}t + b\gamma_1'\bar{t}}{\bar{b}t'\gamma_1 + b\gamma_1'\bar{t'}} \ . \tag{10.4}$$

This LR corresponds to the LR developed in the literature (Evett, 1984).

This BN computes the two probabilities of Y that form the LR in the same way as at the source level: the numerator is the probability of $Y = \Gamma_1$ when the states $X = \Gamma_1$ and F^a are instantiated, and the denominator the probability of $Y = \Gamma_1$ when the states $X = \Gamma_1$ and $\bar{F}^a$ are instantiated (Fig. 10.4). However, unlike the source level evaluation, the resulting

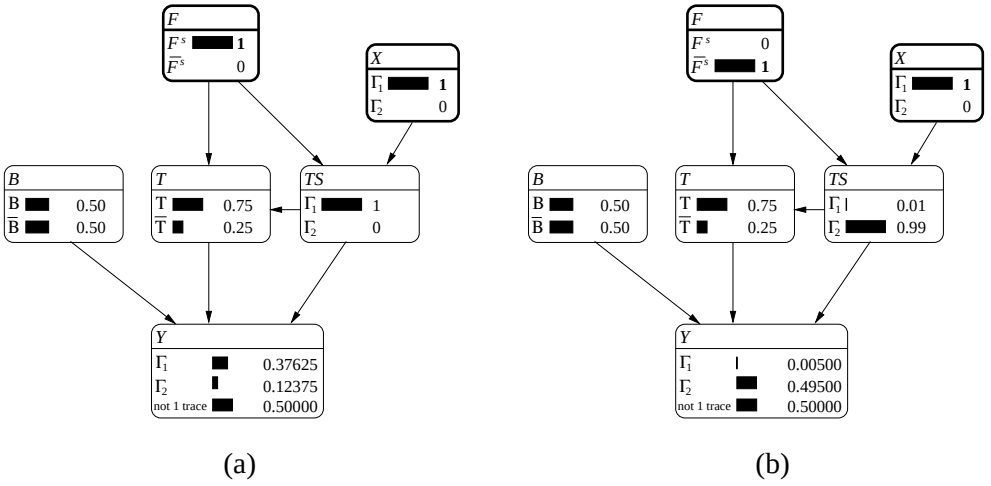


Figure 10.4: *The BN in Fig. 10.3(d) computes the probabilities forming the LR (Eq. (10.4)). Here,* $b = 0.5$, $t = t' = 0.75$, $\gamma_1 = \gamma_1' = 0.01$, *and* $\gamma_2 = \gamma_2' = 0.02$. *The bold contour indicates that the node is instantiated.* **(a)** *The numerator of the LR is the probability of* $Y = \Gamma_1$ *when states* $X = \Gamma_1$ *and* F^a *are instantiated, in this case 0.37625;* **(b)** *the denominator is the probability of* $Y = \Gamma_1$ *when states* $X = \Gamma_1$ *and* $\bar{F}^a$ *are instantiated, in this case 0.005.*

probabilities for $Y = \Gamma_1$ no longer figure in the probability table for Y. This is because in this BN, there are additional nodes separating node Y from nodes X and F. The calculation of the LR takes these intermediate nodes into account by extending the conversation (e.g., Lindley, 1985; Dawid and Evett, 1997) of the probability of Y to the parent variables of this node, that is, to B, T and TS, according to the relationships described in the probability table of node Y (Table 10.4). The first row of this table, corresponding to $Y = \Gamma_1$, tells us that the trace can only have characteristic Γ_1 when the trace was transferred from a source having characteristic Γ_1 in the absence of a background trace (column 4), or when the trace is a background trace and there was no transfer from the assailant (columns 7, 8, and 9).

In all other cases, Table 10.4 defines a probability of 0 for observing $Y = \Gamma_1$, making this analytical result impossible for the combination of the states in that column. The table defines a probability of 1 for $Y = \Gamma_1$ if the trace was transferred, and a probability of γ_1' if the trace is a background trace, such that the LR (Eq. (10.3)) is equal to

$$LR = \frac{1 \times \overbrace{Pr(\bar{B} \cap T \cap TS = \Gamma_1 | X = \Gamma_1, F^a)}^{\text{(a)}} + \gamma_1' \times \overbrace{Pr(B \cap \bar{T} | X = \Gamma_1, F^a)}^{\text{(b)}}}{1 \times \underbrace{Pr(\bar{B} \cap T \cap TS = \Gamma_1 | X = \Gamma_1, \bar{F}^a)}_{\text{(c)}} + \gamma_1' \times \underbrace{Pr(B \cap \bar{T} | X = \Gamma_1, \bar{F}^a)}_{\text{(d)}}} \ , \tag{10.5}$$

where the mathematical symbol $\cap$ means the intersection where both the state on its left and the state on its right are true. The probabilities labelled (a), (b), (c), and (d) are discussed below:

(a) $Pr(\bar{B} \cap T \cap TS = \Gamma_1 | X = \Gamma_1, F^a)$ is the probability that the trace was transferred during the struggle from an assailant having characteristic Γ_1, given that the suspect has characteristic Γ_1 and that the suspect was this assailant who struggled with the victim. This probability is equal to

$$\bar{b} \times t \times Pr(TS = \Gamma_1 | X = \Gamma_1, F^a) \ .$$

According to Table 10.3, $Pr(TS = \Gamma_1 | X = \Gamma_1, F^a) = 1$ (row 1, column 1), so that the above expression reduces to

$$\bar{b} \times t \ .$$

(b) $Pr(B \cap \bar{T} | X = \Gamma_1, F^a)$ is the probability that the trace is a background trace and that there was no transfer from the assailant, given that the suspect, having characteristic Γ_1, was the assailant in the struggle with the victim. As the trace was not transferred from the assailant, this probability is independent of the assailant's characteristic and is equal to

$$b \times \bar{t} \ .$$

(c) $Pr(\bar{B} \cap T \cap TS = \Gamma_1 | X = \Gamma_1, \bar{F}^a)$ is the probability that the trace was transferred during the struggle from an assailant having characteristic Γ_1, given that the suspect has characteristic Γ_1, but that the suspect was not the assailant who struggled with the victim on the crime scene. This probability is equal to

$$\bar{b} \times t' \times Pr(TS = \Gamma_1 | X = \Gamma_1, \bar{F}^a) \ .$$

According to Table 10.3, $Pr(TS = \Gamma_1 | X = \Gamma_1, \bar{F}^a) = \gamma_1$ (row 1, column 4), so that the above expression is equal to

$$\bar{b} \times t' \times \gamma_1 \ .$$

(d) $Pr(B \cap \bar{T} | X = \Gamma_1, \bar{F}^a)$ is the probability that the trace is a background trace and that there was no transfer from the assailant, given that the suspect, with characteristic Γ_1, was not the assailant in the struggle with the victim. As in (b), this probability is independent of the assailant's characteristic and is equal to

$$b \times \bar{t'} \ .$$

Introducing the expressions for (a), (b), (c) and (d) into Eq. (10.5) produces

$$LR = \frac{1 \times \bar{b} \times t + \gamma_1' \times b \times \bar{t}}{1 \times \bar{b} \times t' \times \gamma_1 + \gamma_1' \times b \times \bar{t'}} \ ,$$

which is Eq. (10.4). This calculation validates the structure of the BN in Fig. 10.3(d) for evaluating an activity level LR for a single trace. In the next section, we extend this BN to two traces.

10.4 Constructing a Bayesian network for two traces

This section describes the steps for extending a BN for a single trace to two traces. We first illustrate this concept for the BN in Fig. 10.1, at the source level, and then apply the same reasoning to the BN in Fig. 10.3(d), at the activity level, to address the two-trace transfer problem described at the beginning of the paper.

10.4.1 Constructing a source level Bayesian network for two traces

In a two-trace problem, the evidence consists of the characteristic of the suspect's sample and the characteristics of two traces recovered on the crime scene. We will call the first trace recovered on the crime scene "trace 1," and the second trace recovered on the scene "trace 2."

To represent the characteristics of each of the traces, we duplicate node Y in Fig. 10.1, creating a node for trace 1 called Y_1, and a node for trace 2 called Y_2. As in Fig. 10.1, the characteristic of each of these traces will be identical to the characteristic of the suspect's sample if that trace comes from the suspect. Therefore, Y_1 and Y_2 are each a child of nodes X and F. To distinguish between the two traces, we duplicate node F to create node F_1 for trace 1, and node F_2 for trace 2 (Fig. 10.5(a)), so that F_1 contains the propositions:

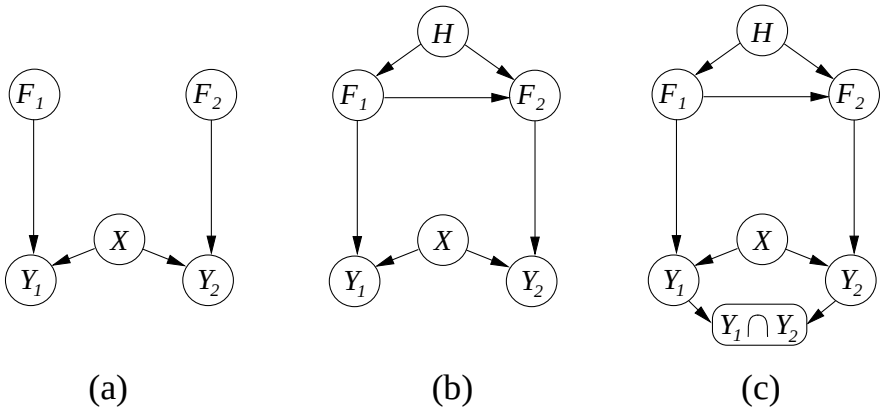


Figure 10.5: *The construction of a BN for evaluating a source level LR for two traces.* **(a)** *Nodes F and Y in Fig. 10.1 are duplicated such that nodes F_1 and Y_1 refer to trace 1, and nodes F_2 and Y_2 to trace 2.* **(b)** *The pair of propositions H^s and $\bar{H}^s$ in node H are added as a parent to nodes F_1 and F_2 (see Tables 10.5 and 10.6 for the conditional probability tables of F_1 and F_2).* **(c)** *The BN is completed by adding a node $Y_1 \cap Y_2$ to compute the numerator and denominator of the LR. This is a modified version of the BN presented in (Taroni et al., 2006a).*

F_1^s—*trace 1 comes from the suspect;*
$\bar{F}_1^s$—*trace 1 does not come from the suspect;*

and F_2 the propositions:

F_2^s—*trace 2 comes from the suspect;*
$\bar{F}_2^s$—*trace 2 does not come from the suspect;*

The probability tables of nodes Y_1 and Y_2 are identical to Table 10.1 for node Y in a one-trace scenario.

Now, the propositions of interest for evaluating an LR for multiple traces are no longer F^s and $\bar{F}^s$, but H^s and $\bar{H}^s$:

H^s—*one of the traces on the crime scene comes from the suspect;*
$\bar{H}^s$—*neither of the traces on the crime scene comes from the suspect.*

Assuming that it is not possible for both of the traces on the crime scene to come from the suspect, and that it is equally probable for either of the traces to come from the suspect when H^s is true (for a source level BN where we relax these assumptions, see Gittelson et al. (2013a)), then either F_1^s or F_2^s must be true when proposition H^s is true, but F_1^s and F_2^s will never both be true at the same time. We therefore add node H, containing states H^s and $\bar{H}^s$, as a parent of nodes F_1 and F_2. In addition, we must add a link between nodes F_1 and F_2 to assure that F_2^s is true (i.e., trace 2 comes from the suspect) when H^s (i.e., one of the traces comes from the suspect) and $\bar{F}_1^s$ (i.e., trace 1 does not come from the suspect) are both true (Fig. 10.5(b)). If proposition $\bar{H}^s$ is true, then it follows that both $\bar{F}_1^s$ and $\bar{F}_2^s$ are true. Tables 10.5 and 10.6 define the probability distributions over F_1^s, $\bar{F}_1^s$, F_2^s and $\bar{F}_2^s$.

Table 10.5: *Probability table for node F_1 in Fig. 10.5. This node indicates whether or not trace 1 comes from the suspect. Given that one of the traces on the crime scene comes from the suspect, this probability table considers it equally probable for this trace to be trace 1 or trace 2.*

	H:	H^s	$\bar{H}^s$
F_1:	F_1^s	0.5	0
	$\bar{F}_1^s$	0.5	1

Table 10.6: *Probability table for node F_2 in Fig. 10.5. This node indicates whether or not trace 2 comes from the suspect. This probability table considers it impossible for both of the traces to come from the suspect, and considers that either trace 1 or trace 2 must come from the suspect if H^s is true.*

	H:	H^s		$\bar{H}^s$	
	F_1:	F_1^s	$\bar{F}_1^s$	F_1^s	$\bar{F}_1^s$
F_2:	F_2^s	0	1	0	0
	$\bar{F}_2^s$	1	0	1	1

For this BN to compute the numerator and denominator of the LR, we add a node that combines the characteristics of both traces, node $Y_1 \cap Y_2$, as a child of Y_1 and Y_2 (Fig. 10.5(c)). The resulting BN is a slightly modified version of the BN presented in Taroni et al. (2006a). (In Taroni et al. (2006a), propositions F_1^s, $\bar{F}_1^s$, F_2^s, and $\bar{F}_2^s$ are all combined in a single node F.) The two versions are logically equivalent and compute an LR of

$$\begin{aligned} LR &= \frac{Pr(Y_1 = \Gamma_1, Y_2 = \Gamma_2 | X = \Gamma_1, H^s)}{Pr(Y_1 = \Gamma_1, Y_2 = \Gamma_2 | X = \Gamma_1, \bar{H}^s)} \\ &= \frac{1}{2\gamma_1} \end{aligned}$$

for $Y_1 = \Gamma_1$, $Y_2 = \Gamma_2$, and $X = \Gamma_1$. This model is, therefore, in perfect agreement with scientific literature (Evett, 1987). For further explanations or examples of a BN treating a two-trace problem at the source level, see Taroni et al. (2006a) and Gittelson et al. (2013a).

10.4.2 Constructing an activity level Bayesian network for two traces

To address the two-trace transfer problem at the activity level, we proceed in the same way as at the source level to extend a BN of a single trace to two traces. We begin by using node X as the center of the new model, and duplicate the rest of the nodes on either side of X, such that we have a set of nodes referring to trace 1 (labeled with a subscript 1) on the left of X, and a set of nodes referring to trace 2 (labeled with a subscript 2) on the right of X (Fig. 10.6(a)). The probability tables for nodes TS_1, TS_2, Y_1, and Y_2 are identical to Tables 10.3 and 10.4, for nodes TS and Y, respectively.

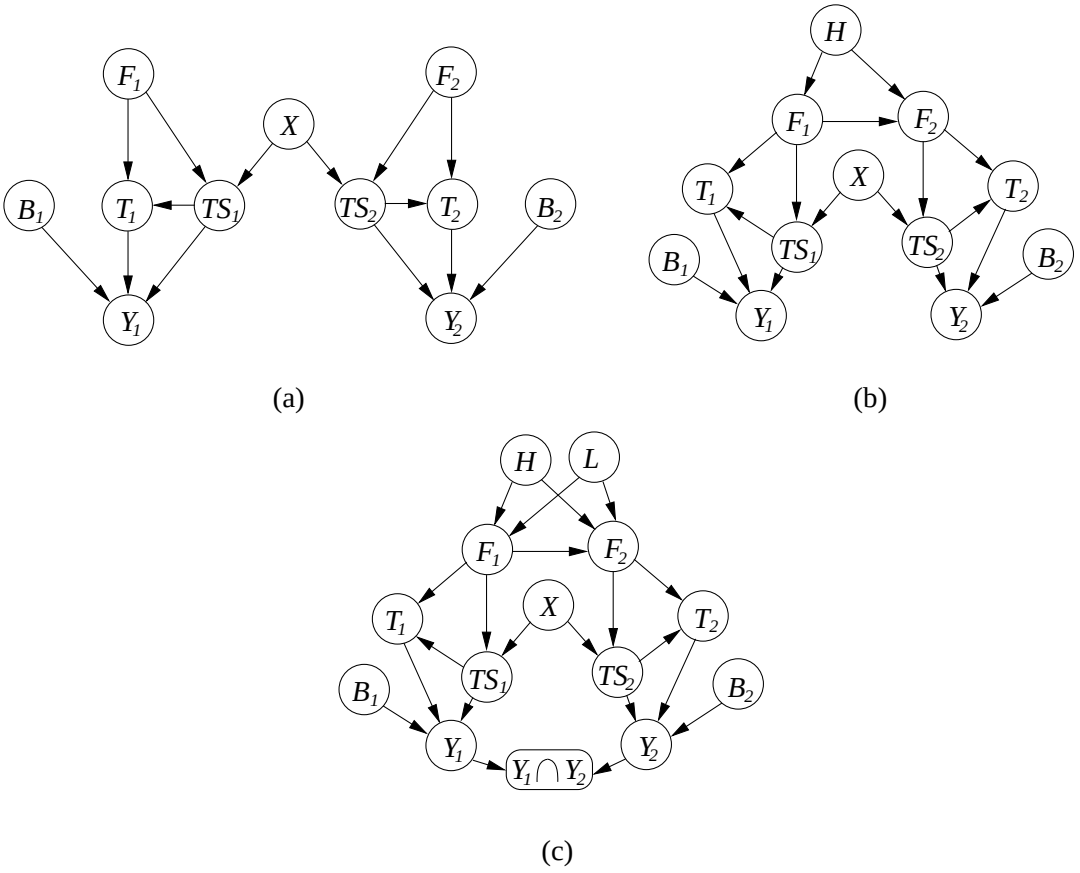


Figure 10.6: *The construction of a BN for evaluating an activity level LR for two traces. We begin with the BN in Fig. 10.3(d).* **(a)** *We duplicate all of the nodes in this model except for X, and use subscripts 1 and 2 to differentiate the nodes referring to trace 1 from the nodes referring to trace 2.* **(b)** *We add a node H and an arrow from F_1 to F_2 just as in Fig. 10.5(b).* **(c)** *We add an additional node L to model the uncertainty on whether the assailant in location 1 was the same person as the assailant in location 2, and a node $Y_1 \cap Y_2$ to compute the numerator and denominator of the LR.*

Again, we introduce a node H as a parent of nodes F_1 and F_2. H contains the propositions of interest for the two-trace transfer problem:

H^a—*the suspect was engaged in a struggle with the victim;*
$\bar{H}^a$—*the suspect was not engaged in a struggle with the victim;*

F_1 the propositions:

F_1^a—*the suspect was engaged in a struggle with the victim in location 1;*
$\bar{F}_1^a$—*the suspect was not engaged in a struggle with the victim in location 1;*

and F_2 the propositions:

F_2^a—*the suspect was engaged in a struggle with the victim in location 2;*
$\bar{F}_2^a$—*the suspect was not engaged in a struggle with the victim in location 2.*

If $\bar{H}^a$ is true, then it follows that both $\bar{F}_1^a$ and $\bar{F}_2^a$ are true. If H^a is true, we assume that either F_1^a, F_2^a, or both F_1^a and F_2^a must be true. Unlike the BN we presented above for a source level evaluation, we will take into account here the possibility that it may have been the same assailant in both locations. For this, we must link F_1 to F_2, as before (Fig. 10.6(b)), and define a new node L containing the states

L—the assailant in location 1 is the same person as the assailant in location 2;
$\bar{L}$—the assailant in location 1 is not the same person as the assailant in location 2;

as a parent to F_1 and F_2 (Gittelson et al., 2013a). We use the probability λ to denote the prior probability that there were two different assailants on the crime scene (Dawid, 2004), that is, $Pr(\bar{L}) = \lambda$. To relax the assumption that it is equally probable for the suspect to have struggled with the victim in either of the two locations in a case where the suspect was one of two assailants, we introduce another probability, which we call τ. This probability represents the prior probability that the suspect was engaged in the struggle in location 1 in the case that the suspect is an assailant and that there were two different assailants in locations 1 and 2 on the crime scene, i.e., $Pr(F_1^a|\bar{L}, H^a) = \tau$. Tables 10.7 and 10.8 give the probability tables for F_1 and F_2.

Table 10.7: *Probability table for node F_1 in Fig. 10.6. This node indicates whether or not the suspect was engaged in a struggle with the victim in location 1. We use τ to denote the probability that the suspect was engaged in the struggle in location 1 given that he is one of two assailants (i.e., $Pr(F_1^a|\bar{L}, H^a)$).*

	H:	H^a		$\bar{H}^a$	
	L:	L	$\bar{L}$	L	$\bar{L}$
F_1:	F_1^a	1	τ	0	0
	$\bar{F}_1^a$	0	$1-\tau$	1	1

Table 10.8: *Probability table for node F_2 in Fig. 10.6. This node indicates whether or not the suspect was engaged in a struggle with the victim in location 2. This is only possible when either the suspect was the assailant in both locations (column 1), or when the suspect was one of two assailants and he was not engaged in a struggle with the victim in location 1 (column 4).*

	H:	H^a				$\bar{H}^a$			
	L:	L		$\bar{L}$		L		$\bar{L}$	
	F_1:	F_1^a	$\bar{F}_1^a$	F_1^a	$\bar{F}_1^a$	F_1^a	$\bar{F}_1^a$	F_1^a	$\bar{F}_1^a$
F_2:	F_2^a	1	n/a	0	1	0	0	0	0
	$\bar{F}_2^a$	0	n/a	1	0	1	1	1	1

To compute the LR,

$$LR = \frac{Pr(Y_1 = \Gamma_1, Y_2 = \Gamma_2|X = \Gamma_1, H^a)}{Pr(Y_1 = \Gamma_1, Y_2 = \Gamma_2|X = \Gamma_1, \bar{H}^a)}\ , \tag{10.6}$$

we add a node $Y_1 \cap Y_2$ as a child of nodes Y_1 and Y_2 (Fig. 10.6(c)). This completes the BN, with all of the nodes and states of the final model given in Table 10.9.

Note that if the same background noise is present in both locations, nodes B_1 and B_2 may be merged into a single node B, parent to both Y_1 and Y_2. This creates an additional probabilistic link between Y_1 and Y_2, which influences the evaluation of the LR. However, this special case is not treated in this paper.

In the next section, we present the algebraic expression that corresponds to the computed LR. We derive this expression to compare the LR provided by this model with the existing formulae in forensic literature.

10.5 Algebraic expression for the LR computed by the Bayesian network in Fig. 10.6(c)

The robust mathematical framework of BNs allows their user to deduce the algebraic formulae of interest. In the forensic context of the two-trace transfer problem, we are interested in

Table 10.9: *Description of the states and prior marginal probabilities of the nodes in Fig. 10.6, with* $i = 1, 2$. *We use* δ *to denote the prior probability that the suspect was engaged in a struggle with the victim (i.e.,* $Pr(H^a) = \delta$*),* λ *to denote the prior probability that the victim struggled with two different assailants in location 1 and location 2 (i.e.,* $Pr(\bar{L}) = \lambda$*), and* τ *to denote the prior probability of the suspect being the assailant who struggled with the victim in location 1 in the case where the suspect was an assailant, and the assailant in location 1 was not the same person as the assailant in location 2 (i.e.,* $Pr(F_1^a|\bar{L}, H^a) = \tau$*). For node* T_i*, we differentiate between* t_i*, the transfer probability under* F_i^a*, and* t_i'*, the transfer probability under* $\bar{F}_i^a$*. For node* Y *(the trace's characteristic), we differentiate between* γ_1 *and* γ_2*, denoting the match probabilities in the population of potential assailants, used in the case the trace was transferred during the alleged activity, and* γ_1' *and* γ_2'*, denoting the match probabilities in the population of background traces, used in the case that the trace is a background trace. For simplicity, we use only three categories to describe the analytical results:* Γ_1*,* Γ_2*, and* Γ_{other} *(where* Γ_{other} *groups together all of the possible analytical results that are neither* Γ_1*, nor* Γ_2*).*

Nodes	States	Prior Marginal Probabilities	Definitions of the States
H	H^a	δ	The suspect was engaged in a struggle with the victim
	$\bar{H}^a$	$1-\delta$	The suspect was not engaged in a struggle with the victim
L	L	$1-\lambda$	The assailant in location 1 is the same person as the assailant in location 2
	$\bar{L}$	λ	The assailant in location 1 is not the same person as the assailant in location 2
F_1	F_1^a	$\delta(1-\lambda+\lambda\tau)$	The suspect was engaged in a struggle with the victim in location 1
	$\bar{F}_1^a$	$\delta\lambda(1-\tau)+1-\delta$	The suspect was not engaged in a struggle with the victim in location 1
F_2	F_2^a	$\delta(1-\lambda+\lambda(1-\tau))$	The suspect was engaged in a struggle with the victim in location 2
	$\bar{F}_2^a$	$\delta\lambda\tau+1-\delta$	The suspect was not engaged in a struggle with the victim in location 2
B_i	B_i	b_i	Presence of a background trace in location i
	$\bar{B}_i$	$\bar{b}_i$	Absence of a background trace in location i
T_i	T_i	t_i or t_i'	There was a transfer from the assailant in location i
	$\bar{T}_i$	$\bar{t}_i$ or $\bar{t}_i'$	There was no transfer from the assailant in location i
X	Γ_1	γ_1	Characteristic of the suspect's sample
	Γ_2	γ_2	
	Γ_{other}	$1-\gamma_1-\gamma_2$	
TS_i	Γ_1	γ_1	Characteristic of trace i's true source if it was transferred during the struggle with the victim
	Γ_2	γ_2	
	Γ_{other}	$1-\gamma_1-\gamma_2$	
Y_i	Γ_1	γ_1 or γ_1'	Characteristic of trace i
	Γ_2	γ_2 or γ_2'	
	Γ_{other}	$1-\gamma_1-\gamma_2$ or $1-\gamma_1'-\gamma_2'$	
$Y_1 \cap Y_2$		Combine Y_1 and Y_2	Characteristics of trace 1 and trace 2

the LR for propositions H^a and $\bar{H}^a$ (Eq. (10.6)). In this section, we introduce the transfer, background and match probabilities into this expression, based on the probabilistic relationships modeled in the BN in Fig. 10.6(c). For this development, we assume that trace 1 has characteristic Γ_1 ($Y_1 = \Gamma_1$) and trace 2 characteristic Γ_2 ($Y_2 = \Gamma_2$). As for the characteristic of the suspect's sample, we will label it X for the time being and specify its characteristic later on.

We begin by applying the third law of probability for dependent events to Eq. (10.6):

$$LR = \underbrace{\frac{Pr(Y_1 = \Gamma_1|X, H^a)}{Pr(Y_1 = \Gamma_1|X, \bar{H}^a)}}_{\text{(i)}} \times \underbrace{\frac{Pr(Y_2 = \Gamma_2|Y_1 = \Gamma_1, X, H^a)}{Pr(Y_2 = \Gamma_2|Y_1 = \Gamma_1, X, \bar{H}^a)}}_{\text{(ii)}} \tag{10.7}$$

This separates the LR into the product of two ratios:

(i) the LR for observing the characteristic of trace 1 for the two competing propositions given the characteristic of the suspect's sample;

(ii) the LR for observing the characteristic of trace 2 for the same propositions, given the characteristic of the suspect's sample *and given that the characteristic of trace 1 has already been observed.*

As the probability tables for nodes Y_1 and Y_2 are identical to the probability table for node Y in the one-trace evaluation, the developments of these two ratios over nodes B_1, T_1 and TS_1, and B_2, T_2 and TS_2, respectively, are identical to Eq. (10.5):

$$LR = \frac{1 \times \overbrace{Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|X, H^a)}^{\text{(a)}} + \gamma_1' \times \overbrace{Pr(B_1 \cap \bar{T}_1|X, H^a)}^{\text{(b)}}}{\underbrace{1 \times Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|X, \bar{H}^a)}_{\text{(c)}} + \gamma_1' \times \underbrace{Pr(B_1 \cap \bar{T}_1|X, \bar{H}^a)}_{\text{(d)}}}$$
$$\times \frac{1 \times \overbrace{Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X, H^a)}^{\text{(e)}} + \gamma_2' \times \overbrace{Pr(B_2 \cap \bar{T}_2|Y_1 = \Gamma_1, X, H^a)}^{\text{(f)}}}{\underbrace{1 \times Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X, \bar{H}^a)}_{\text{(g)}} + \gamma_2' \times \underbrace{Pr(B_2 \cap \bar{T}_2|Y_1 = \Gamma_1, X, \bar{H}^a)}_{\text{(h)}}} . \tag{10.8}$$

Let us first examine the denominators of these two ratios.

10.5.1 Denominators of the ratios in Eq. (10.8)

Under $\bar{H}^a$, neither of the traces was transferred by the suspect, so we assume the suspect's characteristic has no influence on the probabilities of $Y_1 = \Gamma_1$ and $Y_2 = \Gamma_2$, and remove the conditioning on X from these probabilities.

Probability (c) (i.e., $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|\bar{H}^a)$) is the probability that trace 1 was transferred during the struggle from an assailant having characteristic Γ_1, given that the suspect was not an assailant who struggled with the victim on the crime scene. As in the one-trace transfer problem, this probability is equal to

$$\bar{b_1} \times t_1' \times Pr(TS_1 = \Gamma_1|\bar{H}^a) \ .$$

The probability table for node TS_1 is identical to the probability table for node TS shown in Table 10.3, such that $Pr(TS_1 = \Gamma_1|\bar{H}^a) = \gamma_1$, and the above expression is equal to

$$\bar{b_1} \times t_1' \times \gamma_1 \ .$$

Probability (d) (i.e., $Pr(B_1 \cap \bar{T}_1|\bar{H}^a)$) is the probability that trace 1 is a background trace and that there was no transfer from the assailant in location 1, given that the suspect was not an assailant in the struggle with the victim. As in the one-trace transfer problem, this probability is independent of the assailant's characteristic and is equal to

$$b_1 \times \bar{t_1'} \ .$$

Probability (g) (i.e., $Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, \bar{H}^a)$) is the probability that trace 2 was transferred during the struggle from an assailant having characteristic Γ_2, given that trace 1 has characteristic Γ_1 and that the suspect was not an assailant who struggled with the victim on the crime scene. This probability is equal to

$$\bar{b_2} \times t_2' \times Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, \bar{H}^a) \ .$$

The probability table for node TS_2 is identical to the probability table for node TS shown in Table 10.3, such that $Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, \bar{H}^a) = \gamma_2$, and the above expression is equal to

$$\bar{b_2} \times t_2' \times \gamma_2 \ .$$

Probability (h) (i.e., $Pr(B_2 \cap \bar{T}_2|Y_1 = \Gamma_1, \bar{H}^a)$) is the probability that trace 2 is a background trace and that there was no transfer from the assailant in location 2, given that trace 1 has characteristic Γ_1 and that the suspect was not an assailant in the struggle with the victim. As in the development of probability (d), this probability is independent of the assailant's characteristic and is equal to

$$b_2 \times \bar{t_2'} \ .$$

The denominator of ratio (i) is therefore

$$\bar{b_1} t_1' \gamma_1 + \gamma_1' b_1 \bar{t_1'} \ ,$$

and the denominator for ratio (ii),

$$\bar{b_2} t_2' \gamma_2 + \gamma_2' b_2 \bar{t_2'} \ .$$

Each of these expressions is identical to the denominator of the LR published for the one-trace transfer problem (Eq. (10.4)). This is reasonable, because the observation of two different traces makes it impossible for the two traces to have come from the same person, such that the observations of the two traces can be considered independent of each other. Each observation is, therefore, comparable with the observation of a single trace.

Next, let us examine the numerators of the two ratios in Eq. (10.8).

10.5.2 Numerators of the ratios in Eq. (10.8)

Probabilities (a) and (e) (i.e., $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|X, H^a)$ and $Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X, H^a)$) are the probabilities that each of the traces was transferred during the struggle, and probabilities (b) and (f) (i.e., $Pr(B_1 \cap \bar{T}_1|X, H^a)$ and $Pr(B_2 \cap \bar{T}_2|Y_1 = \Gamma_1, X, H^a)$) the probabilities that each of the traces is a background trace and that there was no transfer from the struggle in each of the locations. The developments of the latter are independent of an assailant's characteristic. As in the one-trace transfer problem, probability (b) is equal to

$$b_1 \times \bar{t_1} \ ,$$

and probability (f) to

$$b_2 \times \bar{t_2} \ .$$

Concerning the probabilities that each of the traces was transferred during the struggle under H^a, it is possible that a trace may have been transferred by the suspect. We must, therefore, take into account the probability that the true source of a transferred trace may effectively be the suspect. As a result, we must consider the characteristic of the suspect's sample to develop probabilities (a) and (e).

Theoretically, there are two possibilities for the evidence in a two-trace problem:

(1) the suspect's sample has characteristic Γ_1 (i.e., $X = \Gamma_1$) and matches trace 1; or

(2) the suspect's sample has characteristic Γ_2 (i.e., $X = \Gamma_2$) and matches trace 2.

In the first case, the suspect matches the first trace observed on the scene; in the second case, the suspect matches the second trace observed on the scene. Of course, the order of the observations will not affect the numerical value obtained for the LR. However, the algebraic derivation of the formulae includes conditional probabilities which will differ in these two scenarios. Therefore, we will develop probabilities (a) and (e) twice: first, we will assume that the suspect's sample matches the first trace, and second, that the suspect's sample matches the second trace.

In the case that the suspect's sample matches trace 1 (i.e., $X = \Gamma_1$)

Probability (a), that is, $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1 | X = \Gamma_1, H^a)$, is the probability of the event that trace 1 was transferred during the struggle from an assailant having characteristic Γ_1, given that the suspect has characteristic Γ_1 and that the suspect was an assailant who struggled with the victim on the crime scene. As in the one-trace transfer problem, this probability is equal to

$$\bar{b_1} \times t_1 \times Pr(TS_1 = \Gamma_1 | X = \Gamma_1, H^a) \ . \tag{10.9}$$

To find $Pr(TS_1 = \Gamma_1 | X = \Gamma_1, H^a)$, we must continue to work our way up in the structure of the BN (Fig. 10.6(c)). Between nodes TS_1 and H is node F_1. We must, therefore, extend the conversation to propositions F_1^a and $\bar{F}_1^a$:

$$\begin{aligned} Pr(TS_1 = \Gamma_1 | X = \Gamma_1, H^a) &= 1 \times \underbrace{Pr(F_1^a | H^a)}_{1-\lambda+\lambda\tau} + \gamma_1 \times \underbrace{Pr(\bar{F}_1^a | H^a)}_{\lambda(1-\tau)} \\ &= 1 - \lambda + \lambda\tau + \gamma_1\lambda(1-\tau) \ . \end{aligned}$$

This development results from two possible explanations for the event $TS_1 = \Gamma_1$ (i.e., the event that a transferred trace's true source in location 1 is Γ_1):

- either trace 1 was transferred from the suspect, in which case the probability of $TS_1 = \Gamma_1$ is equal to 1;
- or trace 1 was transferred by the other assailant, who has characteristic Γ_1 with a probability of γ_1.

The weighted sum of these values with the probabilities of each of them occurring given proposition H^a produces the above expression. The first case is possible either when the suspect was the only assailant (probability of $1 - \lambda$), or when the suspect was one of two assailants (probability of λ) and of these two, was the one in location 1 (probability of τ). The second case is only possible when the suspect was one of two assailants (probability of λ), and this time was the one in location 2 (probability of $1 - \tau$).

Inserting these results into Eq. (10.9) gives us the following expression for $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1 | X = \Gamma_1, H^a)$:

$$\bar{b_1} \times t_1 \times \left[1 - \lambda + \lambda\tau + \gamma_1\lambda(1-\tau)\right] \ .$$

Introducing the expressions for (a), (b), (c) and (d) into Eq. (10.8) produces

$$\frac{Pr(Y_1 = \Gamma_1|X = \Gamma_1, H^a)}{Pr(Y_1 = \Gamma_1|X = \Gamma_1, \bar{H}^a)} = \frac{\bar{b_1}t_1\left[1 - \lambda + \lambda\tau + \gamma_1\lambda(1-\tau)\right] + \gamma_1' b_1 \bar{t_1}}{\bar{b_1}t_1'\gamma_1 + \gamma_1' b_1 \bar{t_1'}} \tag{10.10}$$

for ratio (i) in Eq. (10.7).

The expression for (e) is more complex. $Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a)$ is the probability of the event that trace 2 was transferred during the struggle from an assailant having characteristic Γ_2, given that trace 1 and the suspect have characteristic Γ_1 and that the suspect was an assailant who struggled with the victim on the crime scene. This probability is equal to

$$\bar{b_2} \times t_2 \times Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a) \ , \tag{10.11}$$

and the extension of the conversation to propositions F_2^a and $\bar{F}_2^a$ to find $Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a)$ produces

$$Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a) = \gamma_2 \times Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_1, H^a) \ . \tag{10.12}$$

Given that the suspect has characteristic Γ_1, the true source of a transferred trace in location 2 can only have characteristic Γ_2 if the assailant in this location was not the suspect (proposition $\bar{F}_2^a$). In this case, the probability that the true source has characteristic Γ_2 is γ_2. To compute $Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_1, H^a)$, the BN applies Bayes' theorem:

$$Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_1, H^a) = \frac{Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_1, H^a) \times Pr(\bar{F}_2^a|H^a)}{Pr(Y_1 = \Gamma_1|X = \Gamma_1, H^a)} \ . \tag{10.13}$$

This ratio is made up of the following three probabilities:

- The development of probability $Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_1, H^a)$ is identical to that of the numerator of the traditional one-trace transfer scenario (Eq. (10.5)):

$$\begin{aligned} &Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_1, H^a) \\ &\quad = 1 \times Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_1, H^a) + \gamma_1' \times Pr(B_1 \cap \bar{T}_1|H^a) \\ &\quad = \bar{b_1}t_1 + \gamma_1' b_1 \bar{t_1} \ . \end{aligned}$$

- By definition,

$$Pr(\bar{F}_2^a|H^a) = \lambda\tau \ ,$$

  that is, the probability that the assailant in location 2 was not the suspect given that the suspect was an assailant is equal to the probability that there were two different assailants in each of the locations (probability λ) and that the suspect was the assailant in location 1 (probability τ).

- And $Pr(Y_1 = \Gamma_1|X = \Gamma_1, H^a)$ is the numerator of the first ratio in our LR (Eq. (10.10)):

$$Pr(Y_1 = \Gamma_1|X = \Gamma_1, H^a) = \bar{b_1}t_1\left[1 - \lambda + \lambda\tau + \gamma_1\lambda(1-\tau)\right] + \gamma_1' b_1 \bar{t_1} \ .$$

Therefore, Eq. (10.13) is equal to

$$Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_1, H^a) = \frac{\left(\bar{b_1}t_1 + \gamma_1' b_1 \bar{t_1}\right)\lambda\tau}{\bar{b_1}t_1\left[1 - \lambda + \lambda\tau + \gamma_1\lambda\left(1-\tau\right)\right] + \gamma_1' b_1 \bar{t_1}} \ .$$

Inserting this result into Eq. (10.12), and Eq. (10.12) into Eq. (10.11), makes probability (e) in Eq. (10.8) equal to

$$Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a) = \bar{b_2}t_2\gamma_2\left\{\frac{\left(\bar{b_1}t_1 + \gamma_1' b_1 \bar{t_1}\right)\lambda\tau}{\bar{b_1}t_1\left[1 - \lambda + \lambda\tau + \gamma_1\lambda\left(1-\tau\right)\right] + \gamma_1' b_1 \bar{t_1}}\right\} \ .$$

Introducing the expressions for (e), (f), (g) and (h) into Eq. (10.8) produces

$$\frac{Pr(Y_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, H^a)}{Pr(Y_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_1, \bar{H}^a)} = \frac{\bar{b_2} t_2 \gamma_2 \left\{ \frac{(\bar{b_1} t_1 + \gamma_1' b_1 \bar{t_1})\lambda\tau}{\bar{b_1} t_1 [1-\lambda+\lambda\tau+\gamma_1\lambda(1-\tau)] + \gamma_1' b_1 \bar{t_1}} \right\} + \gamma_2' b_2 \bar{t_2}}{\bar{b_2} t_2' \gamma_2 + \gamma_2' b_2 \bar{t_2'}}$$

for ratio (ii) in Eq. (10.7).

Inserting all of the obtained results into Eq. (10.8) leads to the following activity level LR for a case in which the suspect's sample matches the first of two traces recovered on a crime scene:

$$LR = \underbrace{\frac{\bar{b_1} t_1 \left[1 - \lambda + \lambda\tau + \gamma_1 \lambda (1-\tau)\right] + \gamma_1' b_1 \bar{t_1}}{\bar{b_1} t_1' \gamma_1 + \gamma_1' b_1 t_1'}}_{\text{(i)}} \times \underbrace{\frac{\bar{b_2} t_2 \gamma_2 \left\{ \frac{(\bar{b_1} t_1 + \gamma_1' b_1 \bar{t_1})\lambda\tau}{\bar{b_1} t_1 [1-\lambda+\lambda\tau+\gamma_1\lambda(1-\tau)] + \gamma_1' b_1 \bar{t_1}} \right\} + \gamma_2' b_2 \bar{t_2}}{\bar{b_2} t_2' \gamma_2 + \gamma_2' b_2 \bar{t_2'}}}_{\text{(ii)}} .$$

In the case that the suspect's sample matches trace 2 (i.e., $X = \Gamma_2$)

In this case, probability (a), that is, $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|X = \Gamma_2, H^a)$, is the probability of the event that trace 1 was transferred during the struggle from an assailant having characteristic Γ_1, given that the suspect was an assailant who struggled with the victim on the crime scene, but that the suspect has characteristic Γ_2. This probability is equal to

$$\bar{b_1} \times t_1 \times Pr(TS_1 = \Gamma_1|X = \Gamma_2, H^a) .$$

In this case, the true source can only have characteristic Γ_1 if the suspect was not the assailant in location 1 (proposition $\bar{F}_1^a$), and if the assailant in location 1 possesses characteristic Γ_1 (probability of γ_1):

$$\begin{aligned} Pr(TS_1 = \Gamma_1|X = \Gamma_1, H^a) &= \gamma_1 \times \underbrace{Pr(\bar{F}_1^a|H^a)}_{\lambda(1-\tau)} \\ &= \gamma_1 \lambda (1-\tau) . \end{aligned}$$

The expression for $Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|X = \Gamma_2, H^a)$ is therefore $\bar{b_1} \times t_1 \times \gamma_1 \lambda (1-\tau)$.

Introducing the expressions for (a), (b), (c) and (d) into Eq. (10.8) produces

$$\frac{Pr(Y_1 = \Gamma_1|X = \Gamma_2, H^a)}{Pr(Y_1 = \Gamma_1|X = \Gamma_2, \bar{H}^a)} = \frac{\bar{b_1} t_1 \gamma_1 \lambda (1-\tau) + \gamma_1' b_1 \bar{t_1}}{\bar{b_1} t_1' \gamma_1 + \gamma_1' b_1 \bar{t_1'}} \tag{10.14}$$

for ratio (i) in Eq. (10.7).

Again, it is the expression for (e) which is more complex. $Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_2, H^a)$ is the probability of the event that trace 2 was transferred during the struggle from an assailant having characteristic Γ_2, given that trace 1 has characteristic Γ_1, the suspect has characteristic Γ_2 and the suspect was an assailant who struggled with the victim on the crime scene. This probability is equal to

$$\bar{b_2} \times t_2 \times Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_2, H^a) , \tag{10.15}$$

and the extension of the conversation to propositions F_2^a and $\bar{F}_2^a$ to find $Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_2, H^a)$ produces

$$\begin{aligned} &Pr(TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_2, H^a) \\ &\quad = 1 \times Pr(F_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) + \gamma_2 \times Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) . \end{aligned} \tag{10.16}$$

This development results from two possible explanations for the event $TS_2 = \Gamma_2$ (i.e., the event of the true source of a transferred trace in location 2 having characteristic Γ_2):

- either trace 2 was transferred from the suspect, in which case the suspect was the assailant who struggled with the victim in the location of trace 2 (proposition F_2^a);
- or trace 2 was transferred by another assailant (proposition $\bar{F}_2^a$), who has characteristic Γ_2 with a probability of γ_2.

The BN computes the probabilities $Pr(F_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a)$ and $Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a)$ by applying Bayes' theorem:

$$Pr(F_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) = \frac{Pr(Y_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) \times Pr(F_2^a|H^a)}{Pr(Y_1 = \Gamma_1|X = \Gamma_2, H^a)} \ , \tag{10.17}$$

$$Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) = \frac{Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) \times Pr(\bar{F}_2^a|H^a)}{Pr(Y_1 = \Gamma_1|X = \Gamma_2, H^a)} \ . \tag{10.18}$$

These two ratios are made up of the following five probabilities:

- We develop $Pr(Y_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a)$ by extending the conversation over nodes B_1, T_1, and TS_1:

$$\begin{aligned} &Pr(Y_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) \\ &\quad = 1 \times Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) + \gamma_1' \times Pr(B_1 \cap \bar{T}_1|H^a) \\ &\quad = \bar{b_1} t_1 \times Pr(TS_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) + \gamma_1' b_1 \bar{t_1} \ . \end{aligned} \tag{10.19}$$

  To find $Pr(TS_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a)$, we extend the conversation over propositions F_1^a and $\bar{F}_1^a$:

$$Pr(TS_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) = \gamma_1 \times Pr(\bar{F}_1^a|F_2^a, H^a) \ ,$$

  and to find $Pr(\bar{F}_1^a|F_2^a, H^a)$, we apply Bayes' theorem:

$$\begin{aligned} Pr(\bar{F}_1^a|F_2^a, H^a) &= \frac{\overbrace{Pr(F_2^a|\bar{F}_1^a, H^a)}^{1} \times \overbrace{Pr(\bar{F}_1^a|H^a)}^{\lambda(1-\tau)}}{\underbrace{Pr(F_2^a|H^a)}_{1-\lambda+\lambda(1-\tau)}} \\ &= \frac{\lambda(1-\tau)}{1-\lambda+\lambda(1-\tau)} \ . \end{aligned}$$

  Inserting these results into Eq. (10.19) produces

$$Pr(Y_1 = \Gamma_1|F_2^a, X = \Gamma_2, H^a) = \bar{b_1} t_1 \gamma_1 \left[\frac{\lambda(1-\tau)}{1-\lambda+\lambda(1-\tau)}\right] + \gamma_1' b_1 \bar{t_1} \ .$$

- By definition,

$$Pr(F_2^a|H^a) = 1 - \lambda + \lambda(1-\tau) \ ,$$

  that is, given that the suspect was an assailant, the suspect was the assailant in location 2 either when there was only one assailant in both locations (probability $1-\lambda$), or when there were two different assailants (probability λ), and the suspect struggled with the victim in location 2 (probability $1-\tau$).

- $Pr(Y_1 = \Gamma_1|X = \Gamma_2, H^a)$ is the numerator of the first ratio in our LR (Eq. (10.14)):

$$Pr(Y_1 = \Gamma_1|X = \Gamma_2, H^a) = \bar{b_1} t_1 \gamma_1 \lambda(1-\tau) + \gamma_1' b_1 \bar{t_1} \ .$$

- We develop $Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a)$ by extending the conversation over nodes B_1, T_1, and TS_1:

$$\begin{aligned} &Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) \\ &\quad = 1 \times Pr(\bar{B}_1 \cap T_1 \cap TS_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) + \gamma_1' \times Pr(B_1 \cap \bar{T}_1|H^a) \\ &\quad = \bar{b_1}t_1 \times Pr(TS_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) + \gamma_1' b_1 \bar{t_1} \ . \end{aligned} \tag{10.20}$$

  To find $Pr(TS_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a)$, we extend the conversation over propositions F_1^a and $\bar{F}_1^a$:

$$\begin{aligned} Pr(TS_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) &= \gamma_1 \times \underbrace{Pr(\bar{F}_1^a|\bar{F}_2^a, H^a)}_{0} \\ &= 0 \ . \end{aligned}$$

  Given that the suspect has characteristic Γ_2, the true source of a transferred trace in location 1 can only have characteristic Γ_1 if the assailant in location 1 was not the suspect (proposition $\bar{F}_1^a$). This unknown assailant possesses characteristic Γ_1 with a probability of γ_1. However, if the suspect was an assailant on the crime scene, then he must have been the assailant at either location 1 or location 2. According to our definitions of the propositions, it is therefore impossible that the suspect was not the assailant in location 1, given that he did not struggle with the victim in location 2, but was an assailant on the crime scene.

  Inserting this result into Eq. (10.20) produces

$$Pr(Y_1 = \Gamma_1|\bar{F}_2^a, X = \Gamma_2, H^a) = \gamma_1' b_1 \bar{t_1} \ .$$

  If the suspect was an assailant, yet did not struggle with the victim in location 2, he must have been the assailant in location 1. However, because the trace in location 1 does not match the suspect's sample, trace 1 can only have characteristic Γ_1 if it is a background trace.

- And, by definition,

$$Pr(\bar{F}_2^a|H^a) = \lambda\tau \ ,$$

  as in Eq. (10.13).

Therefore, Eq. (10.17) is equal to

$$\begin{aligned} Pr(F_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) &= \frac{\left\{\bar{b_1}t_1\gamma_1\left[\frac{\lambda(1-\tau)}{1-\lambda+\lambda(1-\tau)}\right] + \gamma_1' b_1 \bar{t_1}\right\} \times [1 - \lambda + \lambda(1-\tau)]}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1 \bar{t_1}} \\ &= \frac{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1 \bar{t_1}[1 - \lambda + \lambda(1-\tau)]}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1 \bar{t_1}} \ , \end{aligned}$$

and Eq. (10.18) to

$$Pr(\bar{F}_2^a|Y_1 = \Gamma_1, X = \Gamma_2, H^a) = \frac{\gamma_1' b_1 \bar{t_1}\lambda\tau}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1 \bar{t_1}} \ .$$

Inserting these results in Eq. (10.16), and Eq. (10.16) into Eq. (10.15), gives us the probability of (e):

$$\begin{aligned} &Pr(\bar{B}_2 \cap T_2 \cap TS_2 = \Gamma_2|Y_1 = \Gamma_1, X = \Gamma_2, H^a) \\ &= \bar{b_2}t_2\left\{\frac{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1\bar{t_1}[1 - \lambda + \lambda(1-\tau)]}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1\bar{t_1}} + \gamma_2\left[\frac{\gamma_1' b_1\bar{t_1}\lambda\tau}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1\bar{t_1}}\right]\right\} \\ &= \frac{\bar{b_2}t_2\left\{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1\bar{t_1}[1 - \lambda + \lambda(1-\tau)] + \gamma_2(\gamma_1' b_1\bar{t_1}\lambda\tau)\right\}}{\bar{b_1}t_1\gamma_1\lambda(1-\tau) + \gamma_1' b_1\bar{t_1}} \ . \end{aligned}$$

Introducing the expressions for (e), (f), (g) and (h) into Eq. (10.8) produces

$$\frac{Pr(Y_2=\Gamma_2|Y_1=\Gamma_1,X=\Gamma_2,H^a)}{Pr(Y_2=\Gamma_2|Y_1=\Gamma_1,X=\Gamma_2,\bar{H}^a)}=\frac{\frac{\bar{b}_2t_2\{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1[1-\lambda+\lambda(1-\tau)]+\gamma_2(\gamma_1'b_1\bar{t}_1\lambda\tau)\}}{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1}+\gamma_2'b_2\bar{t}_2}{\bar{b}_2t_2'\gamma_2+\gamma_2'b_2\bar{t}_2'}$$

for ratio (ii) in Eq. (10.7). Combining this result with Eq. (10.14), produces the following activity level LR for a case where the suspect's sample matches the second trace recovered on the crime scene:

$$LR=\underbrace{\frac{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1}{\bar{b}_1t_1'\gamma_1+\gamma_1'b_1\bar{t}_1'}}_{\text{(i)}}\times\underbrace{\frac{\frac{\bar{b}_2t_2\{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1[1-\lambda+\lambda(1-\tau)]+\gamma_2(\gamma_1'b_1\bar{t}_1\lambda\tau)\}}{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1}+\gamma_2'b_2\bar{t}_2}{\bar{b}_2t_2'\gamma_2+\gamma_2'b_2\bar{t}_2'}}_{\text{(ii)}}.$$

Summary

The LR computed by the BN in Fig. 10.6c is the product of two ratios: one pertaining to the observation of the matching trace, and the other to the observation of the non-matching trace. Whatever the order of these observations, the second observation is always conditional on the analytical result of the first trace observed.

If trace 1 matches the suspect's sample, the LR is:

$$LR=\underbrace{\frac{\bar{b}_1t_1\left[1-\lambda+\lambda\tau+\gamma_1\lambda(1-\tau)\right]+\gamma_1'b_1\bar{t}_1}{\bar{b}_1t_1'\gamma_1+\gamma_1'b_1\bar{t}_1'}}_{\textit{matching trace}}\times\underbrace{\frac{\bar{b}_2t_2\gamma_2\left\{\frac{(\bar{b}_1t_1+\gamma_1'b_1\bar{t}_1)\lambda\tau}{\bar{b}_1t_1[1-\lambda+\lambda\tau+\gamma_1\lambda(1-\tau)]+\gamma_1'b_1\bar{t}_1}\right\}+\gamma_2'b_2\bar{t}_2}{\bar{b}_2t_2'\gamma_2+\gamma_2'b_2\bar{t}_2'}}_{\textit{non-matching trace given matching trace}},\tag{10.21}$$

and if trace 2 matches the suspect's sample, the LR is:

$$LR=\underbrace{\frac{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1}{\bar{b}_1t_1'\gamma_1+\gamma_1'b_1\bar{t}_1'}}_{\textit{non-matching trace}}\times\underbrace{\frac{\frac{\bar{b}_2t_2\{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1[1-\lambda+\lambda(1-\tau)]+\gamma_2(\gamma_1'b_1\bar{t}_1\lambda\tau)\}}{\bar{b}_1t_1\gamma_1\lambda(1-\tau)+\gamma_1'b_1\bar{t}_1}+\gamma_2'b_2\bar{t}_2}{\bar{b}_2t_2'\gamma_2+\gamma_2'b_2\bar{t}_2'}}_{\textit{matching trace given non-matching trace}}.\tag{10.22}$$

Note that each ratio in these equations is an extension of the LR published for the one-trace transfer problem (Eq. (10.4)). In fact, the denominator and the second term in the numerator remain unchanged. The extension affects only the first term in the numerator, that is, the product $\bar{b}t$. This product is the probability that the recovered trace was transferred by an assailant during the alleged activity. It must be multiplied by a factor corresponding to the probability of the transferred trace's source having the observed characteristic. This factor will vary according to whether we consider one trace (in Eq. (10.4) for the one-trace transfer problem, this factor is equal to 1) or two traces, or more accurately, whether we consider a case where there is the possibility of more than one assailant (Triggs and Buckleton, 2003). In the latter case, it will further depend on whether we consider the matching trace or the non-matching trace, and the first or the second of the two recovered traces. Note that Eqs. (10.21) and (10.22) contain parameters which are actually prior probabilities. For this reason, the LR that is obtained is not to be intended in its usual sense because it appears that it no longer depends only upon the sample data.

To compare these expressions with each other and with Eq. (10.1), we multiply the two

ratios, and rewrite them in the form of Eq. (10.1). Thus Eq. (10.21) becomes:

$$\begin{aligned} LR &= \frac{\bar{b_2}t_2\gamma_2\left[\left(\bar{b_1}t_1 + b_1\gamma_1'\bar{t_1}\right)\lambda\tau\right] + \gamma_2'b_2\bar{t_2}\left\{\bar{b_1}t_1\left[1-\lambda+\lambda\tau+\gamma_1\lambda\left(1-\tau\right)\right] + \gamma_1'b_1\bar{t_1}\right\}}{\left(\bar{b_1}t_1'\gamma_1 + \gamma_1'b_1\bar{t_1'}\right)\times\left(\bar{b_2}t_2'\gamma_2 + \gamma_2'b_2\bar{t_2'}\right)} \\ &= \frac{\bar{b_1}\bar{b_2}t_1t_2\lambda\tau\gamma_2 + \bar{b_1}b_2t_1\bar{t_2}\left[1-\lambda+\lambda\tau+\lambda\left(1-\tau\right)\gamma_1\right]\gamma_2' + b_1\bar{b_2}\bar{t_1}t_2\lambda\tau\gamma_1'\gamma_2 + b_1b_2\bar{t_1}\bar{t_2}\gamma_1'\gamma_2'}{\bar{b_1}\bar{b_2}t_1't_2'\gamma_1\gamma_2 + \bar{b_1}b_2t_1'\bar{t_2'}\gamma_1\gamma_2' + b_1\bar{b_2}\bar{t_1'}t_2'\gamma_1'\gamma_2 + b_1b_2\bar{t_1'}\bar{t_2'}\gamma_1'\gamma_2'}, \end{aligned} \tag{10.23}$$

and Eq. (10.22):

$$\begin{aligned} LR &= \frac{\bar{b_2}t_2\left\{\bar{b_1}t_1\gamma_1\lambda\left(1-\tau\right) + \gamma_1'b_1\bar{t_1}\left[1-\lambda+\lambda\left(1-\tau\right)\right] + \gamma_2\left(\gamma_1'b_1\bar{t_1}\lambda\tau\right)\right\} + \gamma_2'b_2\bar{t_2}\left[\bar{b_1}t_1\gamma_1\lambda\left(1-\tau\right) + \gamma_1'b_1\bar{t_1}\right]}{\left(\bar{b_1}t_1'\gamma_1 + \gamma_1'b_1\bar{t_1'}\right)\times\left(\bar{b_2}t_2'\gamma_2 + \gamma_2'b_2\bar{t_2'}\right)} \\ &= \frac{\bar{b_1}\bar{b_2}t_1t_2\lambda\left(1-\tau\right)\gamma_1 + \bar{b_1}b_2t_1\bar{t_2}\lambda\left(1-\tau\right)\gamma_1\gamma_2' + b_1\bar{b_2}\bar{t_1}t_2\left[1-\lambda+\lambda\left(1-\tau\right)+\lambda\tau\gamma_2\right]\gamma_1' + b_1b_2\bar{t_1}\bar{t_2}\gamma_1'\gamma_2'}{\bar{b_1}\bar{b_2}t_1't_2'\gamma_1\gamma_2 + \bar{b_1}b_2t_1'\bar{t_2'}\gamma_1\gamma_2' + b_1\bar{b_2}\bar{t_1'}t_2'\gamma_1'\gamma_2 + b_1b_2\bar{t_1'}\bar{t_2'}\gamma_1'\gamma_2'}. \end{aligned} \tag{10.24}$$

These two equations are identical for $\tau = 0.5$. In this case, we do not need to differentiate between trace 1 and trace 2, and only need to distinguish between the matching trace $i \in \{1,2\}$ with characteristic Γ_i, and the non-matching trace $j \in \{1,2\}$, $j \neq i$, with characteristic Γ_j:

$$LR = \frac{\frac{1}{2}\bar{b_i}\bar{b_j}t_it_j\lambda\gamma_j + \bar{b_i}b_jt_i\bar{t_j}\left[1-\lambda+\frac{1}{2}\lambda+\frac{1}{2}\lambda\gamma_i\right]\gamma_j' + \frac{1}{2}b_i\bar{b_j}\bar{t_i}t_j\lambda\gamma_i'\gamma_j + b_ib_j\bar{t_i}\bar{t_j}\gamma_i'\gamma_j'}{\bar{b_i}\bar{b_j}t_i't_j'\gamma_i\gamma_j + \bar{b_i}b_jt_i'\bar{t_j'}\gamma_i\gamma_j' + b_i\bar{b_j}\bar{t_i'}t_j'\gamma_i'\gamma_j + b_ib_j\bar{t_i'}\bar{t_j'}\gamma_i'\gamma_j'}. \tag{10.25}$$

The numerator and denominator of this equation each consist of the four possible combinations of background and transferred traces for the matching and the non-matching trace (i.e., both traces were transferred, only the matching trace was transferred, only the non-matching trace was transferred, and both traces are background traces). In the denominator and in the fourth term of the numerator, the probabilities of the observations consist of the product of the corresponding background, transfer and match probabilities. These describe events that are independent of the suspect's characteristic and of the suspect's possible involvement in the struggle with the victim. The first three terms of the numerator, however, are not independent of the suspect's involvement. These contain the additional probabilities of λ and τ (which, in Eq. (10.25) is equal to $\frac{1}{2}$), in addition to the background, transfer, and match probabilities. More specifically:

- The first term in the numerator considers the event that both traces were transferred during the struggle. Apart from the appropriate background and transfer probabilities, the probability of observing one matching and one non-matching trace, given that the suspect was an assailant, is equal to the match probability of the non-matching trace's characteristic (i.e., γ_j) times the probability that there were two different assailants (i.e., λ) times the probability that the suspect was the assailant in location 1 (in this case, $\frac{1}{2}$).

- The second term in the numerator describes the event that only the matching trace was transferred during the struggle. In this case, there are three possibilities: the suspect could have been the assailant in both locations (probability of $1-\lambda$), the suspect could have been the assailant only in the location of the matching trace (here, probability of $\frac{1}{2}\lambda$), or the suspect could have been the assailant only in the location of the non-matching trace (here, probability of $\frac{1}{2}\lambda$). In the last case, the assailant in the other location also had the matching characteristic with a probability of γ_i. The sum of the probabilities of each of these possible events makes up the additional factor in the square brackets in Eq. (10.25).

- The third term in the numerator describes the event that only the non-matching trace was transferred. As the suspect could not have transferred the non-matching trace, this is only possible if the suspect was the assailant at the location of the matching trace (here, probability of $\frac{1}{2}\lambda$).

The LR computed by the BN thus combines the probabilities defined in this paper in a logical way.

10.6 Discussion

In this section, we analyze Eqs. (10.23) - (10.25). First, we compare these expressions to Eq. (10.1); then, we show how Eq. (10.25) may reduce to both the source level LR for two traces and the activity level LR for a single trace under the appropriate assumptions; and finally, we discuss the extension of the model to n traces.

10.6.1 Comparison with Eq. (10.1)

Rewriting Eq. (10.1) with the subscript i for the probabilities referring to the matching trace, and the subscript j for the probabilities referring to the non-matching trace, produces (Triggs and Buckleton, 2003):

$$LR = \frac{\frac{1}{2}\bar{b}_i\bar{b}_j t_i t_j \left(1-2q\right)\gamma_j + \frac{1}{2}\bar{b}_i b_j t_i \bar{t}_j \left(1+\gamma_i\right)\gamma'_j + \frac{1}{2} b_i \bar{b}_j \bar{t}_i t_j \gamma'_i \gamma_j + b_i b_j \bar{t}_i \bar{t}_j \gamma'_i \gamma'_j}{\bar{b}_i\bar{b}_j t_i t_j \left(1-2q\right)\gamma_i\gamma_j + \bar{b}_i b_j t_i \bar{t}_j \gamma_i \gamma'_j + b_i \bar{b}_j \bar{t}_i t_j \gamma'_i \gamma_j + b_i b_j \bar{t}_i \bar{t}_j \gamma'_i \gamma'_j} \quad . \tag{10.26}$$

A comparison of this equation with Eq. (10.25) shows that this equation assumes $\tau = 0.5$, and uses the variable q whereas we have used the variable λ. The differences are due to the different definitions underlying q and λ. This can be seen by setting $q = 0$ and $\lambda = 1$: in this case, these variables disappear from the equations and the two LRs become identical.

The difference between q and λ is that the definition of q is limited to the event of two transferred traces, whereas λ is defined at the level of the propositions, independently of whether the traces were transferred during the struggle. More specifically, expression $1-2q$ (in Eq. (10.26)) denotes the probability that two transferred traces come from different assailants (Triggs and Buckleton, 2003). It therefore only applies to the first term in the numerator and the first term in the denominator, where the probabilities describe the event of two transferred traces. Probability λ (in Eq. (10.25)), on the other hand, describes the prior probability that the assailant in location 1 was not the same assailant as the assailant in location 2. This definition is not limited to only the transferred traces, and therefore appears in the first, second and third terms of the LR's numerator (Eq. (10.25)). Note that λ does not figure in the LR's denominator, because the denominator considers the observations of the two traces independently of each other.

Owing to these different definitions:

- the additional factor of $1-2q$ in the first term of the denominator of Eq. (10.26) makes this denominator smaller than the denominator of Eq. (10.25);
- the additional factor of $1-\lambda$ in the second term of the numerator in Eq. (10.25) makes this term greater in Eq. (10.25) than in Eq. (10.26); and
- probability λ in the third term of the numerator in Eq. (10.25) makes this term smaller in Eq. (10.25) than in Eq. (10.26).

Numerically, the impact of these differences depends on the values assigned to the background and transfer probabilities. (That is, in most cases Eq. (10.26) will produce a slightly greater LR, yet if large values are assigned to b_i and t_j, and small values to b_j and t_i, Eq. (10.25) may produce the greater LR owing to the greater impact of the numerator's third term.)

10.6.2 Verification of the model

To verify the results produced by our model, we show that Eq. (10.25) reduces to the source level LR for two traces and to the activity level LR for a single trace when assumptions are made to simulate these two situations.

First, we observe that letting the background (b_i and b_j) and transfer probabilities (t_i, t'_i, t_j, and t'_j) tend toward 1 or 0 leads to the expected, logical results: the computed LR is an increasing function of t_i and a decreasing function of t_j. If $t_i = t'_i = t_j = t'_j = 0$, it reduces to 1, and when $t_i = t'_i = t_j = t'_j = 1$ and $\bar{b_i} = \bar{b_j} = 1$, it tends toward the source level LR, which is $\frac{1}{2\gamma_i}$ for $\lambda = 1$ and $\tau = \frac{1}{2}$ (Evett, 1987):

$$\begin{aligned} LR &= \frac{\frac{1}{2}\gamma_j}{\gamma_i\gamma_j} \\ &= \frac{1}{2\gamma_i} \ . \end{aligned}$$

Second, the computations of the BN can also be reduced to the activity level LR for a single trace (Eq. (10.4)). To do this, we consider a scenario in which *another* assailant, matching trace j, has already been found, and one assumes that if trace j was transferred by one of the assailants, then it was transferred by this other assailant. We extended Fig. 10.6(c) to illustrate this new situation by adding a node Z for the second suspect, and a node H^Z for a second pair of general propositions, pertaining to this second suspect (Fig. 10.7). The LR

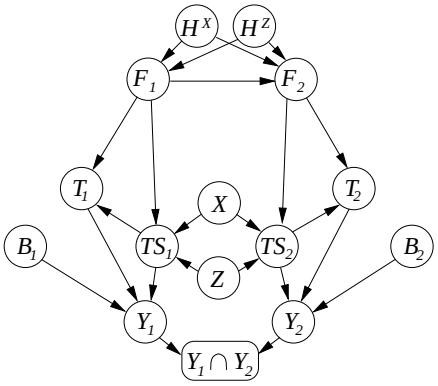


Figure 10.7: *BN extended to include a second suspect, denoted Z, in a case where there were two different assailants. A node H^Z was added containing the same propositions as in H (now renamed H^X), but for this second suspect.*

deduced from this model with respect to the first suspect (see Appendix for the derivation) is now:

$$LR = \frac{\bar{b_i}t_i + \gamma'_i b_i \bar{t_i}}{\bar{b_i}t'_i\gamma_i + \gamma'_i b_i \bar{t_i}} \ ,$$

which corresponds to the activity level LR (Eq. (10.4)) for a single trace (Evett, 1984).

10.6.3 Extension to n traces

In an $n-trace$ transfer problem, we consider n different traces recovered on the crime scene to come from n distinct sources. The organized structure of the BN in Fig. 10.6(c) easily lends it to an extension to any number of traces by transforming it into an *object oriented BN* (OOBN). An OOBN allows the user to evaluate more complex problems by combining different objects in a hierarchical structure (Hepler et al., 2007). An object may be a simple random variable (like the nodes in a regular BN), or a separate, complex model,

such as another BN (Koller and Pfeffer, 1997). Thus, the main advantage of an OOBN is its capacity to differentiate between several hierarchical levels and to combine a set of nodes from different models. This is particularly useful for combining a set of identical network fragments that form a repetitive pattern in a regular BN (Dawid et al., 2007; Kjaerulff and Madsen, 2008).

In the case of the two-trace problem, one can represent each group of variables specific to one trace as a separate object. In this extension of the model, we assume that $\lambda = 1$ (i.e., there are n different perpetrators for a case with n traces), and omit node L in the BN. Thus, the two-trace problem in Fig. 10.6(c) becomes an OOBN with only four objects: two random variables and two subnetworks (Fig. 10.8(a)). The two network fragments hidden

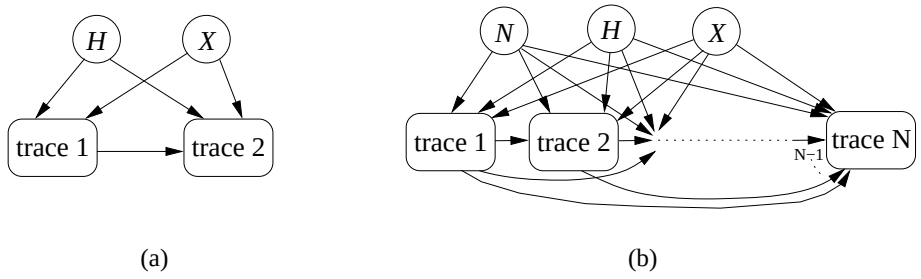


Figure 10.8: *Object-oriented BNs for* **(a)** *two traces, and* **(b)** *n traces. In (b), the additional node N allows the user to specify the number of different traces the BN should consider in the evaluation. Here, the dotted line represents traces 3 to N − 1. Each additional trace has an incoming arrow from nodes N, H, and X, and from each of the previously observed traces.*

in the interface nodes of *trace 1* and *trace 2* in Fig. 10.8(a) are shown in Fig. 10.9(a),(b), respectively. This OOBN has the same structure as the BN in Fig. 10.6(c) (without node

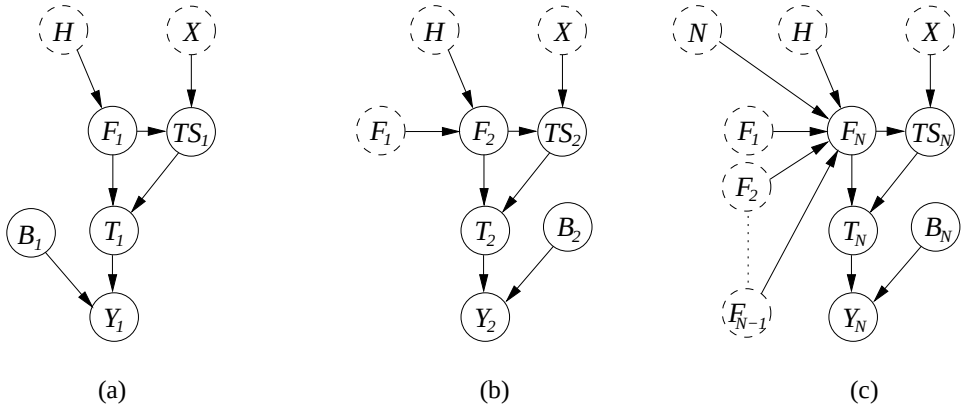


Figure 10.9: *The network fragments hidden in the interface nodes of* **(a)** *trace 1,* **(b)** *trace 2, and* **(c)** *trace N in Fig. 10.8. The nodes with a dashed contour are nodes figuring either in the master network in Fig. 10.8, or in a different network fragment. The dotted line in (c) represents nodes F_3 to F_{N-2}. These are all a parent to node F_N.*

L), only decomposed into three separate elements.

The OOBN structure for two traces suggests a logical way to extend the model to additional traces. New traces are added in the same way trace 2 was added to trace 1: the observation of each new trace's characteristic depends on the general variables H and X, and on the specific hypotheses (contained in nodes of type F) of each of the previously observed traces (Fig. 10.8(b) and 10.9(c)). By this means, one can construct a general model for m different traces ($m \geq n$), and then designate, through an additional node N, the number of different traces n for which one would like the BN to compute an LR.

The program *Hugin Researcher* is only limited by the amount of memory it can use.

This limit lies at 4GB, which is great enough to allow for hundreds of traces to be modelled with this OOBN (the exact number of traces will depend on the number of analytical traits defined as the states of nodes X, Y and TS).

The additional variable N also allows the user to introduce an uncertainty on the number of sources if this is not clearly defined by the circumstantial information of the case. This OOBN clearly describes the dependencies assumed among the variables, and rigorously applies the laws of probability to compute the LR of interest.

10.7 Conclusions

Forensic scientists are faced with the need of addressing increasingly complex inference problems for assessing the value of scientific evidence. Two-trace problems are a typical example for this. They are a realistic problem which, up to now, forensic statisticians have addressed with an algebraic approach for calculating LRs. These applications have led to efficient results for simple evidential assessments, yet quickly lead to mathematically sophisticated expressions when applied to more complex problems. For an increasing number of variables and an increasing number of conditional probabilistic relationships between these variables, purely theoretical developments make it difficult to maintain a transparent and error-free approach. The algebraic approach thus reaches its limits when it is applied to increasingly complex inference problems.

The aim of this study was to investigate a new way for computing LRs, a graphical approach based on the construction of BNs. These graphical models overcome the hurdle of complexity by:

- decomposing complicated events into a set of distinct variables;
- describing and visualizing the assumed dependencies among the variables;
- rigorously handling probabilistic calculations in a mathematically robust environment;
- easily incorporating additional variables into existing models; and
- coherently combining and structuring different aspects of a problem as separate objects in distinct hierarchical levels of an OOBN.

Thus, the construction of BNs provides a transparent approach to inference problems that is not limited by an increasing number of variables and probabilistic relationships, and not prone to careless mathematical errors that may occur when using or developing an algebraic formula.

In the context of the two-trace transfer problem, the development of a BN demonstrated the potential of such graphical probability models by producing a new activity level LR that relaxes assumptions made in previous algebraic developments. In addition, the graphical structure readily presents itself to extensions to more complex problems such as the n-trace problem at the activity level. Thus, the development of BNs allows forensic scientists to progress in the field of evidential interpretation by providing a tool to tackle more complex inference problems in a structured and logical way.

10.8 Appendix: Derivation for the activity level LR for Fig. 10.7

The aim of this derivation is to show that the introduction of a second suspect (denoted Z) matching trace j allows one to reduce the activity level LR for two traces to the activity level LR for a single trace. For this evaluation, we assume that $\lambda = 1$, that is, that there were two different assailants. Further, we assume that the second suspect was engaged in a struggle with the victim in the location of trace j (i.e., proposition $F^a_{j,Z}$). To differentiate the

propositions referring to suspect 1 from those referring to suspect 2, the former now contain an additional subscript X, and the latter a subscript Z (see Table 10.10 for the definitions of these different propositions). We want to obtain the LR for suspect X. With respect

Table 10.10: *Definitions of the additional propositions that figure in the BN shown in Fig. 10.7.*

Nodes	States	Definitions of the States
H^X	H_X^a	Suspect 1 was engaged in a struggle with the victim
	$\bar{H}_X^a$	Suspect 1 was not engaged in a struggle with the victim
H^Z	H_Z^a	Suspect 2 was engaged in a struggle with the victim
	$\bar{H}_Z^a$	Suspect 2 was not engaged in a struggle with the victim
F_i	$F_{i,X}^a$	Suspect 1 was engaged in a struggle with the victim in the location of trace i
	$F_{i,Z}^a$	Suspect 2 was engaged in a struggle with the victim in the location of trace i
	$\bar{F}_i^a$	Neither suspect 1, nor suspect 2, was engaged in a struggle with the victim in the location of trace i
$F_{j\neq i}$	$F_{j\neq i,X}^a$	Suspect 1 was engaged in a struggle with the victim in the location of trace j
	$F_{j\neq i,Z}^a$	Suspect 2 was engaged in a struggle with the victim in the location of trace j
	$\bar{F}_{j\neq i}^a$	Neither suspect 1, nor suspect 2, was engaged in a struggle with the victim in the location of trace j

to suspect X, we shall assume that we observe first the non-matching trace j, and then the matching trace i. The LR is computed for propositions H_X^a and $\bar{H}_X^a$, given proposition $F_{j\neq i,Z}^a$, denoted in the following developments as $F_{j,Z}^a$:

$$LR = \frac{Pr(Y_j|X,Z,F_{j,Z}^a,H_X^a)}{Pr(Y_j|X,Z,F_{j,Z}^a,\bar{H}_X^a)} \times \frac{Pr(Y_i|Y_j,X,Z,F_{j,Z}^a,H_X^a)}{Pr(Y_i|Y_j,X,Z,F_{j,Z}^a,\bar{H}_X^a)} \ . \tag{10.27}$$

We define the observations as: $Y_j = \Gamma_j$, $Y_i = \Gamma_i$, $Z = \Gamma_j$ and $X = \Gamma_i$.

For the numerator of the first ratio, the extension of the conversation over variables B_j, T_j, and TS_j produces:

$$\begin{aligned} Pr(Y_j|X,Z,F_{j,Z}^a,H_X^a) &= 1 \times Pr(\bar{B}_j \cap T_j \cap TS_j = \Gamma_j|Z,F_{j,Z}^a,H_X^a) + \gamma_j' \times Pr(B_j \cap \bar{T}_j|H_X^a) \\ &= \bar{b_j}t_j \times Pr(TS_j = \Gamma_j|Z,F_{j,Z}^a) + \gamma_j' b_j \bar{t_j} \ . \end{aligned}$$

Given $F_{j,Z}^a$ and $Z = \Gamma_j$,

$$Pr(TS_j = \Gamma_j|Z,F_{j,Z}^a) = 1$$

such that

$$Pr(Y_j|X,Z,F_{j,Z}^a,H_X^a) = \bar{b_j}t_j + \gamma_j' b_j \bar{t_j} \ .$$

For the denominator of the first ratio, we observe the same development:

$$\begin{aligned} Pr(Y_j|X,Z,F_{j,Z}^a,\bar{H}_X^a) &= 1 \times Pr(\bar{B}_j \cap T_j \cap TS_j = \Gamma_j|Z,F_{j,Z}^a,\bar{H}_X^a) + \gamma_j' \times Pr(B_j \cap \bar{T}_j|\bar{H}_X^a) \\ &= \bar{b_j}t_j \times Pr(TS_j = \Gamma_j|Z,F_{j,Z}^a) + \gamma_j' b_j \bar{t_j} \ , \end{aligned}$$

with

$$Pr(TS_j = \Gamma_j|Z,F_{j,Z}^a) = 1$$

such that

$$Pr(Y_j|X,Z,F_{j,Z}^a,\bar{H}_X^a) = \bar{b_j}t_j + \gamma_j' b_j \bar{t_j} \ .$$

Given $F_{j,Z}^a$ and $Z = \Gamma_j$, the event $Y_j = \Gamma_j$ is independent of the propositions pertaining to suspect 1 (H_X^a and $\bar{H}_X^a$). This makes the first ratio in Eq. (10.27) equal to 1.

The extension of the conversation for the numerator of the second ratio produces:

$$\begin{aligned} Pr(Y_i|Y_j, X, Z, F^a_{j,Z}, H^a_X) &= 1 \times Pr(\bar{B}_i \cap T_i \cap TS_i = \Gamma_i|X, F^a_{j,Z}, H^a_X) + \gamma'_i \times Pr(B_i \cap \bar{T}_i|H^a_X) \\ &= \bar{b_i} t_i \times Pr(TS_i = \Gamma_i|X, F^a_{j,Z}, H^a_X) + \gamma'_i b_i \bar{t_i} \ . \end{aligned}$$

Given that suspect 1 was one of the assailants (proposition H^a_X), and that suspect 2 was the assailant in the location of trace j (proposition $F^a_{j\neq i,Z}$), suspect 1 must have been the assailant in the location of trace i (proposition $F^a_{i,X}$). Therefore,

$$\begin{aligned} Pr(TS_i = \Gamma_i|X, F^a_{j,Z}, H^a_X) &= Pr(TS_i = \Gamma_i|X, F^a_{i,X}) \\ &= 1 \ , \end{aligned}$$

and

$$Pr(Y_i|Y_j, X, Z, F^a_{j,Z}, H^a_X) = \bar{b_i} t_i + \gamma'_i b_i \bar{t_i} \ .$$

For the denominator of the second ratio, we obtain:

$$\begin{aligned} Pr(Y_i|Y_j, X, Z, F^a_{j,Z}, \bar{H}^a_X) &= 1 \times Pr(\bar{B}_i \cap T_i \cap TS_i = \Gamma_i|\bar{H}^a_X) + \gamma'_i \times Pr(B_i \cap \bar{T}_i|\bar{H}^a_X) \\ &= \bar{b_i} t'_i \times Pr(TS_i = \Gamma_i|\bar{H}^a_X) + \gamma'_i b_i \bar{t'_i} \ . \end{aligned}$$

Under $\bar{H}^a_X$, suspect 1 was not one of the assailants, and trace i must have been transferred by an unknown assailant if it was transferred during the assault. This other assailant has characteristic Γ_i with a probability of γ_i:

$$Pr(TS_i = \Gamma_i|\bar{H}^a_X) = \gamma_i \ ,$$

and

$$Pr(Y_i|Y_j, X, Z, F^a_{j,Z}, \bar{H}^a_X) = \bar{b_i} t'_i \gamma_i + \gamma'_i b_i \bar{t'_i} \ .$$

Therefore, Eq. (10.27) is equal to:

$$\begin{aligned} LR &= \frac{\bar{b_j} t_j + \gamma'_j b_j \bar{t_j}}{\bar{b_j} t_j + \gamma'_j b_j \bar{t_j}} \times \frac{\bar{b_i} t_i + \gamma'_i b_i \bar{t_i}}{\bar{b_i} t'_i \gamma_i + \gamma'_i b_i \bar{t'_i}} \\ &= 1 \times \frac{\bar{b_i} t_i + \gamma'_i b_i \bar{t_i}}{\bar{b_i} t'_i \gamma_i + \gamma'_i b_i \bar{t'_i}} \\ &= \frac{\bar{b_i} t_i + \gamma'_i b_i \bar{t_i}}{\bar{b_i} t'_i \gamma_i + \gamma'_i b_i \bar{t'_i}} \ . \end{aligned}$$

This LR is equal to the activity level LR for a single trace (Eq. (10.4)).

Chapter 11

The Database Search Problem: A Question of Rational Decision Making

Abstract

This paper applies probability and decision theory in the graphical interface of an influence diagram to study the formal requirements of rationality which justify the individualization of a person found through a database search. The decision-theoretic part of the analysis studies the parameters that a rational decision maker would use to individualize the selected person. The modeling part (in the form of an influence diagram) clarifies the relationships between this decision and the ingredients that make up the database search problem, i.e., the results of the database search and the different pairs of propositions describing whether an individual is at the source of the crime stain. These analyses evaluate the desirability associated with the decision of 'individualizing' (and 'not individualizing'). They point out that this decision is a function of (i) the probability that the individual in question is, in fact, at the source of the crime stain (i.e., the state of nature), and (ii) the decision maker's preferences among the possible consequences of the decision (i.e., the decision maker's loss function). We discuss the relevance and argumentative implications of these insights with respect to recent comments in specialized literature, which suggest points of view that are opposed to the results of our study.

11.1 Introduction

The 'classical' database search problem, as it is known throughout forensic and legal theory and practice, relates to a question of the following kind: "What is the strength of the evidence against a given individual found through a database search, when that individual is the only person in the database who presents the same analytical characteristics (such as a DNA profile) as those observed on a crime stain?" After intense and controversial debates, starting in the mid-1990s, and which seemed to have been settled during the last decade, the

Reprinted from Forensic Science International, 222, S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni, The database search problem: A question of rational decision making, pages 186-199, Copyright (2012), with permission from Elsevier.

database search problem has once more become the object of several publications (Schneider et al., 2010; Taroni et al., 2011; Fimmers et al., 2011; Biedermann et al., 2011a). In particular, Schneider et al. (2010) and Fimmers et al. (2011) recently claimed that a single matching profile found through a database search reduces the evidential value of this match compared to a match found by other investigational means (i.e., a situation in which no database search was conducted). However, there are now ample counterarguments demonstrating that this is a misconception (e.g., Balding and Donnelly, 1996; Evett and Weir, 1998; Donnelly and Friedman, 1999; Evett et al., 2000a; Dawid, 2001; Balding, 2002) dating back to the NRC reports (National Research Council (NRC) Committee on DNA Technology in Forensic Science, 1992; National Research Council (NRC) Committee on DNA Forensic Science: An Update, 1996) and writings by Stockmarr (1999). These latter accounts are constructed around a conceptually unsuitable pair of propositions, defined as follows:

H_{db}: the source of the crime stain is in the database;
$\neg H_{db}$: the source of the crime stain is not in the database.

This contradicts probabilistic arguments that demonstrate an increase in the evidential value of a single database match when one considers the conventional and procedurally appropriate pair of propositions, which take the form of:

H_i: the crime stain comes from individual i;
$\neg H_i$: the crime stain comes from someone else unrelated to individual i;

where $i = 1, \ldots, N$, and N is the size of the population of individuals who could have been at the source of the crime stain. The increase for this pair of propositions is due to the exclusion of $n-1$ non-matching profiles (where n denotes the size of the database searched). This argument is now covered to a great extent in existing literature (e.g., Balding and Donnelly, 1996; Evett and Weir, 1998; Donnelly and Friedman, 1999; Evett et al., 2000a; Dawid, 2001; Balding, 2002; Balding and Donnelly, 1995; Kaye, 2009), and currently appears to accumulate the most widespread support.

In their recent publication, Fimmers et al. (2011) seek to take their argument in support of a decrease in the value of a database match a step further: they addressed the act of convicting a suspect and the probability that this conviction is false. That is, they passed from a purely probabilistic discourse to an argument invoking the act of choosing a particular option among several possible options. Their argumentation consists of a hypothetical case, in which investigators search for the individual at the source of a biological stain recovered on a crime scene. The investigators in this case consider a population of 100 million individuals ($N = 10^8$) as the population of potential sources, and possess a database containing the profiles of one million of these individuals ($n = 10^6$). In this population, the DNA profile of the crime stain has a match probability of $\gamma = 10^{-6}$. However, in their example, Fimmers et al. (2011) assume that the true source of the crime stain is *not* in the population considered by the investigators, that is, not among the $N = 10^8$ individuals, and consequently, not among the $n = 10^6$ profiles in the database (since the database contains profiles taken from the population of the $N = 10^8$ individuals). Assuming that "a suspect will certainly be convicted in every case in which the rarity of the corresponding DNA profile is [at least] one in a million" (Fimmers et al., 2011, page 4),[1] that is, when $\gamma \leq 10^{-6}$, they then compare the probability of a false conviction in a *probable cause case*[2] to the probability of a false conviction in a *database search case*[3] for an incriminating profile with a match probability of 10^{-6}:

[1] Translation by the current authors, and words in square brackets added by the current authors.

[2] In a *probable cause case* (also referred to as the *confirmation case* (Donnelly and Friedman, 1999; Kaye, 2009)), a suspect found on the basis of information not related to the DNA profile has a profile that is subsequently found to match the crime mark's profile.

[3] In a *database search case* (also referred to as the *trawl case* (Donnelly and Friedman, 1999; Kaye, 2009)), a search in a database results in a single hit with an individual, who then becomes a suspect in this case because of the matching profile.

- *probable cause case*: "There is a suspect. The DNA profile of that person is determined and found to correspond to that of the crime stain. The person is going to be convicted on the sole basis of this correspondence. Given the assumptions in this example, we know that the conviction is erroneous, because the true author has escaped. How high are the odds, in our scenario, of this to happen by chance? The probability of the DNA profile is 1:1,000,000 and this is the probability for a correspondence by chance with the stain. The probability for a false decision is thus 0.000001." (Fimmers et al., 2011, page 4)[4]

- *database search case*: "There is no suspect. A search in the database is conducted, and exactly one person is found. That person is convicted on the basis of the same argument as that in scenario 1 [the probable cause case]. The conviction is of course, again, false, because the data of the true author are not stored in the database. What is the probability for such a false decision? The answer is somewhat more complicated than that in scenario 1 [the probable cause case]. An error occurs notably when exactly one person is found in the database. (...) We will find exactly one person with a probability of 0.368 (that is in approximately every third similar case), and this person will subsequently be convicted, even though the true author is not in the database. The probability for an error in scenario 2 [the database search case] is therefore considerably greater than in scenario 1 [the probable cause case]." (Fimmers et al., 2011, page 4)[5]

Based on this reasoning, Fimmers et al. argue (page 4):

> "The simple evaluation using a likelihood ratio, as proposed by Taroni et al. (2010) is appropriate for the first scenario [the probable cause case], yet produces an unjustifiably high number of false decisions in the second scenario [the database search case]".[5]

This is questionable, however, because a likelihood ratio in no way amounts to a categorical conclusion with respect to the process of individualization (i.e., the attribution of the trace to a single source to the exclusion of all other potential sources) (Biedermann et al., 2011a). In Fimmers et al. (2011)'s framework, every match results in a wrong individualization. Since every comparison of the crime stain's profile with the profile of an individual in the population has a probability of 10^{-6} of leading to a match, every comparison has a probability of 10^{-6} of leading to a false individualization. It is therefore hardly surprising that Fimmers et al.'s probability of a false individualization increases with the number of comparisons performed. In other words, one comparison in the probable cause case has a probability of

$$\gamma = 10^{-6}$$

of matching the crime stain's profile, whereas one million comparisons in the database search case have a probability of

$$n\gamma(1-\gamma)^{n-1} = 0.368$$

of leading to a match with the crime stain's profile. This reasoning process consists of an unrealistic deduction based solely on the evidence (i.e., the observed match and the match probability of the crime stain's profile). It is combined with an unusual definition of a population of potential sources, which does not contain the true source, and a definition of the decision as a categorical consequence of a match whenever $\gamma \leq 10^{-6}$.

There are many points to discuss regarding the arguments advanced by Fimmers et al. (2011). This paper treats the following three aspects:

A) the decision of 'individualizing' an individual as the source of a crime stain having a match probability of $\gamma = 10^{-6}$ in a population of 100 million potential souces ($N = 10^8$)

[4]Translation by the current authors.

[5]Translation by the current authors, and words in square brackets added by the current authors.

after obtaining a single hit with this individual in a database containing 1 million of these potential sources ($n = 10^6$);

B) the assumption that the true source of the crime stain is *not* in the population considered by the investigators; and

C) the conclusion that a probability of a false individualization is considerably greater in the database search case than in the probable cause case.

Throughout this paper, we will refer to these claims as points A, B and C.

11.2 Structure and contents of this paper

In this paper, we invoke decision theory to analyze the issue of how to decide to 'convict', or rather to 'individualize', the matching individual found in a database. The aim is to compare Fimmers et al. (2011)'s conclusions (points A, B and C) with the results obtained from a decision-theoretic approach to the database search problem. Section 11.3 will present a decision-theoretic approach to the database search problem, using the visual representation of an influence diagram to clarify the relationships between the set of target decisions and the variables that pertain to the database search scenario formulated in its traditional version (which focused solely on probabilistic inference, rather than on decision making). This section uses the influence diagram to examine points A and B. We study the normative decision-theoretic framework in further detail in Section 11.4, in order to determine in which situations a rational decision maker would individualize the selected person. This section discusses point A in further detail. Section 11.5 examines the probability of a false individualization for a rational decision maker, and studies how this probability is regulated by the decision maker's personal objectives and preferences. It concludes by comparing our results with the assertion made in point C. Section 11.6 summarizes the insights provided by this study, and discusses them with regard to Fimmers et al. (2011)'s argument.

Note that the decision of 'convicting' or 'individualizing' is made on the basis of the probability distribution over the propositions, and must therefore be on the same hierarchical level as the propositions in the hierarchy of propositions (Cook et al., 1998b). Fimmers et al. (2011)'s hypothetical case addresses the decision at the crime level in this hierarchy, speaking of 'convicting' as the decision. In this paper, we prefer staying at the source level (the decision of 'individualizing'), since all of our propositions (see Section 11.1) are on this level. Note that, mathematically, the crime level analysis is equivalent to the source level analysis if one assumes maximal relevance for the crime stain (i.e., if one assumes that the crime stain certainly comes from the perpetrator) (e.g., Aitken and Taroni, 2004).

11.3 A decision-theoretic approach to the database search problem

11.3.1 Preliminaries

Decision theory has provided a logical framework for solving several forensic decision problems (Taroni et al., 2005, 2007; Biedermann et al., 2008a; Taroni et al., 2010). Here, we are interested in the process of 'individualization', that is, the attribution of a trace to a single source to the exclusion of all other potential sources. Notably, the act of 'individualizing', or 'not individualizing', a person or an object can be conceptualized as a decision made on the basis of the inferences resulting from the probabilistic evaluation of evidence (Biedermann et al., 2008a). This idea is also applicable to a situation in which an individual is selected through a database search. In this section, we present the database search problem from

such a decision-theoretic point of view. Section 11.3.2 defines the ingredients for a decision-theoretic approach, and Section 11.3.3 translates these concepts into an influence diagram to clarify the logical relationships between the decision of individualizing and the variables in the database search problem.

11.3.2 Elements of decision theory

Decision theory provides a normative model for making rational decisions under uncertainty.[6] It provides a mechanism for combining a measurement of the uncertainty relating to the target propositions that are an integral part of the decision problem, with a measurement of the decision maker's preferences concerning the possible outcomes of the decision (e.g., von Neumann and Morgenstern, 1947; Savage, 1951; Pratt et al., 1964). Its application follows a set of well defined rules invoking the theoretical notions defined and explained below.

Actions

A decision consists of choosing one action from a set of possible actions. For this, the initial set of possible actions must be an exhaustive list of mutually exclusive actions from which the decision maker will choose exactly one.

The decision that interests us here is that of formally individualizing an individual i, $i \in \{1, \ldots, N\}$, as the source of the trace recovered on the crime scene. For the sake of illustration, let us call individual i Mr. Smith. With regard to Mr. Smith, an investigator (e.g., an investigating magistrate, hereafter referred to as the decision maker) has an action space consisting of:

a_i: individualize Mr. Smith as the source of the trace recovered on the crime scene;
$\neg a_i$: not individualize Mr. Smith as the source of the trace recovered on the crime scene.

The difficulty in choosing one of these two actions resides in the uncertainty surrounding the true state of nature (defined below). That is, the outcome of the decision does not depend exclusively on the action chosen, but also upon the true state of one or several variables, on which the decision maker has no influence.

States of nature

In decision theory, the variables that affect the degree of satisfaction the decision maker obtains from choosing a particular action are called the states of nature (or the states of the world). Their true state is unknown to the decision maker at the time of making the decision, and may remain unknown to the decision maker even after having made the decision.

In the case of individualizing a suspect, the satisfaction obtained from the outcome depends on whether the trace truly comes from that individual: that is, a correct individualization is desirable, whereas a wrong individualization is undesirable (Biedermann et al., 2008a). For actions a_i and $\neg a_i$, the states of nature are therefore the pair of propositions:

H_i: the crime stain comes from Mr. Smith;
$\neg H_i$: the crime stain comes from someone else unrelated to Mr. Smith.

Since the actual source of the crime stain is never known with certainty (e.g., Kaplan, 1968), the decision maker requires a probability distribution over the possible states. This probability distribution will depend on the information available to the decision maker at the time of making the decision. In this paper, we consider the decision after having searched the database and found a single match with Mr. Smith. Therefore we consider

[6] Rational decision making is understood here as a coherent decision making process that obeys the laws of probability (Lindley, 1985). We use the term 'rational' as a synonym of 'coherent' meaning the opposite of 'incoherent', in the sense that 'incoherent' decisions may lead to losses such as Dutch books (e.g., Lindley, 1985; Press, 2003).

Table 11.1: *Combining the possible actions, a_i (individualize Mr. Smith) and $\neg a_i$ (not individualize Mr. Smith) with the possible states of nature, H_i (the crime stain comes from Mr. Smith) and $\neg H_i$ (the crime stain does not come from Mr. Smith) forms four possible consequences, denoted C_k, $k = 1, \ldots, 4$. To describe the desirability of each of these consequences, we assign a loss value to each one: a minimal loss of 0 for a correct conclusion, and losses of λ and 1 for incorrect conclusions, such that the loss of a miss (incorrectly not individualizing Mr. Smith) is a fraction λ of the loss of a false individualization.*

	a_i		$\neg a_i$	
	H_i	$\neg H_i$	H_i	$\neg H_i$
consequence:	C_1	C_2	C_3	C_4
loss:	$l(C_1) = 0$	$l(C_2) = 1$	$l(C_3) = \lambda$	$l(C_4) = 0$

a situation where Mr. Smith is an individual registered in the database, that is, where $i \in \{1, \ldots, n\}$. We denote the probabilities of H_i and $\neg H_i$ posterior to the database search as $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and $Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, where M_i, represents a match between the crime stain's DNA profile and Mr. Smith's DNA profile, and X_j, $j = 1, \ldots, n$ and $j \neq i$, a non-match between the crime stain's DNA profile and the DNA profile of individual j.

Consequences

The combination of the states of nature with the possible actions produces the possible outcomes, or consequences, of the decision. In this case, the combination of a_i and $\neg a_i$ with H_i and $\neg H_i$ produces four possible consequences (Table 11.1). We denote these C_k, $k = 1, \ldots, 4$:

C_1: correctly individualize Mr. Smith as the source of the crime stain;
C_2: incorrectly individualize Mr. Smith as the source of the crime stain;
C_3: incorrectly not individualize Mr. Smith as the source of the crime stain;
C_4: correctly not individualize Mr. Smith as the source of the crime stain.

Of these four consequences, the decision maker considers the correct conclusions C_1 and C_4 as desirable outcomes, and the incorrect conclusions C_2 and C_3 as undesirable outcomes. To represent these preferences, the decision maker must quantify the 'satisfaction' obtained from each consequence. Several scales exist for this quantification. What is important is not so much which scale (because one can translate the values from one scale into another scale), but how the different values assigned to the possible consequences relate to each other. Here, we have chosen *losses* with a scale ranging from 0 to 1, because (i) this scale allows us to introduce 0's for the most desirable outcomes, which reduces the complexity of the equations that follow, and (ii) it facilitates the quantification of the desirability or undesirability of intermediate consequences (see Appendix).

Losses

A loss, denoted $l(C_k)$, $k = 1, \ldots, 4$, is defined as a non-negative number, such that the larger this value, the less desirable the consequence it describes (e.g., Lindley, 1985; Press, 2003). Here, we use a scale ranging from 0, for the best consequence, to 1, for the worst consequence.[7] The best consequence therefore has a minimal loss value of 0. Here, we consider the best consequence as that of reaching a correct conclusion, and assign loss values of 0 to consequences C_1 and C_4:

$$l(C_1) = 0 \qquad \text{and} \qquad l(C_4) = 0 \ .$$

[7] For a more detailed and technical explanation of this scale, the loss function and how these relate to utilities, we refer the interested reader to the Appendix.

For the incorrect conclusions, the loss values $l(C_2)$ and $l(C_3)$ reflect the extent of the loss of one false conclusion with regard to the loss of the other false conclusion (Taroni et al., 2010, pp. 337-338). The legal principle that it is better to acquit a guilty offender than to convict an innocent person (e.g., Blackstone, 1765-1769; Hale, 1847; Fortescue, 1917)[8] advocates associating a greater loss with a false individualization (C_2) than with the miss of an individualization (i.e., an incorrect non-individualization, C_3):[9]

$$l(C_2) > l(C_3) \ . \tag{11.1}$$

We therefore assign the maximal loss value of 1 to a false individualization:

$$l(C_2) = 1 \ ,$$

and use the parameter λ to designate the loss of a missed individualization:

$$l(C_3) = \lambda \ .$$

Generally, λ is defined as $0 \leq \lambda \leq 1$, yet if Eq. (11.1) is true and with C_1 and C_4 being the most favorable outcomes, we have $0 < \lambda < 1$. Table 11.1 presents a summary of the losses associated with each of the four possible consequences.

The loss of a miss is a fraction λ of the loss of an incorrect individualization. The exact value for λ will vary from one decision maker to another, because it reflects that decision maker's personal preferences among the possible consequences of the decision. For example, someone who believes that 1 false individualization is as undesirable as 10 missed individualizations will act according to $\lambda = \frac{1}{10}$, and somebody who believes 1 false individualization to be as undesirable as 100 missed individualizations will act according to $\lambda = \frac{1}{100}$. For a more technical explanation of this loss function, see the Appendix.

Minimizing the expected loss

The most rational action is the one that minimizes the expected loss (e.g., Lindley, 1985; Press, 2003). The rational decision maker will therefore choose the action with the minimum expected loss. The expected loss of each action, denoted $\bar{l}(\cdot)$, is the sum of the losses of the possible consequences of that action, weighted by the probabilities of these consequences actually occurring:

$$\begin{aligned} \bar{l}(a_i) &= l(C_1)Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\ &\quad + l(C_2)Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\ &= Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \ , \end{aligned} \tag{11.2}$$

and

$$\begin{aligned} \bar{l}(\neg a_i) &= l(C_3)Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\ &\quad + l(C_4)Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\ &= \lambda \times Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \ . \end{aligned} \tag{11.3}$$

Given that

$$Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 1 - Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \ ,$$

the expected losses are both linear functions of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. The expected loss for a_i is a decreasing function of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, with a constant slope of -1. It does not depend on any other variables. The expected loss for $\neg a_i$

[8] See for example Kaplan (1968).

[9] Here, we have extended the legal principle addressing the crime level to the issue treated at the source level (see Cook et al. (1998b) for an explanation on source and crime levels in the hierarchy of propositions).

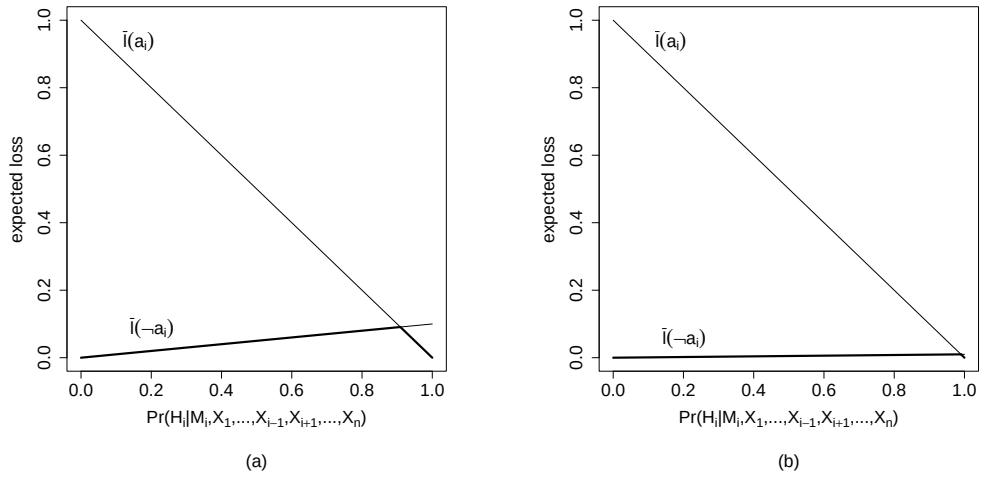


Figure 11.1: *The expected losses of individualizing, a_i, and not individualizing, $\neg a_i$, are linear functions of $Pr(H_i|M_i, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. The rational decision maker chooses the action with the smallest expected loss, indicated here with the bold lines. The point of intersection marking the transition point between choosing $\neg a_i$ and choosing a_i is a function of λ (i.e., the loss associated with missing an individualization). This point corresponds to $\frac{1}{1+\lambda}$.* **(a)** *For $\lambda = \frac{1}{10}$, this transition point lies at $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.9090909$.* **(b)** *For $\lambda = \frac{1}{100}$, this transition is at $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.9900990$.*

is an increasing function of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, with a slope equal to λ. It will therefore vary according to the numerical value specified for λ. Fig. 11.1(a) plots the two expected losses for $\lambda = \frac{1}{10}$, and Fig. 11.1(b) plots them for $\lambda = \frac{1}{100}$. The rational decision maker will always choose the action with the smallest expected loss (indicated by the bold lines in Fig. 11.1). The rational decision maker will therefore individualize the suspect if and only if

$$\bar{l}(a_i) < \bar{l}(\neg a_i) \ ,$$

in other words when

$$\frac{Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}{Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} > \frac{1}{\lambda} \ . \tag{11.4}$$

This corresponds to intuition, which says that an individualization may only be justified when the odds that the crime stain comes from this suspect are high. According to this approach, there is a threshold value for these odds equal to $\frac{1}{\lambda}$, above which the decision maker should individualize the suspect.

This is equivalent to saying, in terms of the probability that the crime stain comes from the selected individual $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, that an individualization is rational for

$$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) > \frac{1}{1+\lambda} \ . \tag{11.5}$$

The expression $\frac{1}{1+\lambda}$ designates the transition point at which the decision maker is indifferent between a_i and $\neg a_i$. This point is the point of intersection of the two functions in each of the graphs in Fig. 11.1 (i.e., 0.9090909 for $\lambda = \frac{1}{10}$ in Fig. 11.1(a), and 0.9900990 for $\lambda = \frac{1}{100}$ in Fig. 11.1(b)). Again in agreement with intuition, these graphs show that the smaller λ, the smaller the range of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ for which the decision maker chooses a_i. In other words, a decision maker who regards 1 false individualization as being as undesirable as 100 missed individualizations ($\lambda = \frac{1}{100}$) requires a higher

$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ to justify an individualization than a decision maker who valuates 1 false individualization as being as undesirable as only 10 missed individualizations ($\lambda = \frac{1}{10}$).

We discuss the meaning and implications of this threshold in further detail in Section 11.4. The next section explains how the decision of individualizing a suspect fits into the entire context of the database search problem.

11.3.3 Influence diagrams

The authors in Biedermann et al. (2011a) presented a Bayesian network modeling the decision as a categorical consequence according to the reasoning described in Fimmers et al.'s example. We can reformulate this model, made up entirely of probabilistic nodes, by introducing elements of decision theory. A decision theoretic approach consists of constructing an influence diagram.

Influence diagrams, or Bayesian decision networks, extend probabilistic models, such as Bayesian networks, to incorporate loss functions and action nodes (e.g., Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008). These allow the user to compute expected losses for the possible actions, in addition to updating probability distributions over the unknown variables. Representing the database search problem as an influence diagram allows us to clarify the relationships existing between the variables in this problem and the decision of individualizing a suspect.

The aim of this section is to illustrate how each of the two pairs of propositions presented in the introduction (see Section 11.1) relates to the decision of individualizing a source. For this, we consider pair $\{H_i, \neg H_i\}$, concerning whether the crime stain comes from Mr. Smith (i.e., individual i), in a node labeled H_i, and pair $\{H_{db}, \neg H_{db}\}$, concerning whether the source of the crime stain is in the database, in a node labeled H_{db}. This database contains the profiles of individuals $1, \ldots, n$. Since we defined $i \in \{1, \ldots, n\}$, the profile of Mr. Smith is registered in the database.

Fig. 11.2 presents two possible influence diagrams for the database search problem. In addition to nodes H_i and H_{db}, these models contain:

- Node N: This numerical node allows the user to specify the size of the population of potential sources considered by the investigators. Here, we set $N = 10^8$.
- Node n: This numerical node indicates the size of the database representing the population of interest. We set $n = 10^6$.
- Node γ: This numerical node gives the match probability of the crime stain's profile in the population of interest. We set $\gamma = 10^{-6}$.
- Node S *in* N: This is a Boolean node that indicates whether the source of the crime stain is actually in the population of size N considered by the investigators as the population of potential sources.
- Node M_i: This node contains state M_i, describing a match between the crime stain's profile and Mr. Smith's profile, and state $\neg M_i$, for the absence of such a match. The probability of the match depends on whether the crime stain comes from Mr. Smith. If it comes from Mr. Smith, we assume that no laboratory errors are possible such that a match is certain. If it comes from someone else, the probability of observing a match with the profile of Mr. Smith is given by the profile's match probability, γ, in the population of potential donors. The probability table for this node is:

	H_i:	H_i	$\neg H_i$
M_i:	$Pr(M_i)$	1	γ
	$Pr(\neg M_i)$	0	$1 - \gamma$

- Node $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$: This node contains the state $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ to describe the event of observing a non-match for individuals $1, \ldots, i-1, i+1, \ldots, n$ in the database. Its complement, $\neg\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$, describes all of the other possible results of a comparison of these registered profiles with the crime stain's profile. The probability of obtaining exactly $n-1$ non-matches in a database search follows a binomial distribution with $n-1$ trials and a success probability of γ. Assuming that no laboratory errors occurred and that there's a match between the crime stain's profile and Mr. Smith's profile, there are only two situations in which it is possible to obtain $n-1$ non-matches: (i) when the crime stain comes from Mr. Smith, and (ii) when Mr. Smith is not the donor of the crime stain *and* the true donor is not in the database. For, if the true donor is in the database and this donor is not Mr. Smith, one would observe $n-2$ non-matches, since there would have to be an additional match with this true donor. Therefore, the conditional probabilities associated with $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ depend on both H_i and H_{db}, as shown in the following probability table:

H_i:	H_i		$\neg H_i$	
H_{db}:	H_{db}	$\neg H_{db}$	H_{db}	$\neg H_{db}$
$X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$:				
$Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$	$(1-\gamma)^{n-1}$	n/a	0	$(1-\gamma)^{n-1}$
$Pr(\neg\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\})$	$1-(1-\gamma)^{n-1}$	n/a	1	$1-(1-\gamma)^{n-1}$

The 'n/a' in the second column of the table stands for 'not applicable'. The combination of the states H_i and $\neg H_{db}$ is in fact impossible because Mr. Smith *is* in the database. In other words, if the crime stain comes from Mr. Smith (proposition H_i), then the source of the crime stain must be in the database (proposition H_{db}), since Mr. Smith is in the database.

- Node A: This is the action node. It contains a_i, the action of individualizing Mr. Smith, and $\neg a_i$, the action of not individualizing Mr. Smith.

- Node L: This node contains the losses defined for each of the possible consequences (see Table 11.1).

There are two possibilities for constructing an influence diagram containing these nodes (Fig. 11.2(a) and 11.2(b)). The difference between the two lies in the direction of the arrow between nodes H_i and H_{db}. This dependence relationship may be modeled in both directions, and produces the same result:

- $H_i \rightarrow H_{db}$ (Fig. 11.2(a)): In this case, we first consider whether Mr. Smith is at the source of the trace, and then, in a second step, specify for proposition $\neg H_i$ whether the 'someone else' is in the database. It is clear that if the source of the crime stain is not in the population of size N considered by the investigators (i.e., S *in* N = *false*), then the marginal probabilities of propositions H_i and H_{db} are both equal to 0 (see Biedermann et al. (2011a)). If, however, the source of the crime stain is in the population considered (i.e., S *in* N = *true*), then we relate the probabilities for H_i and H_{db} to the size of this population of potential sources and the size of the database as follows:

	S *in* N:	*true*
H_i:	$Pr(H_i)$	$\frac{1}{N}$
	$Pr(\neg H_i)$	$1-\frac{1}{N}$

	S *in* N:	*true*	
	H_i:	H_i	$\neg H_i$
H_{db}:	$Pr(H_{db})$	1	$\frac{n-1}{N-1}$
	$Pr(\neg H_{db})$	0	$\frac{N-n}{N-1}$

For a situation in which S *in* N = *true*, the marginal probability of proposition H_i is thus one over the total number of individuals in the population of interest, and

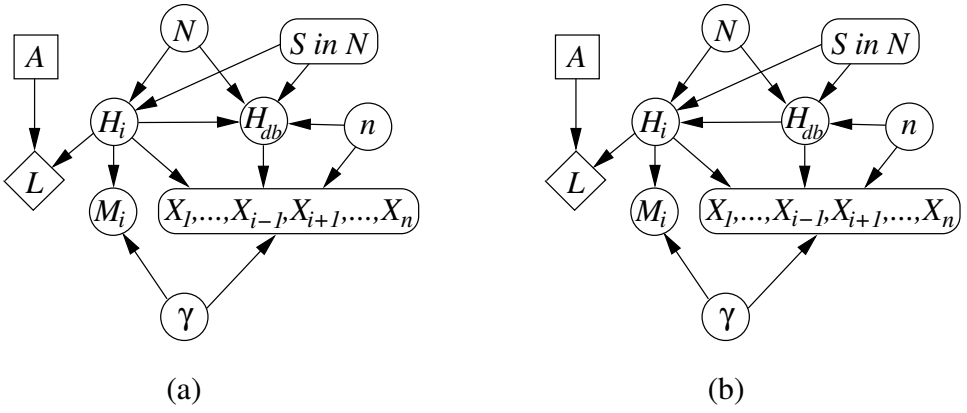


Figure 11.2: *Two influence diagrams representing the decision of individualizing a suspect found as the only match in a database search. The square node A contains the possible actions, in this case, 'individualizing' or 'not individualizing' the suspect as the source of the crime stain, and the diamond-shaped node L the loss values defined by the decision maker for each of the possible consequences. The rest of the nodes are variables. All of the probabilities in these nodes are determined by three variables: the size of the population considered N, the size of the database representative of this population n, and the match probability γ of the crime stain's profile in the population of interest. Node S in N allows the user to introduce uncertainty on whether the crime stain's source is in the population considered. Node M_i describes the event of finding a match between the crime stain's profile and the profile of the suspect Mr. Smith (denoted here as individual i), and node $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ the event of not finding a match with the crime stain's profile among the $n-1$ other profiles in the database (denoted here as individuals $1, \ldots, i-1, i+1, \ldots, n$). Nodes H_i and H_{db} differentiate between two pairs of propositions: node H_i contains the pair of propositions on whether Mr. Smith (individual i) is the source of the crime stain, and node H_{db} the pair of propositions on whether the crime stain's source is in the database. Note that the dependance relationship between these two pairs of propositions may be modeled in either direction, producing two different structures for this influence diagram:* **(a)** $H_i \rightarrow H_{db}$, *and* **(b)** $H_{db} \rightarrow H_i$. *The two influence diagrams produce the exact same results. The text in section 11.3.3 describes the nodes in further detail, and presents the probability tables.*

the marginal probability of proposition H_{db} reduces to the ratio of the number of individuals in the database to the number of individuals in the population of interest:

$$\begin{aligned} Pr(H_{db}|S\ in\ N = true) &= Pr(H_{db}|H_i, S\ in\ N = true) \times Pr(H_i|S\ in\ N = true) \\ &+ Pr(H_{db}|\neg H_i, S\ in\ N = true) \times Pr(\neg H_i|S\ in\ N = true) \\ &= 1 \times \frac{1}{N} + \frac{n-1}{N-1} \times \frac{N-1}{N} \\ &= \frac{n}{N}\ . \end{aligned} \tag{11.6}$$

This ratio is equal to the probability assigned to $Pr(H_{db}|S\ in\ N = true)$ in the probability table of node H_{db} in the influence diagram of Fig. 11.2(b) (see below). It shows that the marginal probability of proposition H_{db} given $S\ in\ N = true$ is identical in the two influence diagrams.

This approach of conditioning the probability distribution over the pair of propositions $\{H_{db}, \neg H_{db}\}$ on the pair of propositions $\{H_i, \neg H_i\}$ corresponds to Kaye's reasoning (Kaye, 2009): the probability of the donor not being in the database given the trace does not come from Mr. Smith, i.e. $\frac{N-n}{N-1}$ in the table above, is equal to Kaye's θ (figuring in the appendix of Kaye (2009)).

- $H_{db} \rightarrow H_i$ (Fig. 11.2(b)): Here, we first specify whether the donor of the crime stain is in the database, and then, in the second step, consider whether this donor is Mr. Smith. Again, the marginal probabilities of propositions H_i and H_{db} are equal to 0 when the source of the crime stain is not in the population considered ($S\ in\ N = false$) (Biedermann et al., 2011a). If the source of the crime stain is in the population ($S\ in\ N = true$), then generic probability assignments to this population of potential sources produces the following probability tables:

	$S\ in\ N$:	$true$	
	H_{db}:	H_{db}	$\neg H_{db}$
H_i:	$Pr(H_i)$	$\frac{1}{n}$	0
	$Pr(\neg H_i)$	$1 - \frac{1}{n}$	1

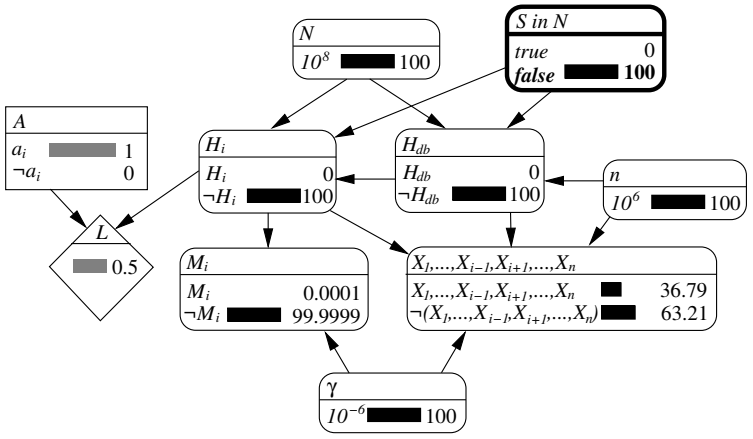
	$S\ in\ N$:	$true$
H_{db}:	$Pr(H_{db})$	$\frac{n}{N}$
	$Pr(\neg H_{db})$	$1 - \frac{n}{N}$

For a situation in which $S\ in\ N = true$, the marginal probability of proposition H_{db} is the number of individuals in the database divided by the number of individuals in the population of interest, and the marginal probability of H_i reduces to one over the size of the population of interest:

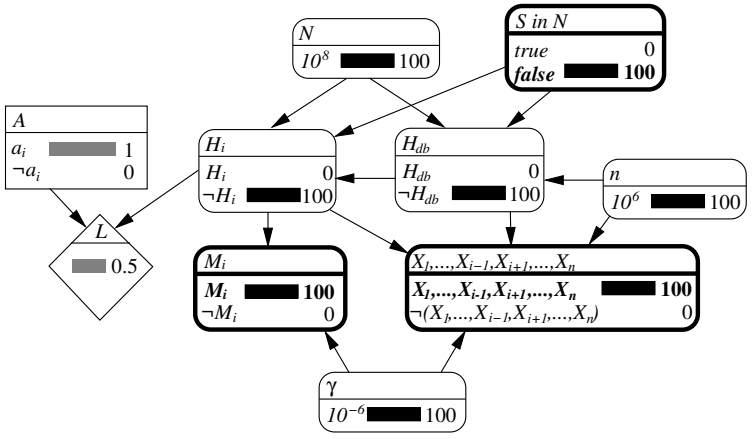
$$\begin{aligned} Pr(H_i|S\ in\ N = true) &= Pr(H_i|H_{db}, S\ in\ N = true) \times Pr(H_{db}|S\ in\ N = true) \\ &+ Pr(H_i|\neg H_{db}, S\ in\ N = true) \times Pr(\neg H_{db}|S\ in\ N = true) \\ &= \frac{1}{n} \times \frac{n}{N} + 0 \times \frac{N-n}{N} \\ &= \frac{1}{N}\ . \end{aligned} \tag{11.7}$$

This ratio is equal to the probability assigned to $Pr(H_i|S\ in\ N = true)$ in the probability table of node H_i in the influence diagram of Fig. 11.2(a) (see above). It shows that the marginal probability of proposition H_i given $S\ in\ N = true$ is identical in the two influence diagrams.

Hence, the approaches in Fig. 11.2(a) and 11.2(b) produce identical results, so that either of them may be used. For the rest of this paper, we have chosen to use the influence diagram in Fig. 11.2(b).

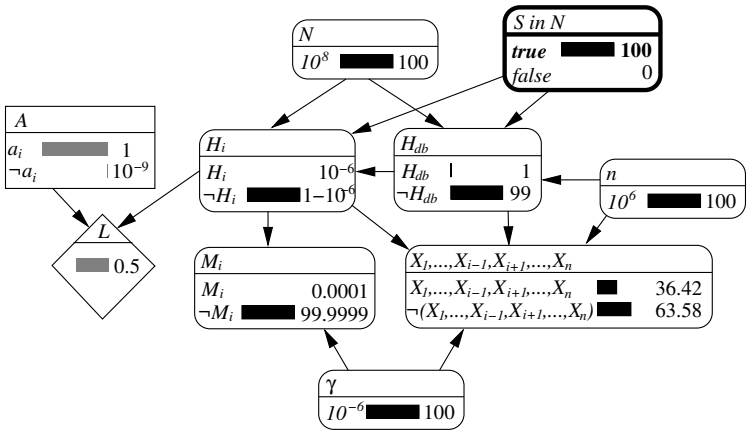


(a)

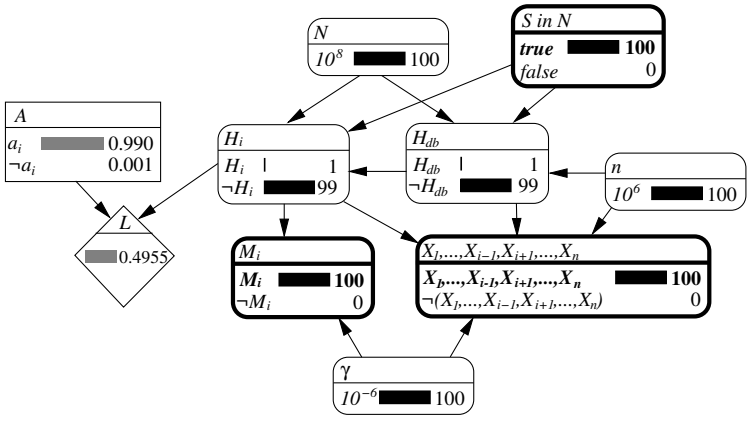


(b)

Figure 11.3: *Expanded representation of the influence diagram presented in Fig. 11.2(b) for Fimmers et al.'s scenario, in which it is assumed that the crime stain's source is not in the population of size N considered by the investigators (i.e., S in N = false). Node A indicates the expected losses on a scale from 0 (the minimum loss) to 1 (the maximum loss), and node L the arithmetic mean of the expected losses of all of the possible actions. The losses are defined in Table 11.1, here with* $\lambda = \frac{1}{10}$. *The rest of the nodes indicate probabilities in %. The nodes with bold borders are the nodes instantiated by the user, i.e., containing states whose probabilities have been set to 1.* **(a)** *The most rational action before searching the database is not to individualize individual i.* **(b)** *After finding a single match in the database with individual i, the most rational action is still to 'not individualize' this person.*



(a)



(b)

Figure 11.4: *Expanded representation of the influence diagram presented in Fig. 11.2(b) for the scenario where the crime stain's source is in the population of size N considered by the investigators (i.e., S in N = true). Node A indicates the expected losses on a scale from 0 (the minimum loss) to 1 (the maximum loss), and node L the arithmetic mean of the expected losses of all of the possible actions. The losses are defined in Table 11.1, here with* $\lambda = \frac{1}{10}$*. The rest of the nodes indicate probabilities in %. The nodes with bold borders are the nodes instantiated by the user, i.e., containing states whose probabilities have been set to 1.* **(a)** *The most rational action before searching the database is not to individualize individual i.* **(b)** *The result of a single match with individual i after searching the database has decreased the expected loss of individualizing this person and increased the expected loss of not individualizing him/her, yet the most rational action remains 'not to individualize'.*

11.3.4 Example

Fig. 11.3 shows the results produced by the influence diagram described in Section 11.3.3 (i.e., Fig. 11.2(b)) for the case presented by Fimmers et al. (2011). Here, we investigate the scenario using the numerical values summarized in point A, with the assumption given in point B, which prescribes that S *in* N = *false*. We oppose this case to a situation in which S *in* N = *true* (Fig. 11.4), in other words, where we relax the assumption of point B, and use only the numerical values of point A. For the purpose of illustration, we set $\lambda = \frac{1}{10}$. Node A gives the expected losses for 'individualizing' and 'not individualizing' Mr. Smith. We observe that:

- For S *in* N = *false* (Fig. 11.3), the result of the database search has no impact on the decision. The expected loss of 'not individualizing' Mr. Smith has the minimal value of 0 before searching the database, and remains 0 after this search. In this case, the rational decision maker will definitely not individualize Mr. Smith. As an alternative representation, Fig. 11.5(a) gives the decision tree for this decision process.

- For S *in* N = *true* (Fig. 11.4), the expected loss for 'individualizing' decreases from 1 to 0.99, and the expected loss for 'not individualizing' increases from 10^{-9} to 10^{-3}. This change is due to the evolution of the probability distribution over propositions H_i and $\neg H_i$, which determines the expected losses according to Eq. (11.2) and (11.3). Notice how the influence diagram clearly demonstrates that this change in expected losses cannot come from the probability distribution over propositions H_{db} and $\neg H_{db}$ since, in this particular case, this probability distribution is the same before and after observing the single match in the database.[10] According to the updated expected losses, however, the most rational decision is still 'not to individualize' Mr. Smith, even after this individual produced the one and only match in the database. This decision process is shown as a decision tree in Fig. 11.5(b).

These observations lead to the following conclusions:

- Between the two pairs of propositions $\{H_i, \neg H_i\}$ and $\{H_{db}, \neg H_{db}\}$, it is pair $\{H_i, \neg H_i\}$ which has a direct impact on the decision of individualizing Mr. Smith. Pair H_{db} only has an impact when it changes the probability distribution in node H_i. That is, given the probability distribution over pair $\{H_i, \neg H_i\}$, the rational decision maker does not need to know the probability distribution over pair $\{H_{db}, \neg H_{db}\}$ to choose between 'individualizing' and 'not individualizing' Mr. Smith. This reflects the states of nature in this decision problem, which are $\{H_i, \neg H_i\}$, not $\{H_{db}, \neg H_{db}\}$.

- Clearly, it is unreasonable to individualize the suspect in the scenario described by the numerical values in point A. In the situation presented by Fig. 11.3 (which includes point B), we know that the suspect is not the source. In the situation presented by Fig. 11.4 (without point B), the suspect only has a probability of 0.01 of being the source. In either case, it is unsafe to 'individualize' the selected person.

In view of these analyses, it thus appears that Fimmers et al. (2011) should not proceed with an individualization in their hypothetical scenario. The question that now remains is in what situations one could individualize a single matching individual found through a database search. We explore this question in the next section.

[10] This is a coincidence due to the numerical values of $N = 10^8$, $n = 10^6$ and $\gamma = 10^{-6}$. Before the database search, the marginal probability of H_{db} is equal to $\frac{n}{N}$, and after finding a single hit in the database, this probability is $\frac{1}{1+\gamma(N-n)}$ (Balding and Donnelly, 1995). With $N = 10^8$, $n = 10^6$ and $\gamma = 10^{-6}$, this gives us $\frac{n}{N} = \frac{10^6}{10^8} = \frac{1}{100}$ and $\frac{1}{1+\gamma(N-n)} = \frac{1}{1+10^{-6}(10^8-10^6)} = \frac{1}{100}$. Notice also how the influence diagram presents the concept of post-data equivalence between the two pairs of propositions. That is, the database search updates the marginal probability of H_i from $\frac{1}{N}$ to $\frac{1}{1+\gamma(N-n)}$ after finding a single match in the database. This is the same posterior probability as for H_{db}, because the exclusion of all the other individuals in the database has made the propositions logically equivalent (Dawid, 2001).

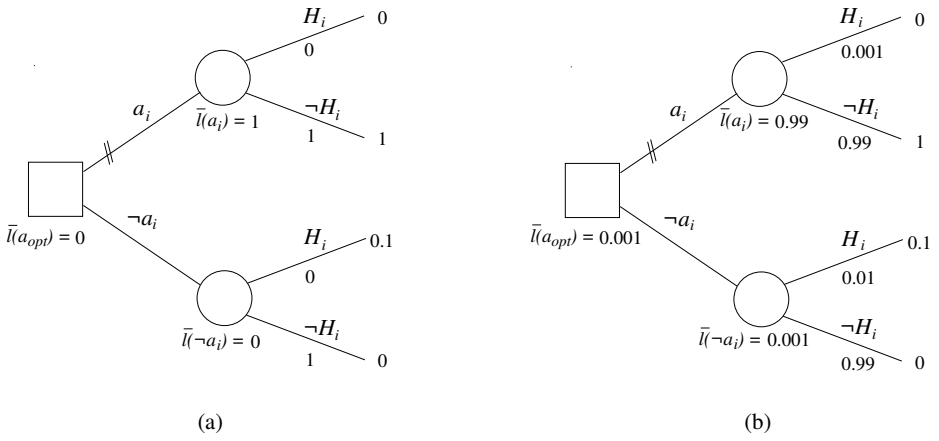


Figure 11.5: *The decision tree for individualizing or not individualizing Mr. Smith, with* $\lambda = \frac{1}{10}$. *The branches branching off of a square indicate the possible actions, and the branches branching off of a circle the states of nature. The numbers below the latter indicate the probability distribution over the states of nature. The values at the end of each branch, on the right-hand side, indicate the loss of the consequence obtained by following the branches from left to right to that end of the tree. The expected losses for each action are given at the end of each action branch, below each circle. The rational decision maker chooses the action which minimizes the expected loss, denoted here as the optimal action* $a_{opt} \in \{a_i, \neg a_i\}$. *This action's expected loss is indicated below the square. The double lines bar the branch of the irrational action (i.e., the action that does not minimize the expected loss).* **(a)** *The tree for* S *in* $N =$ *false, corresponding to the situation modeled by the influence diagram in Fig. 11.3.* **(b)** *The tree for* S *in* $N =$ *true, corresponding to the situation modeled by the influence diagram in Fig. 11.4.*

11.4 Decision-theoretic analysis of individualizing a suspect found through a database search

11.4.1 Preliminaries

According to Figs. 11.3 and 11.4 where the loss function is specified for $\lambda = \frac{1}{10}$, it is not rational to individualize a suspect found through a database search given the numerical values presented in Fimmers et al.'s hypothetical case (i.e., $N = 10^8$, $n = 10^6$ and $\gamma = 10^{-6}$ as given in point A). If the assumption in point B holds (i.e., S *in* $N =$ *false*), it is impossible for the crime stain to come from Mr. Smith, because the crime stain does not come from someone in the population considered by the investigators, and Mr. Smith is in this population considered by the investigators. Therefore the assumption in point B produces

$$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0 \ ,$$

such that Eq. (11.2) and (11.3) are equal to

$$\bar{l}(a_i) = 1 \qquad \text{and} \qquad \bar{l}(\neg a_1) = 0 \ ,$$

respectively. In this case, the most rational action is always $\neg a_i$, regardless of the value of λ. If we relax the assumption made in point B such that the crime stain's source *is* in the population considered by the investigators (an assumption we shall maintain throughout the rest of this paper to make a decision-theoretic analysis of the problem possible), the decision of individualizing Mr. Smith rests on two values (Eq. (11.5)):

- $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, the probability that Mr. Smith is the source of the crime stain, and
- λ, the relative loss of missing an individualization with regard to the loss of falsely individualizing a source.

In this section, we analyze the decision of individualizing Mr. Smith with regard to each of these parameters. Section 11.4.2 examines the impact of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ on this decision, and Section 11.4.3 the impact of λ.

11.4.2 Impact of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ on the decision of individualizing

Consider a decision maker for whom $\lambda = \frac{1}{10}$ (i.e., this decision maker considers 1 false individualization to be as undesirable as 10 missed individualizations). For a decision maker with a predefined value for λ, the decision of individualizing will depend entirely on the posterior probability of the person in question being the source of the trace. According to Eq. (11.5), the decision maker will individualize a person whenever

$$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) > 0.9090909 \ .$$

Assuming that each member of the population of potential sources initially has the same probability of being at the source of the crime stain, a single hit in the database with Mr. Smith produces a posterior probability of

$$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = \frac{1}{1 + \gamma(N - n)} \tag{11.8}$$

that this individual is at the source of the crime stain (Balding and Donnelly, 1995). The posterior probability for H_i is therefore a function of the match probability of the crime stain's profile, the number of individuals in the population of potential sources, and the number of these individuals who are registered in the database searched. Note that this posterior probability is also a function of the prior probability of the selected individual being the source of the crime stain, $Pr(H_i)$.[11] Here, this analysis assigns a prior probability of $\frac{1}{N}$, so that the dependence of the posterior probability of H_i on the population size N also reflects the dependence of this posterior probability on its prior probability.

First, let us consider a crime stain profile with a match probability of $\gamma = 10^{-6}$ in the population of interest. In this case, the decision maker will only individualize the single matching individual if the search in the database has excluded enough individuals to raise the posterior probability above the threshold of 0.9090909. For $N = 10^8$, this is attained when the database size exceeds 99.8999989% of the population of interest. Therefore $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ increases as the proportion of the members of the population of interest who are registered in the database increases. For the purpose of illustration, Fig. 11.6(a) shows $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ in function of this proportion, i.e. $\frac{n}{N}$, for $\gamma = 10^{-6}$ and $N = 10^8$. In a population of 100 million ($N = 10^8$), 99.8999989% of this population corresponds to a database of size $n = 99,899,999$. This is the size necessary to individualize a single matching individual in the database search. It leaves us with 100'001 people outside the database about whom we know nothing on the

[11] Bayes' theorem, together here with the law of total probability, provides the relationship between the posterior probability $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ and the prior probability $Pr(H_i)$ (e.g., Aitken and Taroni, 2004):

$$\begin{aligned}
&Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \\
&\quad = \frac{Pr(M_i|H_i) \ \times \ Pr(H_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}{Pr(M_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} \\
&\quad = \frac{Pr(M_i|H_i) \ \times \ Pr(H_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}{Pr(M_i|H_i) \ \times \ Pr(H_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + Pr(M_i|\neg H_i) \ \times \ Pr(\neg H_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} \\
&\quad = \frac{Pr(M_i|H_i) \frac{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|H_i) \ \times \ Pr(H_i)}{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}}{Pr(M_i|H_i) \ \times \ \frac{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|H_i) \ \times \ Pr(H_i)}{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} + Pr(M_i|\neg H_i) \ \times \ \frac{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|\neg H_i) \ \times \ Pr(\neg H_i)}{Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}} \\
&\quad = \frac{Pr(M_i|H_i)Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|H_i)Pr(H_i)}{Pr(M_i|H_i)Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|H_i)Pr(H_i) + Pr(M_i|\neg H_i)Pr(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n|\neg H_i)Pr(\neg H_i)} \ ,
\end{aligned}$$

with $Pr(\neg H_i) = 1 - Pr(H_i)$.

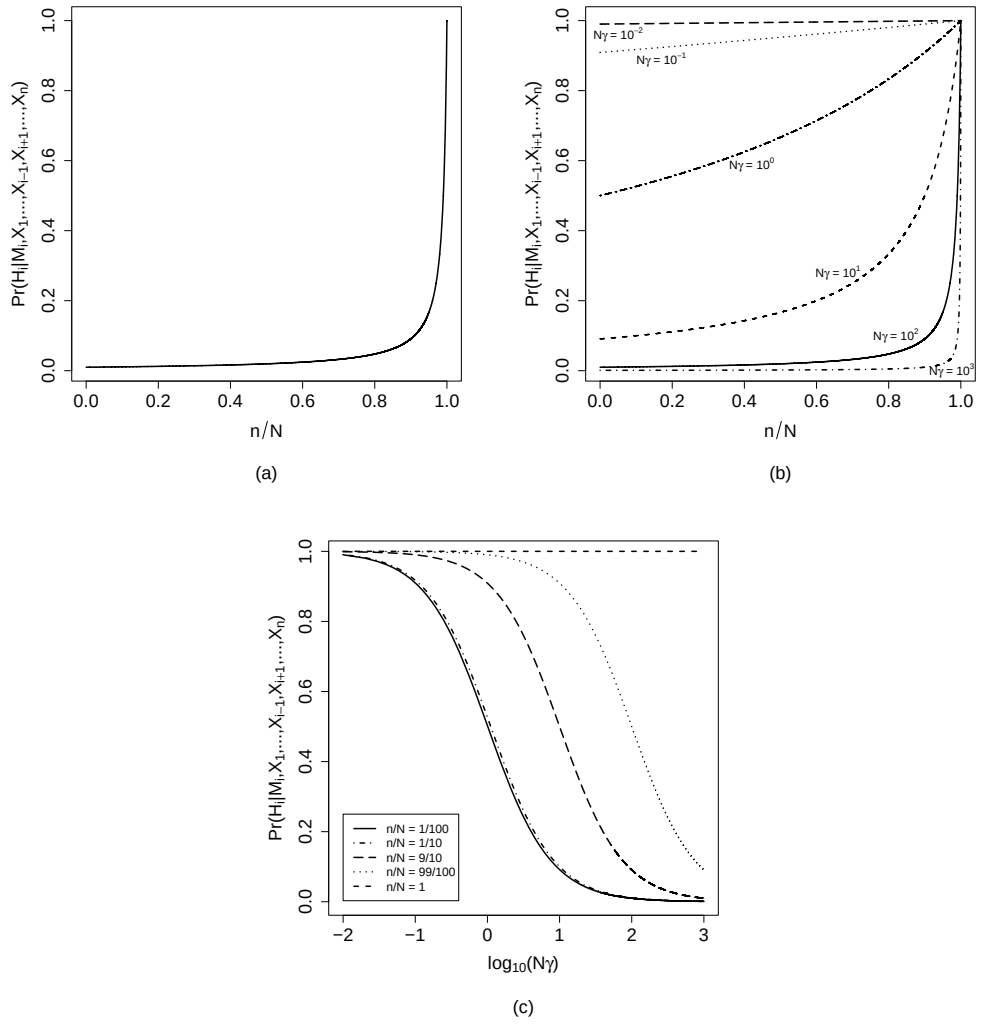


Figure 11.6: *The probability of a single matching individual in a database to be at the source of the crime stain is a function of the size of the population of potential crime stain donors N, the number of these individuals who are registered in the database n, and the match probability of the crime stain's profile γ.* **(a)** *For $\gamma = 10^{-6}$ and $N = 10^8$, the proportion of individuals in the database has a small impact on this probability, unless the database covers almost the entire population of interest.* **(b) and (c)** *In general, the individualization threshold (Eq. (11.5)) may be reached by either increasing $\frac{n}{N}$ (x-axis in (b) and the different curves in (c)) or decreasing $N\gamma$ (x-axis in (c) and the different curves in (b)).*

matching status. This is not unreasonable, because the prior probabilities are very low for each person. Thus, in this case, a rational decision maker will only individualize a single matching individual from a database search when the database contains practically the entire population of potential crime stain donors.

The above calculations for the database size are for $\gamma = 10^{-6}$. This corresponds to a trace whose match probability is 100 times greater than the inverse of the population size (i.e., $\gamma = 100 \times N^{-1}$). In other words, one would expect to obtain about 100 matches (i.e., $N\gamma$) in this population if the crime stain's profile were compared with every one of its members. If the crime stain's profile is rarer in the population considered (i.e., the match probability is smaller), the proportion of individuals from the population of interest who are registered in the database can be smaller and still justify the individualization of the matching individual. Fig. 11.6(b) shows $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ in function of $\frac{n}{N}$ for $N\gamma$ equal to 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, and 10^3. The smaller γ, the smaller the proportion $\frac{n}{N}$ required to attain a given $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. Thus, the individualization threshold can be reached by either increasing $\frac{n}{N}$, or decreasing $N\gamma$.

Second, let us concentrate on the latter point, and study the impact of γ on the probability $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. Fig. 11.6(c) illustrates how the difference in the order of magnitude between the match probability and the size of the population of potential sources (represented here by $log_{10}(N\gamma)$) affects $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$. For example, for $N = 10^8$ and $n = 10^6$ (this corresponds to a proportion of $\frac{1}{100}$ members of the population of interest registered in the database, and is depicted by the curve with the continuous line in Fig. 11.6(c)), a value of

$$\gamma < 1.010101 \times 10^{-9}$$

raises $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ above the threshold of 0.9090909. This means that for $\frac{n}{N} = \frac{1}{100}$, a rational decision maker for whom $\lambda = \frac{1}{10}$ will individualize a single matching individual in the database whenever the match probability is approximately ten times smaller than the inverse of the population size.

When the proportion of individuals registered in the database increases to much higher values, the match probability required for individualizing increases (curves with non-continuous lines in Fig. 11.6(c)). That is, an individualization may be justified for a more common profile if almost the entire population of potential sources is registered in the database. In the extreme case where the entire population is in the database, the match probability has no impact on $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, which attains certainty after obtaining only one match (this situation is depicted by the dashed, horizontal line for '$\frac{n}{N} = 1$' in Fig. 11.6(c)).

Hence, it is possible to justify an individualization of a matching source found through a database search when either n or γ attain extreme values with regard to N. This is the case when $\frac{n}{N}$ tends towards 1, and when γ becomes particularly small with regard to the population size, that is, when $N\gamma$ tends towards 0. Practically speaking, it is difficult to imagine a case satisfying either of these criteria.[12] A rational decision maker will therefore require further evidence supporting proposition H_i in order to justify an individualization.

In this section, we analyzed the decision of an individualization for a decision maker with $\lambda = \frac{1}{10}$. A decision maker with a different assessment of λ will require a different individualization threshold. In the next section, we continue investigating point A by maintaining the numerical values for N, n and γ fixed, and evaluating how large λ must be to justify an individualization in Fimmers et al.'s hypothetical case.

[12] This is even more so if one takes into account the false positive probability (which is highly recommended for an accurate approach to a real case (Thompson et al., 2003)), because this decreases the posterior probability of H_i, with an increasing effect for smaller match probabilities (see Thompson et al. (2003) for a detailed explanation on the impact of the false positive probability on the value of a DNA match).

11.4.3 Impact of λ on the decision of individualizing

The value of λ determines the threshold value of the probability of H_i marking the boundary between choosing a_i and choosing $\neg a_i$ (Eq. (11.5)). For a given $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, the rational decision maker will choose $\neg a_i$ if this probability is below the threshold value, and a_i if this probability is above this threshold.

With $N = 10^8$, $n = 10^6$ and $\gamma = 10^{-6}$ (point A), the probability of H_i after finding a single match with Mr. Smith in the database is equal to 0.01 (Eq. (11.8)). In this case, there is a probability of $1 - 0.01 = 0.99$ that Mr. Smith is *not* the source of the crime stain. To justify an individualization for such a small value of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, λ must be very large: λ must be so large that the intersection of the two expected loss functions presented in the graphs of Fig. 11.1 is to the left of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.01$.[13] To find this value for λ, we solve Eq. (11.5) for λ:

$$\lambda > \frac{1}{Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} - 1 \ .$$

This gives us the expression for calculating the value of λ for an individualization occurring for a particular probability of H_i. With $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.01$, we obtain

$$\begin{aligned} \lambda &> \frac{1}{0.01} - 1 \\ &> 99 \ . \end{aligned}$$

The individualization of a suspect in this case is only justified when λ is greater than 99. This means that the decision maker who individualizes under these conditions perceives the loss of missing an individualization as at least 99 times *greater* than the loss of a false individualization.

However, this contradicts our loss function that ranges from 0 to 1 (see the definition of the losses in Section 11.3.2). According to our loss function, $0 \leq \lambda \leq 1$ when the condition in Eq. (11.1) is relaxed, the loss associated with a missed individualization cannot be smaller than the minimum loss associated with a correct conclusion (0), nor greater than the maximum loss associated with a false individualization (1). The maximum value for λ is therefore 1. When $\lambda = 1$, this means that the loss of a missed individualization is equal to the loss of a false individualization. In this case, we obtain the expected loss functions plotted in Fig. 11.7. With both false conclusions being equally undesirable, and both correct conclusions remaining equally desirable, the threshold for individualizing the selected individual lies at $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.5$. For $0 \leq \lambda \leq 1$, this is the minimum probability for H_i that may lead the decision maker to individualize Mr. Smith. For all $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) < 0.5$, the most rational action is therefore always $\neg a_i$, that is, not to individualize Mr. Smith. This can be seen in Fig. 11.7, which shows action $\neg a_i$ as having the minimum expected loss for all probabilities ranging from 0.0 to 0.5. Hence, an individualization can never be rational for $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) = 0.01$, since $0.0 < 0.01 < 0.5$ and our loss function prescribes that action $\neg a_i$ will always minimize the expected loss for this probability of H_i.

A situation where $\lambda > 1$ would mean that the loss of missing an individualization is *greater* than the loss of a false individualization. This contradicts Eq. (11.1), where we specified that the loss of missing an individualization (incorrectly not individualizing the suspect) is *smaller* than that of a false individualization (i.e., $\lambda < 1$). This analysis reveals that the decision maker who individualizes in the above described case (with $N = 10^8$, $n = 10^6$ and $\gamma = 10^{-6}$ as described in point A) must have objectives and preferences that are in opposition with the fundamental legal concept that it is better to let a guilty person go free than to convict an innocent person. If this is the case, the decision maker will have a loss function

[13] That is, we can shift the point of intersection between the two expected loss functions towards the left in the graph by increasing the slope of the expected loss function of action $\neg a_i$. This slope is λ.

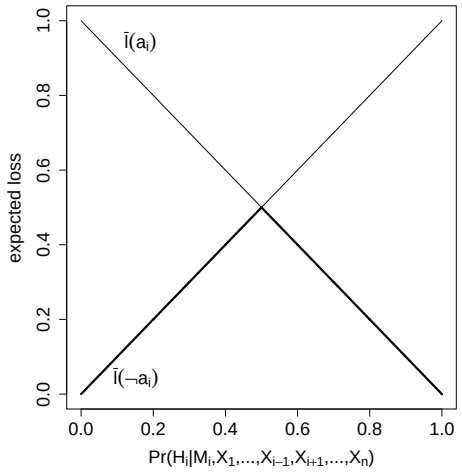


Figure 11.7: *For the maximum value $\lambda = 1$ when the condition in Eq. (11.1) is relaxed (i.e., the loss of missing an individualization is equal to the loss of a false individualization), the rational decision maker will choose not to individualize Mr. Smith, $\neg a_i$, for $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) < 0.5$, and individualize Mr. Smith, a_i, whenever $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) > 0.5$ (indicated by the bold lines). Increasing the value of λ has moved the transition point to the left (compared with $\lambda = \frac{1}{10}$ and $\lambda = \frac{1}{100}$ in Fig. 11.1), increasing the range of $Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ for which the decision maker chooses a_i.*

different from the one described in this paper, and this loss function will lead to different choices. A different loss function is therefore the only thing that will lead a rational decision maker to individualize an individual when the posterior odds for this individual being at the source of the trace are actually in favor of the alternative proposition.

In the next section, we examine the probability of a false individualization, and how this probability relates to the parameters discussed in this section. This will allow us to address point C.

11.5 Probability of a false individualization

Fig. 11.8 extends the influence diagram presented in Section 11.3.3 to include the probability of a false individualization in a node labelled C, in the same way as was done in Biedermann et al. (2011a). This node describes the event of a correct conclusion as a Boolean variable that takes the state of

$$\textit{true} \text{ for } \begin{cases} H_i \cap a_i \\ \neg H_i \cap \neg a_i \end{cases}, \text{ and} \tag{11.9}$$

false otherwise.

This influence diagram shows that, logically, the probability of a false individualization is equal to $Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ in node H_i, in a situation where the decision maker chooses action a_i.

According to Eq. (11.5), the decision maker chooses a_i when

$$Pr(H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) > \frac{1}{\lambda + 1} \ .$$

In terms of $Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, the decision maker chooses a_i whenever

$$Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) < \frac{\lambda}{1 + \lambda} \ . \tag{11.10}$$

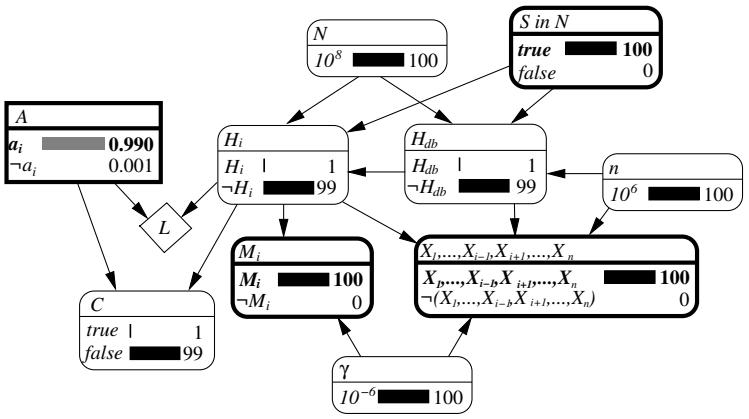


Figure 11.8: *The influence diagram in Fig. 11.4(b) with the additional Boolean node C describing a correct decision. The nodes instantiated by the user are shown with bold borders. Node A gives the expected losses on a scale from 0 (the minimum loss) to 1 (the maximum loss), and the other nodes indicate probabilities in %. If the decision maker individualizes a source found as the only match in a database search, in a case where* $N = 10^8$, $n = 10^6$, $\gamma = 10^{-6}$ *and the crime stain's source is in the population considered, there is a probability of 0.01 (state* true *of node C) that this is a correct individualization. There is therefore a probability of 0.99 (state* false *of node C) that this is a false individualization.*

This means that the probability of a false individualization is limited to the range of probabilities smaller than $\frac{\lambda}{1+\lambda}$ (Fig. 11.9).

Since the value of λ is defined by the decision maker, the decision maker fixes the maximum probability of $\neg H_i$ (Eq. (11.10)) for which he or she will conclude an individualization. Therefore, it is the decision maker himself, or herself, who determines the maximum probability of falsely individualizing a person.

Recall that λ represents the decision maker's preferences among the false conclusions. It describes the extent of the loss perceived when obtaining one false conclusion relative to the loss perceived when obtaining the other false conclusion.[14] By defining λ, the decision maker thus determines his or her risk of concluding a false individualization. The greater λ, the greater the loss of missing an individualization and the smaller the loss of falsely individualizing a source.[15] Now, the smaller the loss for a false individualization (relative to the loss of a missed individualization) the greater is the probability of obtaining a false individualization. For $0 < \lambda < 1$, the probability of a false individualization will lie between 0 and 0.5 (Fig. 11.9).

If, however, a decision maker were to associate a greater loss with a missed individualization than with a false individualization, i.e., $\lambda > 1$, this decision maker's probability of falsely individualizing a source would exceed 0.5. This result, however, would only be possible with a loss function different from the one presented in this paper.

Thus, it is the fixation of λ which determines the probability of a false individualization, not the means by which the suspect has been selected as a possible source of the crime stain (i.e., probable cause case vs. database search case), as was asserted by point C. The decision maker establishes the probability of a false individualization when he or she assigns the numerical value to λ. This value is the same for the probable cause case as for the database search case. Therefore, the probability of a false individualization is exactly the same in both scenarios, contradicting point C.

[14] Recall, for example, that $\lambda = \frac{1}{10}$ meant that the decision maker considers the loss of 1 false individualization to be equal to the loss of 10 missed individualizations (see Section 11.3.2).

[15] We refer the interested reader to the Appendix for a more theoretical explanation of the definition of λ.

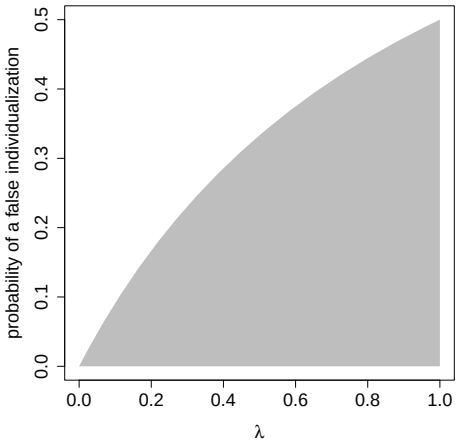


Figure 11.9: *The shaded area represents the range of values for the probability of a false individualization in function of* λ*. The probability of a false individualization is equal to* $Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ *when the decision maker chooses action* a_i*. Since the decision maker will only choose* a_i *when* $Pr(\neg H_i|M_i, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) < \frac{\lambda}{1+\lambda}$*, the probability of a false individualization lies between 0 and* $\frac{\lambda}{1+\lambda}$*.*

11.6 Discussion and conclusions

Fimmers et al. (2011) concluded that the probability of a false individualization[16] is greater in the database search case than in the probable cause case. Their conclusion (point C) is the result of a hypothetical case analysis (summarized in points A and B), in which the decision is a categorical consequence of observing a match for a profile with a match probability of $\gamma \leq 10^{-6}$. According to this line of reasoning, the probability of a false individualization increases with the number of comparisons, causing the database search case (where the number of comparisons is equal to the size of the database) to have a higher probability of a false individualization than the probable cause case (where there is only a single comparison).

The decision-theoretic approach presented in this paper reaches a different conclusion. According to this approach, the decision of individualizing individual i, $i \in \{1, \ldots, n\}$, as the source of the crime stain is a function of

(i) the probability of proposition H_i, that is, the probability that individual i is the source of the crime stain (i.e., this is the probability distribution over the states of nature); and

(ii) the decision maker's perception of the losses associated with falsely individualizing and falsely not individualizing a suspect (i.e., this is the decision maker's loss function).

A rational decision maker individualizes individual i whenever the probability of H_i exceeds a particular threshold. Note that, in agreement with Balding and Donnelly (e.g., 1996); Evett and Weir (e.g., 1998); Donnelly and Friedman (e.g., 1999); Evett et al. (e.g., 2000a); Dawid (e.g., 2001), it is the probability distribution over propositions $\{H_i, \neg H_i\}$ which is relevant, not the probability distribution over propositions $\{H_{db}, \neg H_{db}\}$. For an individual selected through a database search, the decisive threshold may be attained when

[16] In the orginal paper (Fimmers et al., 2011), they spoke of 'conviction', which is the crime level equivalent of 'individualization'. Here, we have preferred to stay at the source level to remain coherent with the propositions, which we formulated at the source level (Cook et al., 1998b). A source level analysis produces the same results as a crime level analysis assuming maximal relevance for the crime stain (e.g., Aitken and Taroni, 2004).

- $\frac{n}{N} \to 1$, that is, when almost the entire population of potential crime stain sources is registered in the database searched; or when
- $N\gamma \to 0$, that is, when the crime stain has an extremely rare characteristic in the population of potential sources, so that the expected number of matches in this population tends towards zero.

The exact transition probability marking the threshold may vary from one decision maker to another. It is entirely determined by the decision maker's loss function. This loss function describes the decision maker's preferences among the two types of false conclusions. In other words, a decision maker who considers a false individualization a lot less desirable than a missed individualization will have a high decision threshold, whereas a decision maker who considers a false individualization only slightly less desirable than a missed individualization will have a much lower decision threshold. This threshold is independent of the circumstances and probabilities in a particular case, and is therefore the same for the database search case as for the probable cause case. The use of a loss function therefore provides a coherent approach to both of these different scenarios (Lindley, 1976). Thus, the only difference between these two cases lies in the probability of H_i: the database search case involves different prior probabilities and observations than a probable cause case, which inevitably lead to a different evaluation of the posterior probability of H_i. Given the decision maker's loss function, which determines the threshold, the decision is based on the probability of H_i, regardless of whether this value results from evidence from a database search, or from evidence from other leads in the case. As a result, the probability of concluding a false individualization is independent of how individual i became a suspect in the case.

With a decision-theoretic approach, the probability of falsely individualizing a suspect found through a database search is exactly the same as the probability of falsely individualizing a suspect found by other investigational means. This probability is determined by the probability of $\neg H_i$ when the decision maker chooses to individualize individual i. It therefore depends on the probability of H_i and on the decision threshold. Since the decision maker fixes the threshold when he or she defines the loss function, it is actually the decision maker who imposes his or her own maximum probability of a false individualization. For example, a decision maker with a high decision threshold for individualizing will have low probabilities of a false individualization, and a decision maker with a lower decision threshold will have higher probabilities of a false individualization.

The results of a database search are items of evidence, which update the probability distribution over the states of nature $\{H_i, \neg H_i\}$ according to the same logic as any other type of evidence. The framework used to evaluate evidence in a probable cause case is therefore, not only an adequate framework for evaluating the evidence in a database search case, but the method of choice for a coherent evaluation in view of the decision of individualizing the suspect.

11.7 Appendix: Deriving the loss function

In this paper, we quantify the satisfaction obtained by each consequence in the form of a loss function. The formal definition of the *loss* associated with a consequence k, that is $l(C_k)$, defines this value as the difference between the utility of the most desirable consequence, which we will call $u(C*)$, and the utility of consequence k, denoted $u(C_k)$ (e.g., Lindley, 1985; Press, 2003):

$$l(C_k) = u(C*) - u(C_k) \ . \tag{11.11}$$

Any utility function may therefore be translated into a loss function. In terms of describing the desirabilities of the consequences, the two functions are identical. The loss function is therefore based on the utility scale, and its properties follow from the properties of the utility

function. In the following paragraphs, we describe these properties and explain how they appear in our loss function.

The usual utility scale ranges from 0, describing the least desirable outcome, which we will denote $c*$, to 1, describing the most desirable outcome $C*$. According to this scale, the utility of any intermediate consequence C_k, $u(C_k)$, is equal to the probability $Pr(C*)$ which makes the decision maker indifferent between the following two gambles (von Neumann and Morgenstern, 1947; Pratt et al., 1964; Lindley, 1976):

gamble 1 - obtain consequence C_k for sure;

gamble 2 - obtain consequence $C*$ with a probability of $Pr(C*)$ and consequence $c*$ with a probability of $Pr(c*) = 1 - Pr(C*)$.

In other words, it can be proved that

$$\begin{aligned} u(C_k) &= u(C*) \times Pr(C*) + u(c*) \times Pr(c*) \\ &= 1 \times Pr(C*) + 0 \times Pr(c*) \\ &= Pr(C*) \ . \end{aligned} \tag{11.12}$$

This definition provides the decision maker with a coherent way for evaluating the utility values associated with a decision's possible consequences.

Translating the utility function into a loss function using Eq. (11.11), produces a scale ranging from 0, describing this time the most desirable outcome $C*$, to 1, describing the least desirable outcome $c*$:

$$\begin{aligned} l(C*) &= u(C*) - u(C*) \\ &= 1 - 1 \\ &= 0 \ , \end{aligned} \tag{11.13}$$

and

$$\begin{aligned} l(c*) &= u(C*) - u(c*) \\ &= 1 - 0 \\ &= 1 \ . \end{aligned} \tag{11.14}$$

From Eq. (11.11) and (11.12), it follows that the loss assigned to consequence C_k, $l(C_k)$, is equal to the probability $Pr(c*)$ which makes the decision maker indifferent between the above two gambles:

$$\begin{aligned} l(C_k) &= 1 - u(C_k) \\ &= 1 - Pr(C*) \\ &= Pr(c*) \ . \end{aligned}$$

In other words, with Eq. (11.13) and (11.14), it can be proved that:

$$\begin{aligned} l(C_k) &= l(C*) \times Pr(C*) + l(c*) \times Pr(c*) \ , \\ &= 0 \times Pr(C*) + 1 \times (Pr(c*)) \\ &= Pr(c*) \ . \end{aligned} \tag{11.15}$$

We can thus define the desirabilities of the consequences as either utility or loss values. For the consequences of individualizing and not individualizing a selected person, the most desirable consequences C_1 and C_4 (i.e., the consequences describing correct conclusions) take utility values of 1 and loss values of 0, and the least desirable consequence C_2 (i.e., a false individualization) has a utility of 0 and a loss of 1:

	a_i		$\neg a_i$	
	H_i	$\neg H_i$	H_i	$\neg H_i$
consequence:	C_1	C_2	C_3	C_4
utility:	1	0	$u(C_3)$	1
loss:	0	1	$l(C_3)$	0

The satisfaction obtained by consequence C_3 (i.e., a missed individualization) usually lies somewhere inbetween the two extremes of a correct conclusion and a false individualization. We quantify the desirability of this intermediate outcome by comparing the following two gambles:

gamble 1 - obtain consequence C_3 for sure;

gamble 2 - obtain a correct conclusion (i.e., consequence C_1 or C_4) with a probability of $(1-\lambda)$ and a false individualization (consequence C_2) with a probability of λ;

where $\lambda \in [0,1]$. According to Eq. (11.12), the utility value for C_3 corresponds to the probability of obtaining a correct conclusion when the decision maker is indifferent between gamble 1 and gamble 2:

$$\begin{aligned} u(C_3) &= u(C_1 \text{ or } C_4) \times (1-\lambda) + u(C_2) \times \lambda \\ &= 1 \times (1-\lambda) + 0 \times \lambda \\ &= 1-\lambda \ . \end{aligned}$$

According to Eq. (11.15), the loss associated with C_3 is equal to the probability of obtaining a false individualization when the decision maker is indifferent between gamble 1 and gamble 2:

$$\begin{aligned} l(C_3) &= l(C_1 \text{ or } C_4) \times (1-\lambda) + l(C_2) \times \lambda \\ &= 0 \times (1-\lambda) + 1 \times \lambda \\ &= \lambda \ . \end{aligned} \tag{11.16}$$

Therefore λ is the loss value associated with a missed individualization. This parameter describes the decision maker's preferences among the false conclusions: the decision maker finds a missed individualization (C_3) as undesirable as risking a probability of λ of obtaining a false individualization (C_2).

For assigning a numerical value λ to define the decision maker's loss function, we speak, in this paper, of the decision maker's indifference between 1 false individualization and a number λ^{-1} of missed individualizations. This relationship is obtained by solving Eq. (11.16) for $l(C_2)$, given $l(C_1) = l(C_4) = 0$:

$$l(C_2) = l(C_3) \times \lambda^{-1} \ .$$

Thus, the decision maker perceives one false individualization as producing the same loss as λ^{-1} missed individualizations.

Chapter 12

Decision Analysis for the Genotype Designation in *Low-template-DNA* Profiles

Abstract

What genotype should the scientist specify for conducting a database search to try to find the source of a *low-template-DNA* (*lt-DNA*) trace? When the scientist answers this question, he or she makes a decision. Here, we approach this decision problem from a normative point of view by defining a decision-theoretic framework for answering this question for one locus. This framework combines the probability distribution describing the uncertainty over the trace's donor's possible genotypes with a loss function describing the scientist's preferences concerning false exclusions and false inclusions that may result from the database search. According to this approach, the scientist should choose the genotype designation that minimizes the expected loss. To illustrate the results produced by this approach, we apply it to two hypothetical cases: (1) the case of observing one peak for allele x_i on a single electropherogram, and (2) the case of observing one peak for allele x_i on one replicate, and a pair of peaks for alleles x_i and x_j, $i \neq j$, on a second replicate. Given that the probabilities of allele drop-out are defined as functions of the observed peak heights, the threshold values marking the turning points when the scientist should switch from one designation to another are derived in terms of the observed peak heights. For each case, sensitivity analyses show the impact of the model's parameters on these threshold values. The results support the conclusion that the procedure should not focus on a single threshold value for making this decision for all alleles, all loci and in all laboratories.

12.1 Introduction

12.1.1 Genotype designation in forensic DNA profiling

DNA database searches have become a popular investigative tool for searching for possible sources of biological traces recovered on a crime scene. Such a database search requires the scientist to specify the genotype of the trace's DNA profile. For this, the scientist observes

S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni. Decision analysis for the genotype designation in *low-template-DNA* profiles. *Forensic Science International: Genetics*, in revision, 2013.

the trace's alleles[1] for each locus on the electropherogram (EPG), and chooses a homozygous genotype designation of type $\{x_i, x_i\}$ for the observation of a single peak for allele x_i, and a heterozygous genotype designation of type $\{x_i, x_j\}$ for the observation of a pair of peaks for alleles x_i and x_j. This way of proceeding correctly describes the DNA profile of the crime stain's donor for traces not subject to stochastic effects in the DNA's amplification process. However, for DNA traces that have been called *low-template-DNA* (*lt-DNA* for short) Gill and Buckleton (2010), stochastic effects of *allele drop-out* and *allele drop-in* may cause the trace's resulting EPG to contain peaks for a different allelic configuration than that of its donor Taberlet et al. (1996); Gill et al. (2000); Whitaker et al. (2001):

- If *allele drop-out* occurs, an allele present in the donor's profile does not appear in the crime stain's profile. For example, the donor may be heterozygote $\{x_i, x_j\}$, yet if allele x_j has dropped-out in the crime stain's profile, this leads to only a single peak for allele x_i at this locus.

- If *allele drop-in* occurs, a signal for an allele not possessed by the donor shows up in the crime stain's profile. For example, the donor may be homozygote $\{x_i, x_i\}$, yet if allele x_j has dropped-in in the crime stain's profile, this produces two peaks, for alleles x_i and x_j, at this locus.

As a result, the observation of a single peak for x_i in an *lt-DNA* profile may come either from a homozygous donor $\{x_i, x_i\}$, or from a heterozygous donor where the non-x_i allele has dropped-out. Conversely, the observation of two peaks, for x_i and x_j, in an *lt-DNA* profile may come from a heterozygous donor $\{x_i, x_j\}$, or from a homozygous donor $\{x_i, x_i\}$ or $\{x_j, x_j\}$ where, respectively, an allele x_j or an allele x_i has dropped-in. In both of these cases, the designation of the donor's genotype is not straightforward. This raises an important issue regarding the genotype assignment for *lt-DNA* profiles in view of a DNA database search: which genotype should the investigator specify for conducting the database search in order to try to narrow down the population of the crime stain's possible donors?[2]

There have been several pragmatic approaches to this question. For a locus where the trace's profile shows only a single peak, let us say for allele x_i, one approach consists of comparing this peak's height (or area), denoted here by h_{x_i}, with a pre-determined threshold T Buckleton et al. (2005); Gill et al. (2009): if $h_{x_i} > T$, the locus is designated as a homozygote $\{x_i, x_i\}$, and if $h_{x_i} < T$, the locus is designated as $\{x_i, F\}$, where F represents any one of the alleles at that locus so that $\{x_i, F\}$ matches any profile at that locus with at least one allele x_i Gill et al. (2000). The use of F in the designation of the genotype indicates that allele drop-out is possible. Another approach, this time for a locus for which the trace's profile shows one or several peaks, is the multiple tube approach. In this case, the scientist obtains several EPGs for the same crime stain prior to retaining a consensus profile formed by the alleles observed a particular number of times over all of the replicates Taberlet et al. (1996).

In all of these cases, the designation of the genotype is made in view of the perceived risk of false exclusions and inclusions produced by a database search Gill et al. (2009); Gill and Buckleton (2009):

- A false exclusion occurs for example when a crime stain coming from a heterozygous donor $\{x_i, x_j\}$ is designated as being a homozygote $\{x_i, x_i\}$ or $\{x_j, x_j\}$, or any genotype

[1] Note that here and in the rest of this paper, we consider only the electropherogram's peaks that the scientist has designated as alleles.

[2] Note that the genotype designation problem is not restricted to database search cases. It is also present in what have been called probable cause cases Balding and Donnelly (1996) (or confirmation cases Donnelly and Friedman (1999)), that is, cases where the profile of a potential source is available for comparison purposes from the case's circumstancial information or investigational leads unrelated to the DNA profile. To avoid biaises when comparing the trace's profile with the potential source's profile, the scientist should also designate the trace's genotype before comparing the two profiles. The ideas presented here for a database search case therefore also apply to probable cause cases, yet the latter are not treated in this paper.

containing an allele different from x_i and x_j, so that a database search does not retain individuals having the genotype $\{x_i, x_j\}$ as possible sources of the crime stain.[3]

- False inclusions occur whenever the designation $\{x_i, F\}$ is chosen for a donor who possesses at least one copy of allele x_i, because the search results in the database will include all profiles with at least one allele x_i, instead of only the individuals with the donor's genotype.
  An extreme case of false inclusions occurs if the scientist decides not to specify anything for the locus in question when conducting the search. This corresponds to choosing the designation $\{F, F\}$. In this case, the search results for this locus include all the profiles in the database. Among these, all of the individuals that do not have the donor's genotype are considered false inclusions.

The two pragmatic approaches described above (i.e., comparing the observed peak height to a threshold or counting the number of occurrences of an allele in multiple replicates) are empirical approaches established on the basis of controlled laboratory experiments with the aim of reducing the risks of false exclusions and false inclusions. The true allelic constitution of the trace's donor, however, remains something that is unknown. For this reason, any designation of the genotype amounts to a *choice* made by the scientist. Fundamentally, this problem is therefore one of making a decision under conditions of uncertainty. In this paper, we formalize a normative approach for this decision problem.

12.1.2 Aim and outline of this paper

The aim of this paper is to present a normative approach to the decision of choosing between several possible genotype designations for an *lt-DNA* profile. A normative analysis of this problem falls into the domain of decision theory, and may be modeled by an influence diagram.

The model we present considers the genotype of a single locus of a biological trace. Because it is important to add further complications in a stepwise and informed way—so as to keep track of the model's increasing complexity—the combination of the genotypes of several loci is beyond the scope of this paper. Currently, this model is for a trace supposed to come from a single contributor. It is not applicable for a mixture of several contributors.[4] Other than the alleles coming from the crime stain's source, the model considers that the only other peaks that are possible are due to drop-in alleles, which are considered to be spurious and independent appearances of single alleles coming from multiple sources (Gill et al., 2007, p. 130). These initial assumptions may appear restrictive, but turn out to be crucial for developing a tractable, normative approach to the problem treated here. This is demonstrated by the subtle technicalities involved in the formal development outlined in the forthcoming sections.

This paper is organized as follows: Section 12.2 describes the decision-theoretic framework. We then discuss the application of this model through two examples. In the first example, we examine the case treated in Gill et al. (2009) of observing a single EPG with a single peak for the locus in question (Section 12.3). In the second example, we extend this case to the observation of two replicates, where the first replicate shows a single peak and the second replicate a pair of peaks (the one observed in the first replicate plus another one) for this locus (Section 12.4). The discussion and conclusions follow in Section 12.5.

[3] Note that even if the stain would be correctly designated as $\{x_i, x_j\}$, and the database search would not 'exclude' an individual with genotype $\{x_i, x_j\}$, this does not mean that this individual is the source of the crime stain.

[4] Traces from a single contributor affected by gross contamination (i.e., contamination from a single source that affects the entire profile Gill and Buckleton (2010)) are also considered mixtures of several contributors, because they produces a mixed profile.

12.2 Normative model for the designation of the genotype

12.2.1 Preliminaries

What genotype should the scientist assign to the source of a crime stain based on the observations made on the EPG(s) obtained from a DNA analysis of the trace? A normative approach to this decision problem models the decision as a function of the scientist's uncertainty on the genotype of the trace's source and on the utilities or losses the scientist associates with the decision's possible outcomes. This approach shows what designation the scientist should choose so as to maximize his or her expected utility or minimize his or her expected loss Lindley (1985). This section defines the elements necessary for applying this approach. Section 12.2.2 presents the notation for the donor's genotype, the possible actions and the possible outcomes. Section 12.2.3 then associates losses with each of the possible outcomes. Section 12.2.4 presents the expressions for the expected losses. Section 12.2.5 shows how this approach takes into account the DNA profiling results. And Section 12.2.6 presents this approach in the form of a graphical model.

12.2.2 Donor's genotype, possible actions and outcomes

The decision-theoretic model constructed here considers a single locus $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$, for which $x_1, x_2, \ldots, x_n$ is the exhaustive list of this locus's possible alleles.[5] The genotype of the crime stain's donor at this locus consists of a maternal allele, which we will call Θ_m, and a paternal allele, which we will call Θ_p. Each of these is equal to one of $\mathcal{X}$'s alleles. The genotype $\{\Theta_m = x_v, \Theta_p = x_v\},\ v \in \{1, 2, \ldots, n\}$, describes a homozygous donor, and $\{\Theta_m = x_v, \Theta_p = x_w\},\ v, w \in \{1, 2, \ldots, n\},\ v \neq w$, a heterozygous donor. We will abbreviate the donor's genotype with a single Θ followed by the genotype's two alleles as two subscripts to Θ as follows:

$$\Theta_{x_v x_v} = \{\Theta_m = x_v, \Theta_p = x_v\},\ v \in \{1, 2, \ldots, n\}, \quad \text{for a homozygous donor, and}$$

$$\Theta_{x_v x_w} = \{(\Theta_m = x_v, \Theta_p = x_w)\},\ v, w \in \{1, 2, \ldots, n\},\ v \neq w, \quad \text{for a heterozygous donor.}$$

Since the profile of the crime stain's donor is unknown to the scientist, there is a probability distribution over the donor's possible genotypes, reflecting the scientist's knowledge at the moment when he or she must decide what genotype to search for in the database.

In addition to this probability distribution, the model defines a space of possible actions. Let $\mathcal{A}$ be this action space. The actions in $\mathcal{A}$ are the genotype designations the scientist can specify for a database search. These cover all of the genotypes formed by $\mathcal{X}$'s alleles, all combinations of one of $\mathcal{X}$'s alleles with a wildcard F (where F represents any one of $\mathcal{X}$'s alleles), and the designation formed by two wildcards F (this option represents not specifying any alleles for locus $\mathcal{X}$ when conducting the database search). Let a lowercase letter a, followed by the genotype designation as a subscript, denote each individual action as follows:

$a_{x_i x_i}$ - designating the genotype as a homozygote $\{x_i, x_i\}$, $i \in \{1, 2, \ldots, n\}$,
$a_{x_i x_j}$ - designating the genotype as a heterozygote $\{x_i, x_j\}$, $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$,
$a_{x_i F}$ - designating the genotype as $\{x_i, F\}$, $i \in \{1, 2, \ldots, n\}$,
a_{FF} - designating the genotype as $\{F, F\}$.

The combination of each of the possible actions in the action space with each of the possible genotypes of the DNA's source produces outcomes that fall into one of the following

[5] According to this notation, x_i, $\forall i \in \{1, 2, \ldots, n\}$, is the number naming the allele (Sections 12.3 and 12.4 will present concrete examples for loci TH01 and D21S11).

four categories:

O_1 - a correct genotype designation without using F,
O_2 - an incorrect genotype designation (i.e., at least one of the alleles is incorrect),
O_3 - a correct genotype designation using one F (i.e., the non-F allele is correct), and
O_4 - an uninformative designation $\{F, F\}$.

For example, if the donor's genotype is heterozygous $\{x_v, x_w\}$ (i.e., state $\Theta_{x_v x_w}$ holds), for $v, w \in \{1, 2, \ldots, n\}$, $v \neq w$, then $a_{x_i x_j}$ leads to outcome O_1 whenever $i = v$ and $j = w$, otherwise it leads to outcome O_2. Action $a_{x_i F}$ leads to outcome O_3 whenever $i = v$ or $i = w$, otherwise it leads to outcome O_2. Action $a_{x_i x_i}$ always leads to outcome O_2, and action a_{FF} always leads to outcome O_4, regardless of the donor's genotype. Note that the dimension of a decision table formed by combining the four possible actions $a_{x_i x_i}$, $a_{x_i x_j}$, $a_{x_i F}$ and a_{FF} with each of the donor's possible genotypes is $4 \times \left[n + \frac{n(n-1)}{2}\right]$,[6] with n equal to the locus's total number of alleles, yet each of these combinations leads to one of the four outcomes described above.

12.2.3 Losses

Associated with each of these possible outcomes, there is a value (i.e., a utility or a loss) representing how desirable this outcome is for the scientist making the decision. We will need these values to formulate a decision-making criteria in Section 12.2.4. These values depend on the scientist's objective when conducting the database search. The purpose of a database search is to narrow down the population of potential crime stain donors by excluding individuals with a DNA profile different from that of the crime stain's donor (e.g., Biedermann et al., 2011a). Hence, the scientist perceives the above outcomes as more desirable or less desirable in function of how the database search results narrow down this population. Ideally, the search only retains those individuals who have the genotype of the trace's source. This is what we obtain with outcome O_1. The scientist therefore perceives O_1 as the most desirable of the four outcomes defined above. Conversely, a database search with an incorrect genotype designation (outcome O_2) does not retain the individuals with the genotype of the trace's source. This is a highly undesirable consequence because it does not correctly delimit the population of potential sources: if the trace's donor is registered in the database, the database search leads to a false exclusion. Outcomes O_3 and O_4, on the other hand, both avoid the possibility of a false exclusion. A correct designation using the wildcard F (outcome O_3) narrows down the population of potential sources to a population including individuals with the genotype of the trace's donor, yet this reduced population also contains individuals having a genotype with a one-allele difference to the donor's genotype, that is, it produces false inclusions. Having these false inclusions in the population of potential sources is undesirable because it increases the number of individuals to investigate. Finally, an uninformative designation $\{F, F\}$ (outcome O_4) does not narrow down the population of potential sources at all. All of the individuals in the database are retained as potential sources. The increased number of false inclusions in this case with regard to outcome O_3 make outcome O_4 less desirable than O_3.

The exact preference ordering of these outcomes depends on the scientist's objectives and preferences regarding a false exclusion and false inclusions. This preference ordering may depend on the severity of the case, the number of adventitious matches in the database expected from outcome O_3, and the probability that the trace's source is registered in the database. Here, we will consider the incorrect designation (outcome O_2) as the least desirable outcome, followed by the uninformative designation (outcome O_4), the correct designation using F (outcome O_3), and finally the desired, correct designation (outcome O_1). We there-

[6] This decision table's dimension is $4 \times n^2$ when the donor's genotype is represented by Θ_m and Θ_p, as is the case in the influence diagram presented in Section 12.2.6.

fore assume that:

$$O_1 \succ O_3 \succ O_4 \succ O_2 \ ,$$

where the symbol '$\succ$' means 'is preferred to'.

To qualify the desirability of each of these outcomes, we assign a loss value to each. The use of losses, rather than utilities, allows us to highlight the scientist's negative perception of the false exclusions and inclusions the database search may lead to. A loss, denoted by $l(\cdot)$, ranges here on a scale from 0 (i.e., minimal loss) to 1 (i.e., maximal loss). We assign the minimum loss value of 0 to the most desirable outcome, O_1, of a correct genotype designation:

$$l(O_1) = 0 \ . \tag{12.1}$$

The maximum loss value of 1 is assigned to the least desirable outcome, O_2, of an incorrect genotype designation:

$$l(O_2) = 1 \ . \tag{12.2}$$

The losses associated with the intermediate outcomes O_3 and O_4 are somewhere between 0 and 1. To determine these values, the scientist may compare two gambles von Neumann and Morgenstern (1947). Loss $l(O_3)$ is equal to the probability η_1 that makes the decision maker indifferent between choosing gamble 1 and choosing gamble 2, where these gambles are defined as follows:

gamble 1 - obtain outcome O_3 (a correct designation with an F) for sure;

gamble 2 - obtain the worst outcome, O_2 (an incorrect designation), with a probability of η_1 and the best outcome, O_1 (a correct designation without F), with a probability of $1 - \eta_1$.

Analogously, loss $l(O_4)$ is equal to the probability η_2 that makes the decision maker indifferent between choosing gamble 1 and choosing gamble 2, when these gambles are defined, this time, as:

gamble 1 - obtain outcome O_4 (an uninformative designation) for sure;

gamble 2 - obtain the worst outcome, O_2 (an incorrect designation), with a probability of η_2 and the best outcome, O_1 (a correct designation without F), with a probability of $1 - \eta_2$.

Values assigned to η_1 and η_2 should be coherent, so they must also make the decision maker indifferent between the following two pairs of gambles (e.g., Taroni et al., 2010):

gamble 1 - obtain outcome O_3 (a correct designation with an F) for sure;

gamble 2 - obtain outcome O_4 (an uninformative designation) with a probability of $\frac{\eta_1}{\eta_2}$ and the best outcome, O_1 (a correct designation without F), with a probability of $1 - \frac{\eta_1}{\eta_2}$;

and

gamble 1 - obtain outcome O_4 (an uninformative designation) for sure;

gamble 2 - obtain the worst outcome, O_2 (an incorrect designation), with a probability of $\frac{\eta_2 - \eta_1}{1 - \eta_1}$ and outcome O_3 (a correct designation with an F), with a probability of $1 - \frac{\eta_2 - \eta_1}{1 - \eta_1}$.

In the decision analyses that follow in Sections 12.3 and 12.4, we will use, as an example, a loss function with $\eta_1 = 0.2$ and $\eta_2 = 0.5$. Note that a different choice for η_1 and η_2, representing different objectives for the decision maker, will lead to different decision strategies than the ones we present in Sections 12.3 and 12.4, yet the underlying model remains the same.

12.2.4 Expected losses

A normative approach assumes that the scientist wants to minimize the loss produced by the database search. But since the scientist does not know the genotype of the trace's source, he or she can only choose the action that minimizes the *expected* loss. The action that minimizes the expected loss is called the Bayes action, and is the action we will determine in the decision analyses in Sections 12.3 and 12.4. An action's expected loss, denoted by $\bar{l}(\cdot)$, is the sum of the losses of each of that action's possible outcomes, weighted by the probabilities of each of these outcomes actually occurring if the action in question is chosen. For $a_{x_ix_i}$, $a_{x_ix_j}$, a_{x_iF} and a_{FF}, the expected losses are given by the following equations:

$$\begin{aligned}\bar{l}(a_{x_ix_i}) &= l(O_1) \times Pr(\Theta_{x_ix_i}) + l(O_2) \times [1 - Pr(\Theta_{x_ix_i})] \\ &= 1 - Pr(\Theta_{x_ix_i}) \ , \end{aligned} \tag{12.3}$$

$$\begin{aligned}\bar{l}(a_{x_ix_j}) &= l(O_1) \times Pr(\Theta_{x_ix_j}) + l(O_2) \times [1 - Pr(\Theta_{x_ix_j})] \\ &= 1 - Pr(\Theta_{x_ix_j}) \ , \end{aligned} \tag{12.4}$$

$$\begin{aligned}\bar{l}(a_{x_iF}) &= l(O_3) \times Pr(\Theta_{x_i-}) + l(O_2) \times [1 - Pr(\Theta_{x_i-})] \\ &= 1 + (\eta_1 - 1)Pr(\Theta_{x_i-}) \ , \end{aligned} \tag{12.5}$$

$$\begin{aligned}\bar{l}(a_{FF}) &= l(O_4) \times 1 \\ &= \eta_2 \ , \end{aligned} \tag{12.6}$$

where $Pr(\Theta_{x_i-})$ denotes the probability that the donor's genotype contains at least one copy of allele x_i. Note that here and from now on we will make the notation easier by using $Pr(\Theta_{x_ix_i}) = Pr(\Theta_{x_vx_v})$ and $Pr(\Theta_{x_ix_j}) = Pr(\Theta_{x_vx_w})$. The probabilities $Pr(\Theta_{x_ix_i})$, $Pr(\Theta_{x_ix_j})$ and $Pr(\Theta_{x_i-})$ are conditional probabilities of the donor's genotype. A priori, these probabilities reflect the occurrence of the possible genotypes in the population of potential crime stain donors based on the background information in the case.[7] A DNA analysis of the crime stain produces an EPG containing signals for locus $\mathcal{X}$'s alleles, which update this original probability distribution through Bayesian inference. For brevity, the mathematical notation that follows will omit the background information, and only indicate the observations made on the EPG(s) after a conditioning bar in the expressions for the probabilities and expected losses in Eqs. (12.3)-(12.6).

12.2.5 DNA profiling results

The model represents the results of the DNA analysis as a Boolean vector describing whether or not the scientist observes an allelic peak[8] for each of $\mathcal{X}$'s alleles on the EPG. Let us call this result R. For example,

$$R = \{x_1 = 0, x_2 = 0, \ldots, x_{i-1} = 0, x_i = 1, x_{i+1} = 0, \ldots, x_{n-1} = 0, x_n = 0\}$$

describes the observation of a single peak for allele x_i. For brevity, we will abbreviate this vector with an R followed by the observed allele(s) as subscrips to the R. So $R = \{x_1 = 0, x_2 = 0, \ldots, x_{i-1} = 0, x_i = 1, x_{i+1} = 0, \ldots, x_{n-1} = 0, x_n = 0\}$ becomes R_{x_i}. Analogously, $R_{x_ix_j}$ represents the observation of a pair of peaks for alleles x_i and x_j. The result(s) observed on the EPG(s) figure after the conditioning bar in the expressions for the expected losses and the conditional probabilities in Eqs. (12.3)-(12.6).

The results of a DNA analysis depend on the genotype of the DNA's donor and on the stochastic effects of allele drop-out and allele drop-in. Here, we have modeled these results based on the following assumptions:[9]

[7] Here, the model assumes a Hardy-Weinberg equilibrium for this distribution.
[8] Signals that the scientist has not designated as an allele are not considered.
[9] For the technical details, see 12.6.

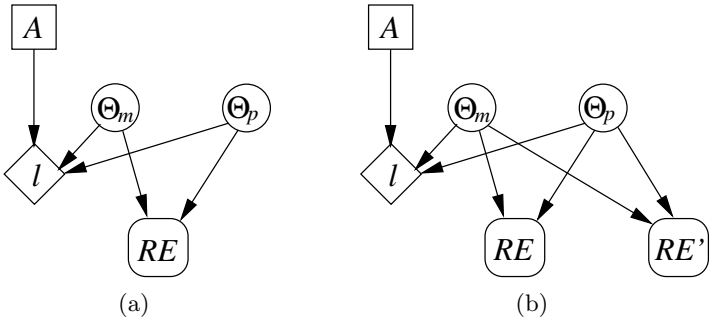


Figure 12.1: *Object-oriented influence diagrams for the genotype designation of a trace's source (node A) based on the trace's DNA profiling results:* **(a)** *These results consist of a single replicate (instance node RE).* **(b)** *These results consist of two replicates (instance nodes RE and RE′—the prime symbol is used to distinguish the second replicate from the first). These influence diagrams evaluate the expected losses for each of the possible genotype designations based on the probability distributions over the donor's maternal and paternal alleles (nodes Θ_m and Θ_p, respectively), and the losses assigned to the decision's possible outcomes (node l). The instance nodes RE and RE′ belong to the class* ***Result*** *presented in 12.6.*

- when there is no allele drop-out and no allele drop-in, the DNA profiling results accurately reflect the donor's genotype—uncertainty on the occurrence of allele drop-out and allele drop-in induces uncertainty on the true donor's genotype given the profiling results;
- the probability of allele drop-out is an increasing function of an allele's mean peak height calculated over all of the analyzed loci on the EPG (we will call this mean peak height H) according to the relationship described by Tvedebrink et al. Tvedebrink et al. (2009), and does not depend on the drop-out probabilities for previous replicates;
- the appearance of a drop-in allele is an independent event,[10] for which we assign here a probability of 0.05 according to Taberlet et al. (1996); Gill et al. (2007);
- there can be no more than two drop-in alleles in one replicate for the locus in question.

12.2.6 Graphical model

To handle the complexity caused by the numerous random variables and probabilistic relationships necessary to model the DNA results, and to combine this probabilistic part with the outcomes' losses for evaluating each action's expected loss, we constructed an object-oriented influence diagram (Fig. 12.1). An influence diagram, or Bayesian decision network, combines probability and decision theory in a graphical model (e.g., Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008). It provides the user with an interface to reason about decision making under uncertainty. Object-orientation presents an organized way to model a repetitive pattern of identical components as objects Kjaerulff and Madsen (2008). In Figures 12.1(a) and 12.1(b), the square node A contains the action space, the circular nodes Θ_m and Θ_p are the random variables representing the donor's maternal and paternal alleles, respectively, and a diamond-shaped node l contains the loss function over the space of possible outcomes (formed by combining A, Θ_m and Θ_p). Every analytical result obtained from the DNA analysis of the trace updates the probability distribution over the donor's genotype

[10] The model considers the appearance of a drop-in allele to be independent of the observed peak heights in the EPG and of the donor's true genotype. Concerning the latter, studies have shown that oversized stutters (sometimes described as drop-in alleles) are another stochastic effect that occurs in lt-DNA profiles (e.g., Gill et al., 2000; Petricevic et al., 2010; Benschop et al., 2011; Grisedale and Daal, 2012; Mitchell et al., 2012). However, the present model only considers spurious and independent appearances of single alleles as drop-in alleles, leaving this oversized-stutter effect for a future extension of this model.

(i.e., the probability distributions over the donor's possible maternal and paternal alleles). The rectangle *RE* is an object, or instance, representing one such analytical result (i.e., the observations made on one replicate). It is an instance of a network class **Result**, which models the results of the DNA analysis in function of the donor's genotype, the occurrence of 0, 1 or 2 allele drop-outs, and the presence of 0, 1 or 2 drop-in alleles (see 12.6 for the complete network structure). Here, Figure 12.1(a) shows the influence diagram's structure for a single replicate, and Figure 12.1(b) its structure for two replicates. When there is more than one replicate, we assume that the result of the additional replicate is independent of the result of the previous replicate given the genotype of the trace's donor. With this assumption, the extension of the model only requires the user to add another instance of the network **Result** to obtain one copy for each replicate. Here, and in the rest of this paper, we will use a prime symbol to distinguish the elements of the second replicate from the elements of the first replicate. The next two sections apply this model to two hypothetical cases.

12.3 Case 1: Observation of a single peak for allele x_i

Consider a case where the crime stain's DNA profiling results consist of a single EPG. For locus $\mathcal{X}$, the scientist observes a single peak on this EPG, let us say for allele x_i. Based on this result, denoted by R_{x_i}, we want to find the Bayes action.

The influence diagram in Fig. 12.1(a) evaluates the expected losses of each of the actions in $\mathcal{A}$ (in node A) when the user sets $x_i = 1$ and $x_k = 0$, $\forall\ k \in \{1, 2, \ldots, n\} \backslash \{i\}$. This updates the probability distribution over the donor's possible genotypes, which in turn updates the expected losses of the actions in $\mathcal{A}$ according to Eqs. (12.3)-(12.6).

Following the model's assumptions presented in Section 12.2, the updated probability distribution for the donor's genotype depends on:

- the allele probabilities Puch-Solis et al. (2012) of the locus's alleles in the population of the crime stain's potential sources, in particular the allele probability of the observed allele, which we will call γ_{x_i};
- the allele probability of the observed allele in the population of possible drop-in alleles, which we will call γ_{c,x_i};
- the drop-out probabilities, which are determined by a locus-specific parameter β_0 and H, the mean peak height of a single allele in the EPG (see 12.6 for further details).

For the sake of illustration in this case study, we reduce the number of parameters in this problem by assuming that the allele probability of the observed allele in the population of possible drop-in alleles is equal to the allele probability of the observed allele in the population of the crime stain's potential sources, so that $\gamma_{c,x_i} = \gamma_{x_i}$. This leaves us with the allele probabilities of the locus's alleles, the locus, and an allele's mean peak height as parameters.

The observation of a single peak for allele x_i reduces the expected loss for all of the genotype designations containing allele x_i. In most cases, the Bayes action will be either the homozygous designation for x_i, action $a_{x_i x_i}$, or the designation that uses a wildcard F to say that the genotype has at least one copy of x_i, action $a_{x_i F}$. However, for a scientist who is very reluctant to risk the possibility of a false designation, the Bayes action may also be action a_{FF}. All the other possible actions in $\mathcal{A}$ will have an expected loss that is greater than at least one of these three actions (i.e., $a_{x_i x_i}$, $a_{x_i F}$, a_{FF}), meaning that they will never be the Bayes action in this situation. For this reason, we will not comment them here. We will thus focus on the possible Bayes actions $a_{x_i x_i}$, $a_{x_i F}$ and a_{FF}, and analyze the effect of the parameters on these three actions' expected losses. Section 12.3.1 describes the parameters' effect on the expected losses $\bar{l}(a_{x_i x_i}|R_{x_i})$ and $\bar{l}(a_{x_i F}|R_{x_i})$. The expected loss $\bar{l}(a_{FF}|R_{x_i})$, on the other hand, is equal to η_2 (Eq. 12.6), and is therefore insensitive to variation in the parameters. Section 12.3.2 then compares the three expected losses for the

observation of a selection of alleles of loci TH01 and D21S11 and presents the Bayes action for each situation.

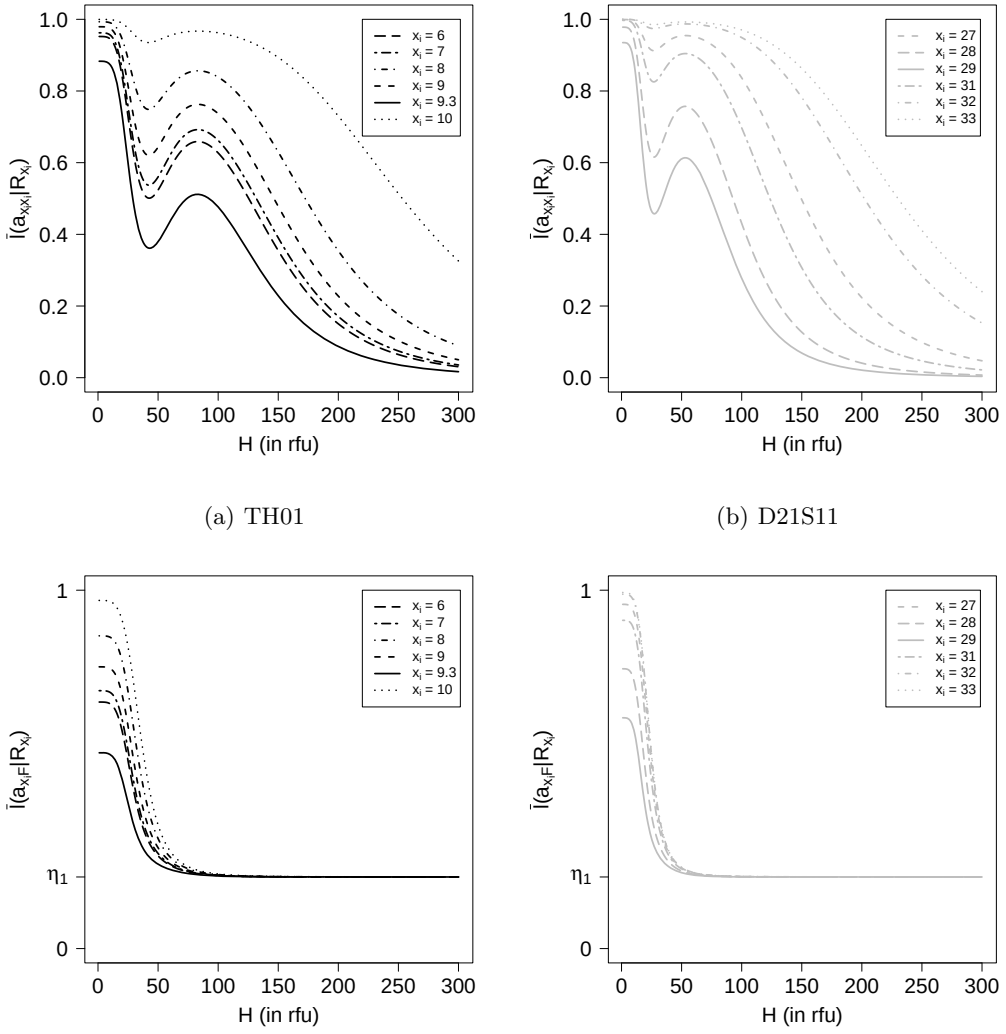
12.3.1 Sensitivity analyses of the expected losses $\bar{l}(a_{x_ix_i}|R_{x_i})$ and $\bar{l}(a_{x_iF}|R_{x_i})$

The expected losses $\bar{l}(a_{x_ix_i}|R_{x_i})$ and $\bar{l}(a_{x_iF}|R_{x_i})$ are functions of an allele's mean peak height H, the observed allele's probability γ_{x_i}, and the locus-specific parameter β_0. In addition, $\bar{l}(a_{x_iF}|R_{x_i})$ also depends on the value of η_1 (Eq. 12.5). Figure 12.2 illustrates the expected losses' dependence on H and γ_{x_i} by plotting $\bar{l}(a_{x_ix_i}|R_{x_i})$ and $\bar{l}(a_{x_iF}|R_{x_i})$ in function of H for six different allele probabilities γ_{x_i}. Figure 12.3 highlights the expected losses' dependence on β_0 by comparing the expected loss functions for two alleles with similar allele probabilities, but of different loci. The chosen loci are D21S11 and TH01, because these present minimum and maximum values for β_0, generating minimum and maximum drop-out probabilities Tvedebrink et al. (2009). The alleles' probabilities in this study are based on the population data for the Swiss Caucasian population Centre Universitaire Romand de Médecine Légale (CURML) (2008). The expected losses $\bar{l}(a_{x_ix_i}|R_{x_i})$ and $\bar{l}(a_{x_iF}|R_{x_i})$ depend on H, γ_{x_i} and β_0 as follows:

Peak height H: The expected loss $\bar{l}(a_{x_ix_i}|R_{x_i})$ generally increases as the mean peak height decreases (Fig. 12.2(a) and 12.2(b)). This relationship is due to the increasing drop-out probabilities for decreasing peak heights (see Eqs. (A.1) and (A.2) in 12.6). However, the increase is not constant, for there is a drop in each of the curves, where $\bar{l}(a_{x_ix_i}|R_{x_i})$ *decreases* as the mean peak height decreases. This is a consequence of the relationship between the probability of a single heterozygous allele dropping out and the peak height H. That is, first, as H decreases, the probability of a single, heterozygous allele dropping out increases, which decreases $Pr(\Theta_{x_ix_i}|R_{x_i})$ and increases the expected loss $\bar{l}(a_{x_ix_i}|R_{x_i})$. Then, as H continues to decrease, there comes a moment where the probability of a single, heterozygous allele dropping out *decreases* due to the increase of the probability that both heterozygous alleles drop out. This increases $Pr(\Theta_{x_ix_i}|R_{x_i})$, decreasing $\bar{l}(a_{x_ix_i}|R_{x_i})$. Then, as the probability of two allele drop-outs increases, the probability that we are acually observing a drop-in allele increases, and this decreases $Pr(\Theta_{x_ix_i}|R_{x_i})$, provoking again an increase in $\bar{l}(a_{x_ix_i}|R_{x_i})$.
The expected loss $\bar{l}(a_{x_iF}|R_{x_i})$ also increases as the mean peak height decreases, yet this increase only occurs for very low peak heights (Fig. 12.2(c) and 12.2(d)). For high values of H, $Pr(\Theta_{x_i-}|R_{x_i})$ is very close to 1, so that $\bar{l}(a_{x_iF}|R_{x_i})$ can be approximated with the constant function $\bar{l}(a_{x_iF}|R_{x_i}) = \eta_1$. For smaller peak heights, the drop-out probabilities become so important that they reduce $Pr(\Theta_{x_i-}|R_{x_i})$, causing a drastic increase in $\bar{l}(a_{x_iF}|R_{x_i})$.

Allele probability γ_{x_i}: The smaller γ_{x_i}, the greater are the expected losses $\bar{l}(a_{x_ix_i}|R_{x_i})$ and $\bar{l}(a_{x_iF}|R_{x_i})$ (e.g., compare the expected losses of alleles 9.3 ($\gamma_{9.3}$ = 0.342) and 10 (γ_{10} = 0.018) in Fig. 12.2(a) and 12.2(c), or the expected losses of alleles 29 (γ_{29} = 0.255) and 33 (γ_{33} = 0.004) in Fig. 12.2(b) and 12.2(d)). That is, the rarer allele x_i is in the population of interest, the smaller are $Pr(\Theta_{x_ix_i}|R_{x_i})$, the probability that the donor's genotype is a homozygote for this allele, and $Pr(\Theta_{x_i-}|R_{x_i})$, the probability that the donor's genotype contains at least one copy of this allele. The smaller these probabilities, the greater are the expected losses of $a_{x_ix_i}$ and a_{x_iF}.
For $\bar{l}(a_{x_ix_i}|R_{x_i})$, the allele probability γ_{x_i} also has a second effect: the smaller γ_{x_i}, the less pronounced is the drop in the curve due to the decrease of the probability of a single, heterozygous allele dropping out. This is due to a decrease of the probability $Pr(\Theta_{x_i-})$ which reduces the influence of the probability of a single, heterozygous allele dropping out on the resulting probability $Pr(\Theta_{x_ix_i}|R_{x_i})$, and hence on the expected loss $\bar{l}(a_{x_ix_i}|R_{x_i})$.

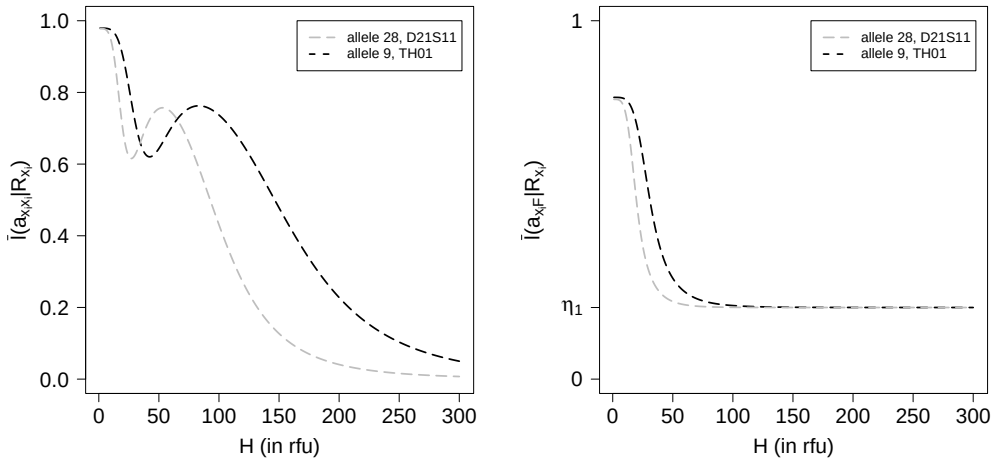
Locus (parameter β_0): According to Tvedebrink et al. (2009), the drop-out probability varies according to the locus. To take this variation into account, the equations proposed



(a) TH01 (b) D21S11

(c) TH01 (d) D21S11

Figure 12.2: **(a) and (b)** *The graphes plot the expected loss* $\bar{l}(a_{x_i x_i}|R_{x_i})$ *in function of an allele's mean peak height H (in rfu) for* $x_i = 6, 7, 8, 9, 9.3, 10$ *of locus TH01 and for* $x_i = 27, 28, 29, 31, 32, 33$ *of locus D21S11, respectively.* **(c) and (d)** *The graphes plot the expected loss* $\bar{l}(a_{x_i F}|R_{x_i})$ *in function of H for the same alleles of locus TH01 and the same alleles of locus D21S11, respectively. The probabilities of the alleles are based on the population data for Caucasians in Switzerland Centre Universitaire Romand de Médecine Légale (CURML) (2008): for TH01,* $\gamma_6 = 0.219$, $\gamma_7 = 0.194$, $\gamma_8 = 0.083$, $\gamma_9 = 0.144$, $\gamma_{9.3} = 0.342$ *and* $\gamma_{10} = 0.018$*; and for D21S11,* $\gamma_{27} = 0.025$, $\gamma_{28} = 0.148$, $\gamma_{29} = 0.255$, $\gamma_{31} = 0.054$, $\gamma_{32} = 0.007$, $\gamma_{33} = 0.004$.



(a) (b)

Figure 12.3: *The locus-specific parameter β_0 (figuring in Eqs. (A.1) and (A.2) in 12.6) is responsible for the horizontal translation of the curves for the expected losses in function of H: the larger β_0, the greater the drop-out probability for a particular H. Here, we compare the expected losses for $x_i = 28$ of locus D21S11 and for $x_i = 9$ of locus TH01, two alleles with similar allele probabilities ($\gamma_{28} = 0.148$ and $\gamma_9 = 0.144$ Centre Universitaire Romand de Médecine Légale (CURML) (2008)) from loci with different β_0 values: $\beta_0 = 17.45$ for D21S11 and $\beta_0 = 19.40$ for TH01 Tvedebrink et al. (2009).* **(a)** *The graph plots $\bar{l}(a_{x_i x_i}|R_{x_i})$.* **(b)** *The graph plots $\bar{l}(a_{x_i F}|R_{x_i})$.*

by Tvedebrink et al. Tvedebrink et al. (2009) (see Eqs. (A.1) and (A.2) in 12.6) and used in this model include a locus-specific parameter β_0. This parameter is greater for TH01 ($\beta_0 = 19.40$) than for D21S11 ($\beta_0 = 17.45$) Tvedebrink et al. (2009). The greater β_0, the greater the drop-out probabilities for a particular H. Thus, locus TH01 has greater drop-out probabilities than locus D21S11. A greater drop-out probability translates the curve for the expected loss in function of the peak height to the right, that is, towards greater values of H (Figure 12.3(a) and 12.3(b)). As a consequence, the general increase of the expected loss for locus TH01 occurs for higher values of H than for locus D21S11.

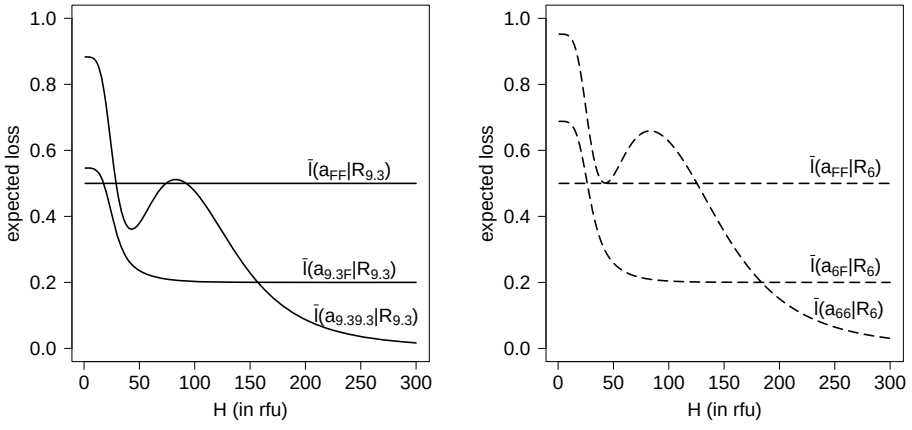
To summarize, $\bar{l}(a_{x_i x_i}|R_{x_i})$ is low when an allele's mean peak height is high, and high when the mean peak height is low, and $\bar{l}(a_{x_i F}|R_{x_i})$ tends towards a constant function equal to η_1 when the mean peak height is high, and increases for a very low mean peak height.

The next section compares $\bar{l}(a_{x_i x_i}|R_{x_i})$, $\bar{l}(a_{x_i F}|R_{x_i})$ and $\bar{l}(a_{FF}|R_{x_i})$ to determine which action the scientist should choose in which situation.

12.3.2 Bayes action

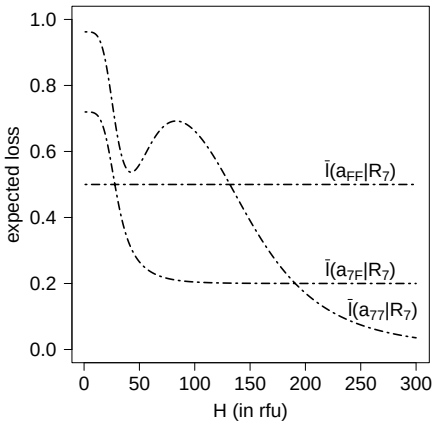
Since $\bar{l}(a_{x_i x_i}|R_{x_i})$ and $\bar{l}(a_{x_i F}|R_{x_i})$ depend on the locus, the observed allele, the mean peak height of a single allele in the profile, and the value defined for η_1, so does the Bayes action. In addition, the Bayes action also depends on η_2, since $\bar{l}(a_{FF}|R_{x_i}) = \eta_2$. Here we will determine the Bayes action for a loss function given by $\eta_1 = 0.2$ and $\eta_2 = 0.5$.

For this loss function, Figs. 12.4 and 12.5 plot the expected losses $\bar{l}(a_{x_i x_i}|R_{x_i})$, $\bar{l}(a_{x_i F}|R_{x_i})$ and $\bar{l}(a_{FF}|R_{x_i})$ in function of H for the observation of each of the alleles presented in Fig. 12.2. These graphs show that the Bayes action is $a_{x_i x_i}$ for high peak heights, $a_{x_i F}$ for lower peak heights, and a_{FF} for extremely low peak heights. The points of intersection between the different curves represent the threshold values in terms of the mean peak height H for which the scientist should switch from one designation to another. The values of these

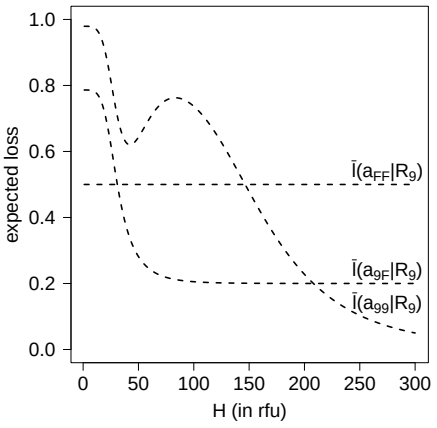


(a) TH01 allele 9.3

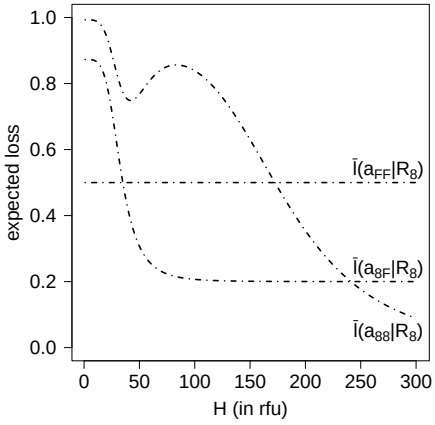
(b) TH01 allele 6



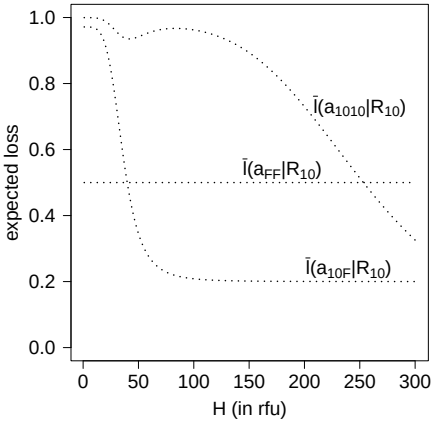
(c) TH01 allele 7



(d) TH01 allele 9

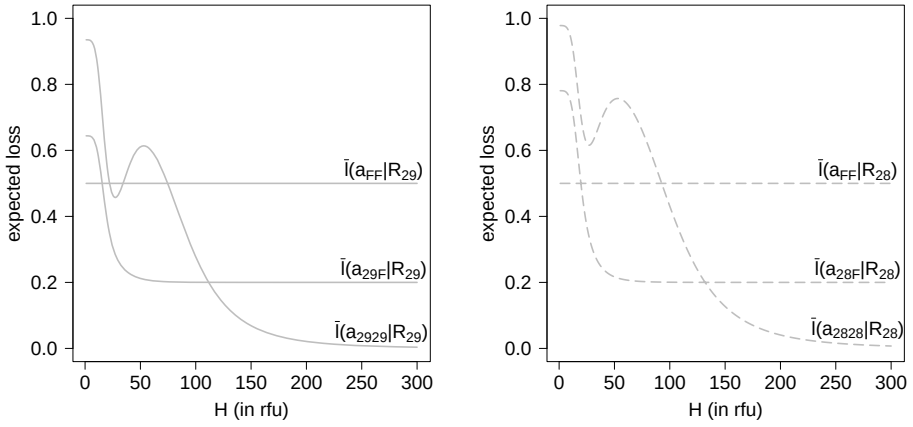


(e) TH01 allele 8



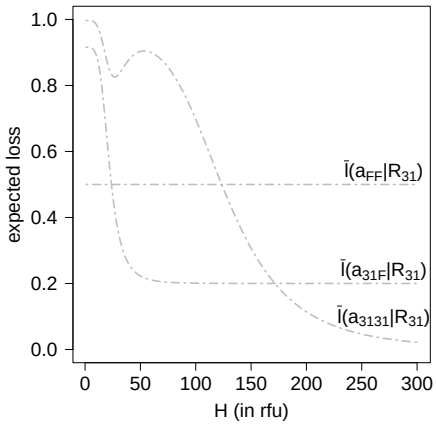
(f) TH01 allele 10

Figure 12.4: *Graphs plotting $\bar{l}(a_{x_i x_i}|R_{x_i})$, $\bar{l}(a_{x_i F}|R_{x_i})$ and $\bar{l}(a_{FF}|R_{x_i})$ in function of H for each of the six alleles of locus TH01 presented in Fig. 12.2(a) and 12.2(c), ordered here from the most common allele to the rarest allele. The loss function is defined here by $\eta_1 = 0.2$ and $\eta_2 = 0.5$. Table 12.1 presents the threshold values that mark the turning points where the Bayes action changes. Note that the peak heights for which these turning points occur are different for each of the alleles.*

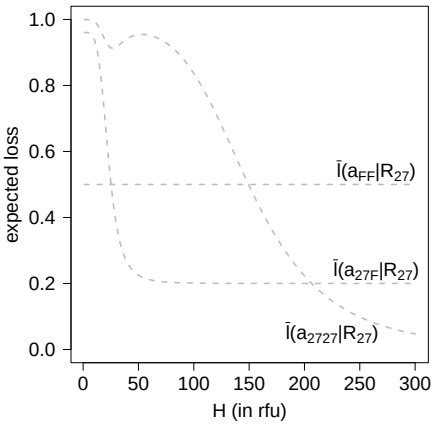


(a) D21S11 allele 29

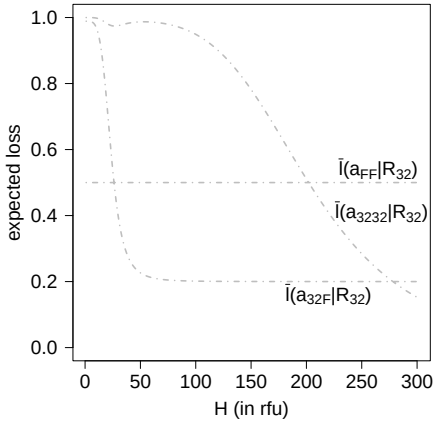
(b) D21S11 allele 28



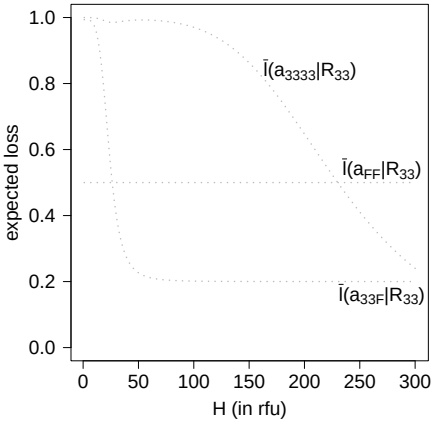
(c) D21S11 allele 31



(d) D21S11 allele 27



(e) D21S11 allele 32



(f) D21S11 allele 33

Figure 12.5: *Graphs plotting* $\bar{l}(a_{x_i x_i}|R_{x_i})$, $\bar{l}(a_{x_i F}|R_{x_i})$ *and* $\bar{l}(a_{FF}|R_{x_i})$ *in function of H for each of the six alleles of locus D21S11 presented in Fig. 12.2(b) and 12.2(d), ordered here from the most common allele to the rarest allele. The loss function is defined here by* $\eta_1 = 0.2$ *and* $\eta_2 = 0.5$. *Table 12.1 presents the threshold values that mark the turning points where the Bayes action changes. Note that the peak heights for which these turning points occur are different for each of the alleles.*

Table 12.1: *The threshold values for H (rounded to the nearest rfu) that minimize the expected loss in the choice between $a_{x_ix_i}$, a_{x_iF} and a_{FF} for $\eta_1 = 0.2$ and $\eta_2 = 0.5$. These values correspond to the points of intersection of the expected losses plotted in Fig. 12.4 and 12.5.*

$\mathcal{X}$	x_i	γ_{x_i}	a_{FF} is the Bayes action for	a_{x_iF} is the Bayes action for	$a_{x_ix_i}$ is the Bayes action for
TH01	9.3	0.342	$H \leq 17$ rfu	18 rfu $\leq H \leq$ 157 rfu	158 rfu $\leq H$
	6	0.219	$H \leq 26$ rfu	27 rfu $\leq H \leq$ 183 rfu	184 rfu $\leq H$
	7	0.194	$H \leq 27$ rfu	28 rfu $\leq H \leq$ 190 rfu	191 rfu $\leq H$
	9	0.144	$H \leq 30$ rfu	30 rfu $\leq H \leq$ 208 rfu	209 rfu $\leq H$
	8	0.083	$H \leq 35$ rfu	36 rfu $\leq H \leq$ 241 rfu	242 rfu $\leq H$
	10	0.018	$H \leq 39$ rfu	40 rfu $\leq H \leq$ 349 rfu	350 rfu $\leq H$
D21S11	29	0.255	$H \leq 15$ rfu	16 rfu $\leq H \leq$ 111 rfu	112 rfu $\leq H$
	28	0.148	$H \leq 19$ rfu	20 rfu $\leq H \leq$ 131 rfu	132 rfu $\leq H$
	31	0.054	$H \leq 23$ rfu	24 rfu $\leq H \leq$ 171 rfu	172 rfu $\leq H$
	27	0.025	$H \leq 25$ rfu	26 rfu $\leq H \leq$ 206 rfu	207 rfu $\leq H$
	32	0.007	$H \leq 25$ rfu	26 rfu $\leq H \leq$ 278 rfu	279 rfu $\leq H$
	33	0.004	$H \leq 26$ rfu	27 rfu $\leq H \leq$ 316 rfu	317 rfu $\leq H$

thresholds are locus-specific, allele-specific, and depend on the loss function. Table 12.1 presents these values for the situations plotted in Figs. 12.4 and 12.5:

- for locus TH01, the threshold between $a_{x_ix_i}$ and a_{x_iF}, above which the scientist should choose $a_{x_ix_i}$, ranges from 157 rfu (for the observation of a common allele) to 349 rfu (for the observation of a rare allele), and the threshold between a_{x_iF} and a_{FF}, below which the scientist should choose a_{FF}, from 18 rfu (for the observation of a common allele) to 40 rfu (for the observation of a rare allele);
- for locus D21S11, these thresholds range from 111 rfu (for the observation of a common allele) to 316 rfu (for the observation of a rare allele) for the choice between $a_{x_ix_i}$ and a_{x_iF}, and from 16 rfu (for the observation of a common allele) to 27 rfu (for the observation of a rare allele) for the choice between a_{x_iF} and a_{FF}.

In summary, both of the thresholds are lower for a locus with smaller drop-out probabilities, and for the observation of a more common allele.

12.4 Case 2: Observation of a single peak for allele x_i in the first replicate and a pair of peaks for alleles x_i and x_j in the second replicate

Consider now a case where the crime stain's DNA profiling results consist of two replicates. The aim of this multiple tube approach is not to produce a concensus profile, but to use the results produced by each of the replicates to produce a more informed probability distribution over the possible genotypes of the crime stain's donor Gill et al. (2000); Curran et al. (2005); Gill et al. (2007); Balding and Buckleton (2009). So, imagine that for locus $\mathcal{X}$, the scientist observes a single peak for allele x_i on the first EPG, and a pair of peaks for alleles x_i and x_j, $i \neq j$, on the second EPG. The observation on the first replicate is denoted by R_{x_i}, and the observation on the second replicate by $R'_{x_ix_j}$ (remember that the prime symbol is used to distinguish the elements of the second replicate from those of the first replicate). Again, we want to find the scientist's Bayes action.

The influence diagram in Fig. 12.1(b) evaluates these expected losses when the user sets $x_i = 1$ and $x_k = 0$, $\forall\ k \in \{1, 2, \ldots, n\}\backslash\{i\}$ in the instance network RE, and $x'_i = 1$, $x'_j = 1$ and $x'_k = 0$, $\forall\ k \in \{1, 2, \ldots, n\}\backslash\{i, j\}$ in the instance network for the second replicate RE'. The

expected losses of the actions in $\mathcal{A}$ are evaluated according to Eqs. (12.3)-(12.6) using the updated probability distribution over the donor's possible genotypes given R_{x_i} and $R'_{x_ix_j}$.

In addition to being dependent on the allele probabilities γ_{x_i} and γ_{c,x_i}, the locus-specific parameter β_0 and the mean peak height of a single allele in the EPG H, the probability distribution over the donor's possible genotypes also depends on the allele probabilities γ_{x_j} and γ_{c,x_j}, since the scientist observed a peak for allele x_j in the second replicate. Furthermore, the fact that the observations were made on two separate replicates implies that there is an H, a γ_{c,x_i} and a γ_{c,x_j} for the first replicate and an H', a γ'_{c,x_i} and a γ'_{c,x_j} for the second replicate. To reduce the number of parameters, we assume that each of the allele probabilities of the observed alleles in the population of possible drop-in alleles is equal to the allele probabilities of these alleles in the population of the crime stain's potential sources, so that $\gamma_{c,x_i} = \gamma'_{c,x_i} = \gamma_{x_i}$ and $\gamma_{c,x_j} = \gamma'_{c,x_j} = \gamma_{x_j}$. This leaves us with the alleles' probabilities, the locus, an allele's mean peak height in the first EPG H, and an allele's mean peak height in the second EPG H' as parameters.

The observation of R_{x_i} and $R'_{x_ix_j}$ makes the Bayes action one of the following genotype designations:

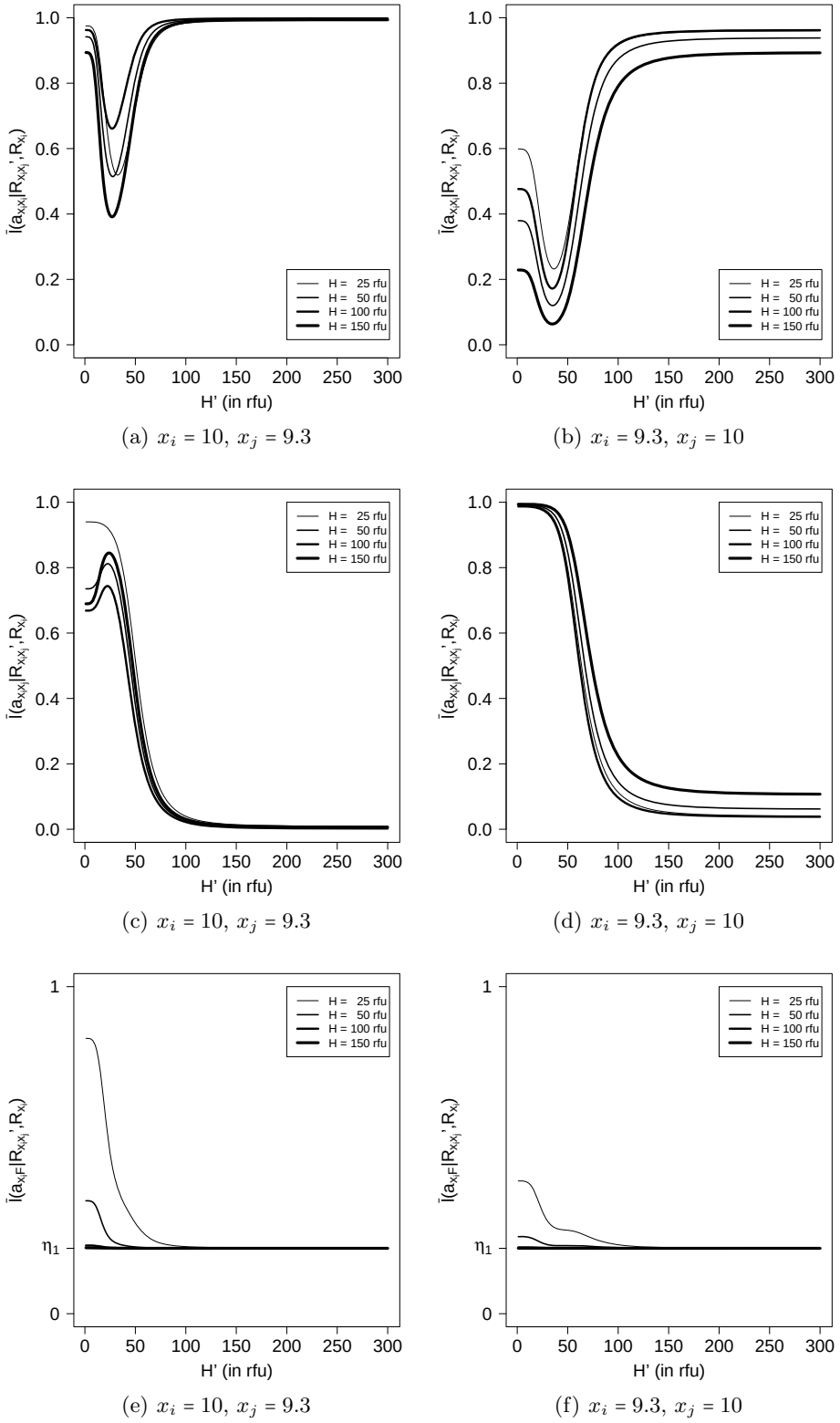
- the homozygous designation, $a_{x_ix_i}$: in this case, x_j in the second replicate is considered to be a drop-in allele;
- the heterozygous designation, $a_{x_ix_j}$: in this case, there was a drop-out of x_j in the first replicate;
- the use of a wildcard F to say that the genotype has at least one copy of allele x_i, a_{x_iF}: this action reflects some uncertainty on whether x_j is a drop-in allele or an allele present in the donor's genotype;
- the uninformative designation a_{FF}.

The decision analysis in this section therefore focuses on these four possible actions. Section 12.4.1 studies the impact of the parameters on the expected losses $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$. Remember that the expected loss $\bar{l}(a_{FF}|R'_{x_ix_j}, R_{x_i})$ is equal to η_2 and remains insensitive to variation in the parameters, and is therefore not studied in Section 12.4.1. The Bayes action is then presented for a selection of situations for locus TH01 by comparing the expected losses for all four actions in Section 12.4.2.

12.4.1 Sensitivity analyses of the expected losses $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$

The expected losses $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ are functions of the mean peak heights H and H', the allele probabilities γ_{x_i} and γ_{x_j}, and the locus-specific parameter β_0. We have already discussed the impact of β_0 in Section 12.3.1, and will not repeat it here. The focus will be on the impact of the two mean peak heights, H and H', and of the allele probabilities of the two observed alleles, γ_i and γ_j. Figure 12.6 illustrates the variation due to these parameters by plotting each of the expected losses in function of H', for different values of H, γ_{x_i} and γ_{x_j}. In particular, the expected losses' dependence on each of these parameters is the following:

Peak height H': The lower the mean peak height H', the lower is the probability of there not being any allele drop-out in the second replicate. The probability of $\Theta_{x_ix_j}$, given that the EPG shows peaks for alleles x_i and x_j, is therefore high for high values of H' and low for low values of H'. For a low H', the observation of this pair of peaks becomes more probable if $\Theta_{x_ix_i}$ is true, that is, if there was no drop-out of the homozygous peak of allele x_i, and peak x_j comes from a drop-in allele, because the event of no allelic drop-out is more probable for a homozygous peak than for heterozygous peaks. This decreases the



(a) $x_i = 10$, $x_j = 9.3$ (b) $x_i = 9.3$, $x_j = 10$

(c) $x_i = 10$, $x_j = 9.3$ (d) $x_i = 9.3$, $x_j = 10$

(e) $x_i = 10$, $x_j = 9.3$ (f) $x_i = 9.3$, $x_j = 10$

Figure 12.6: *Graphs (a) and (b) plot the expected loss* $\bar{l}(a_{x_i x_i}|R'_{x_i x_j}, R_{x_i})$ *in function of* H'*, (c) and (d) the expected loss* $\bar{l}(a_{x_i x_j}|R'_{x_i x_j}, R_{x_i})$*, and (e) and (f) the expected loss* $\bar{l}(a_{x_i F}|R'_{x_i x_j}, R_{x_i})$*. Each graph plots the expected loss for the observation of alleles 9.3 and 10 of locus TH01, with* $\gamma_{9.3} = 0.342$ *and* $\gamma_{10} = 0.018$ *Centre Universitaire Romand de Médecine Légale (CURML) (2008). Graphs (a), (c) and (e) plot the expected losses for* $x_i = 10$ *(a rare allele) and* $x_j = 9.3$ *(a very common allele) for* $H = 25, 50, 100, 150$ *rfu, and graphs (b), (d) and (f) the expected losses for* $x_i = 9.3$ *(a very common allele) and* $x_j = 10$ *(a rare allele) for* $H = 25, 50, 100, 150$ *rfu.*

expected loss $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$ and increases the expected loss $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ for lower values of H'. Apart from extremely low peak heights for H' (i.e., $H' \leq 50$ rfu), this makes the expected loss of the homozygous designation, $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, an increasing function of H' (Figs. 12.6(a) and 12.6(b)), and the expected loss of the heterozygous designation, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$, a decreasing function of H' (Figs. 12.6(c) and 12.6(d)). For extremely low values of H', the drop-out probability of a homozygous peak increases, increasing the expected loss of the homozygous designation.
As for the expected loss $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$, it tends towards η_1 as H' increases (Figs. 12.6(e) and 12.6(f)). This expected loss is slightly greater than η_1 for low peak heights H' and H because the increased probability of allele drop-out of the donor's alleles gives more weight to the possibility that the observed allele x_i is actually a drop-in allele, which increases the possibility of a_{x_iF} being a false designation. As H' increases, the drop-out probabilities decrease, so that $Pr(\Theta_{x_i-}|R'_{x_ix_j}, R_{x_i}) \to 1$, and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i}) \to \eta_1$.

Peak height H: The dependence of $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$ on H in Fig. 12.6(b) corresponds to the relationship described in the previous section (see, for example, Fig. 12.2(a)): the expected loss is generally greater for small values of H than for large values of H, yet presents a drop around 50 rfu, which is greater when x_i is a common allele than when x_i is a rare allele (see explanations on page 161). This behavior is due to the probability that one of the alleles in a heterozygous profile drops-out, which first increases as H decreases, then decreases due to the increase of the probability that both alleles in a heterozygous profile drop-out.
The opposite behavior is observed for $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ in Fig. 12.6(d): the expected loss is generally smaller for small values of H than for large values of H, because the drop-out of a heterozygous allele in the first replicate is more probable for smaller values of H. The increase of the probability of an allele drop-out in a heterozygous profile raises the probability of observing a single peak x_i given that the donor's genotype is $\Theta_{x_ix_j}$, decreasing $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$. Instead of a drop of the expected loss around 50 rfu, there is now an increase around 50 rfu, noticeable when x_i is a common allele. This is because the decrease of the probability of a single allele drop-out in a heterozygous profile, due to the increase of the probability of both alleles dropping out, decreases the probability of R_{x_i} given $\Theta_{x_ix_j}$, thus increasing $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$.
The expected loss $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ in Figs. 12.6(e) and 12.6(f) tends towards η_1 as H increases. As explained above under *Peak height H'*, the expected loss is slightly greater than η_1 for low peak heights H and H' because the increased probability of allele drop-out of the donor's alleles gives more weight to the possibility that the observed allele x_i is actually a drop-in allele, which increases the possibility of a_{x_iF} being a false designation. When H increases, the drop-out probabilities decrease, so that $Pr(\Theta_{x_i-}|R'_{x_ix_j}, R_{x_i}) \to 1$, and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i}) \to \eta_1$.

Allele probabilities γ_{x_i} and γ_{x_j}: With regard to allele x_i, the expected loss $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$ is greater for a rare allele, because the probability of this genotype is smaller for a rare x_i. For a rare allele x_i, it is more probable to observe a heterozygote than a homozygote. This increases the probability of the donor being a heterozygote, and decreases the expected loss $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$. As for allele x_j, its allele probability has practically no impact on the expected loss $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, yet the expected loss $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ is greater for a rare allele (in particular for low peak heights H'). A rare allele x_j reduces the probability of the heterozygous genotype $\Theta_{x_ix_j}$, and this reduces the probability that $a_{x_ix_j}$ leads to a correct designation, especially when the observation $R'_{x_ix_j}$ is subject to increased stochastic effects.
The increase of the expected loss $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ for low values of H' and H is greater for a rare allele x_i then for a common allele x_i, as discussed earlier in Section 12.3.1.

To summarize, $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$ is low when an allele's mean peak height is low in the second replicate and allele x_i has a high allele probability in the population of interest, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$ is low when an allele's mean peak height is high in the second replicate, and $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ tends towards a constant function equal to η_1 as an allele's mean peak height in either replicate increases.

The next section compares $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ and $\bar{l}(a_{FF}|R'_{x_ix_j}, R_{x_i})$ in several situations to determine the Bayes action.

12.4.2 Bayes action

The Bayes action depends on the locus, the observed alleles, the mean peak heights of an allele in each of the EPGs, and on the values of η_1 and η_2. Like for Case 1, we determine the Bayes action for a loss function given by $\eta_1 = 0.2$ and $\eta_2 = 0.5$.

Figure 12.7 plots the expected losses $\bar{l}(a_{x_ix_i}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_ix_j}|R'_{x_ix_j}, R_{x_i})$, $\bar{l}(a_{x_iF}|R'_{x_ix_j}, R_{x_i})$ and $\bar{l}(a_{FF}|R'_{x_ix_j}, R_{x_i})$ in function of H' for the observation of alleles 9.3 and 10 of locus TH01, for H values ranging from 50 rfu to 150 rfu. According to these graphs, the Bayes action is $a_{x_ix_j}$ for high values of H', and a_{x_iF} for lower values of H'. If allele x_i is very common in the population of potential sources, then action $a_{x_ix_i}$ is the Bayes action for a limited range of very low peak heights H'. Table 12.2 presents the threshold values in terms of H' for which the scientist should change from one designation to another for the situations plotted in Fig. 12.7. In these cases, the thresholds take the following values:

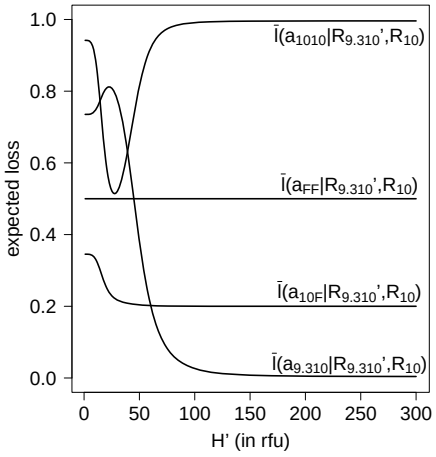
- for $x_i = 10$ and $x_j = 9.3$ (i.e., x_i is rare), the threshold between a_{x_iF} and $a_{x_ix_j}$, above which the scientist should choose $a_{x_ix_j}$, ranges from 57 rfu to 62 rfu;
- for $x_i = 9.3$ and $x_j = 10$ (i.e., x_i is very common), the threshold between a_{x_iF} and $a_{x_ix_j}$, above which the scientist should choose $a_{x_ix_j}$, ranges from 80 rfu to 105 rfu, and there are two further thresholds delimiting the range for which the Bayes action is $a_{x_ix_i}$, with the lower limit ranging from 14 rfu to 29 rfu, and the upper limit ranging from 41 rfu to 55 rfu.

In this scenario, the greatest influence on the thresholds determining the Bayes action comes from the rarity of the observed alleles in the population of potential sources. For a rare x_i, the Bayes action is a heterozygous designation, yet for a common x_i, the Bayes action may also be a homozygous designation when H' is low.

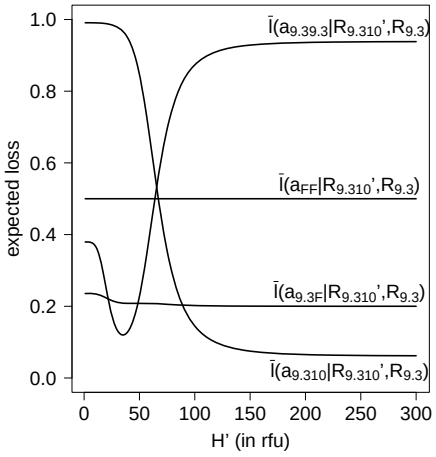
This case study illustrates how complex this decision problem can get for the observation of several replicates. In this case, the different results in each of the replicates make the decision problem particularly challenging. The Bayes action cannot be clearly stated based on the observed alleles alone. It depends on the parameters, and will therefore vary from one situation to another.

12.5 Discussion and conclusions

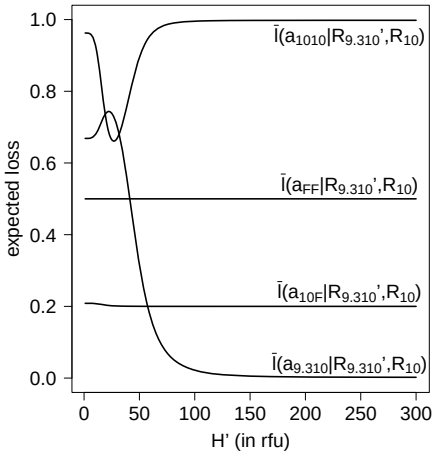
The genotype designation of a trace's source is a decision. Applying a normative approach to this decision problem shows that there is no single, universal threshold for deciding between any two designations. Here, we have shown how this threshold varies in function of the locus, the observed alleles' probabilities in the population of potential sources, and an allele's mean peak height in the EPGs of the previously obtained replicates. In addition, the thresholds depend on the model's assumptions for allele drop-outs and drop-ins, and on the scientist's preferences regarding false exclusions and false inclusions. These parameters vary in function of the case, the scientist and the laboratory: the observed alleles and mean peak heights in the replicates are case-specific; the model's assumptions for allele drop-outs and drop-ins will vary from one laboratory to another, because they should represent each laboratory's



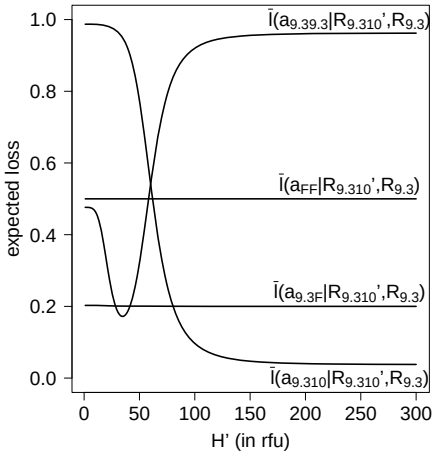
(a) H = 50 rfu, x_i = 10, x_j = 9.3



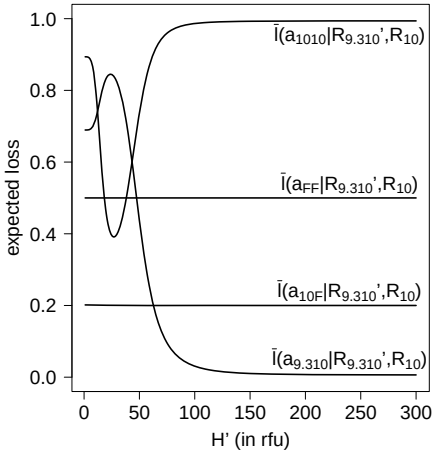
(b) H = 50 rfu, x_i = 9.3, x_j = 10



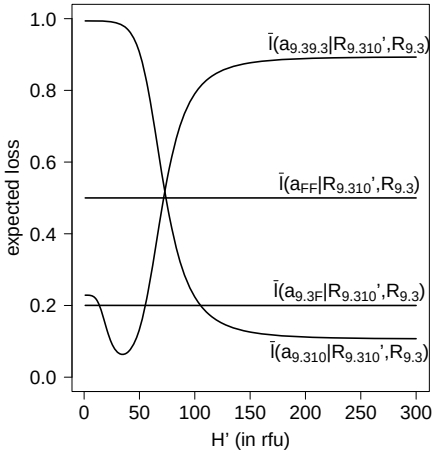
(c) H = 100 rfu, x_i = 10, x_j = 9.3



(d) H = 100 rfu, x_i = 9.3, x_j = 10



(e) H = 150 rfu, x_i = 10, x_j = 9.3



(f) H = 150 rfu, x_i = 9.3, x_j = 10

Figure 12.7: *Graphs plotting* $\bar{l}(a_{x_i x_i}|R'_{x_i x_j}, R_{x_i})$, $\bar{l}(a_{x_i x_j}|R'_{x_i x_j}, R_{x_i})$, $\bar{l}(a_{x_i F}|R'_{x_i x_j}, R_i)$ *and* $\bar{l}(a_{FF}|R'_{x_i x_j}, R_i)$ *in function of* H' *for the observation of alleles 9.3 and 10 of locus TH01, with* $\gamma_{9.3}$ = 0.342 *and* γ_{10} = 0.018 *Centre Universitaire Romand de Médecine Légale (CURML) (2008). Graphs (a) and (b) are for* H = 50 *rfu, graphs (c) and (d) for* H = 100 *rfu, and graphs (e) and (f) for* H = 150 *rfu. Graphs (a), (c) and (e) plot the expected losses when* x_i = 10 *(a rare allele) and* x_j = 9.3 *(a very common allele), and graphs (b), (d) and (f) the expected losses when* x_i = 9.3 *(a very common allele) and* x_j = 10 *(a rare allele). These two combinations represent the minimum and maximum expected losses among the possible combinations of the alleles presented in Fig. 12.2(a) and (c). Table 12.2 indicates the optimal decision strategy for each of the* H *values and allele combinations represented here.*

Table 12.2: *The threshold values for H' (rounded to the neares rfu) that minimize the expected loss in the choice between $a_{x_ix_i}$, $a_{x_ix_j}$, a_{x_iF} and a_{FF} for $\eta_1 = 0.2$ and $\eta_2 = 0.5$. These values correspond to the points of intersection of the curves minimizing the expected loss in the graphs of Fig. 12.7.*

H	x_i and x_j	a_{x_iF} is the Bayes action for	$a_{x_ix_i}$ is the Bayes action for	a_{x_iF} is the Bayes action for	$a_{x_ix_j}$ is the Bayes action for
50 rfu	$\begin{cases} x_i = 10\ ,\ x_j = 9.3 \\ x_i = 9.3\ ,\ x_j = 10 \end{cases}$	-	-	$H' \leq 60$ rfu	61 rfu $\leq H'$
		$H' \leq 21$ rfu	22 rfu $\leq H' \leq 48$ rfu	49 rfu $\leq H' \leq 88$ rfu	89 rfu $\leq H'$
100 rfu	$\begin{cases} x_i = 10\ ,\ x_j = 9.3 \\ x_i = 9.3\ ,\ x_j = 10 \end{cases}$	-	-	$H' \leq 57$ rfu	58 rfu $\leq H'$
		$H' \leq 28$ rfu	29 rfu $\leq H' \leq 41$ rfu	42 rfu $\leq H' \leq 80$ rfu	81 rfu $\leq H'$
150 rfu	$\begin{cases} x_i = 10\ ,\ x_j = 9.3 \\ x_i = 9.3\ ,\ x_j = 10 \end{cases}$	-	-	$H' \leq 62$ rfu	63 rfu $\leq H'$
		$H' \leq 13$ rfu	14 rfu $\leq H' \leq 55$ rfu	56 rfu $\leq H' \leq 105$ rfu	106 rfu $\leq H'$

performances; and the scientist's preferences will vary from one scientist to another, and may also vary from one case to another in function of the severity of the case. As a result, the numerous parameters involved make that this decision problem has no single threshold that leads to the Bayes action in every case.

According to the decision-theoretic framework proposed in this paper, the Bayes action is a function of the scientist's loss function and the probability distribution over the trace's donor's possible genotypes *in a particular case*. This model provides a case-specific approach, that can be adapted to incorporate each laboratory's drop-out and drop-in assumptions. When the model contains numerous random variables with complex probabilistic relationships, the construction of an influence diagram provides a practical means to evaluate the expected losses for each of the possible genotype designations. An object-oriented approach further allows the scientist to overcome the difficulty caused by the increasing complexity of the calculations for an increasing number of replicates. With object-orientation, this model allows the scientist to "copy and paste" the network substructure for the result of a single replicate (an instance of the network **Result**) as many times as required, and perform calculations for as many replicates as desired. This model therefore provides a logical, case-specific approach for addressing the genotype designation problem for *lt-DNA* profiles without requiring a universally pre-defined stochastic threshold.

12.6 Appendix: Network class 'Result'

The network class **Result** models the probability distribution over the DNA analysis's possible results (i.e., the observed and the unobserved alleles) in function of the donor's genotype and the stochastic effects. The instance nodes RE and RE' in the influence diagrams shown in Fig. 12.1 are instances of this class. Figure 12.8 shows the network class **Result** we used in this study (the nodes are defined in Table 12.3). This is one possible structure for this

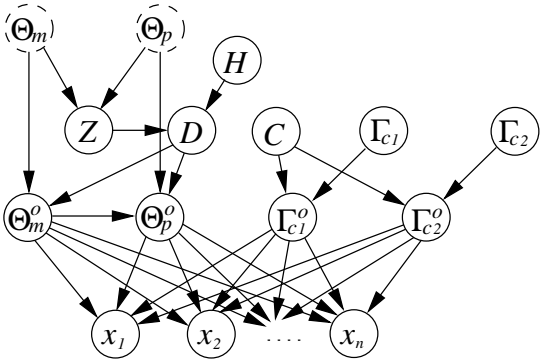


Figure 12.8: *The network class* ***Result****. This network is composed of the nodes with the continuous contours (Table 12.3 presents the definitions of these nodes). The nodes with the dashed contours represent nodes Θ_m and Θ_p in the influence diagrams in Fig. 12.1, which are input nodes for this network.*

network class, inspired by Mortera et al. Mortera et al. (2003) for nodes $x_1, x_2, \ldots, x_n$, and by Dawid et al. (2002, 2007) for the nodes of the observed alleles. Nodes $x_1, x_2, \ldots, x_n$, representing each of $\mathcal{X}$'s alleles, are Boolean nodes for the presence or absence of each allele in the trace's DNA profile. The network assumes that a maximum of four alleles may be observed at the same time: these are the donor's maternal and paternal alleles for a heterozygous donor in a case with no allele drop-outs plus two drop-in alleles. Nodes Θ_m^o and Θ_p^o indicate the observed alleles coming from the donor, and nodes Γ_{c1}^o and Γ_{c2}^o the observed drop-in alleles. Node x_k, $k = 1, 2, \ldots, n$, will take the value of *true* whenever at least one of

Table 12.3: *Definitions of the nodes in the network class* ***Result*** *shown in Fig. 12.8.*

Node	Definition
Θ^o_m	Observed donor's maternal allele
Θ^o_p	Observed donor's paternal allele
Z	Zygosity of the donor's genotype
H	Mean peak height of a single allele in the EPG
D	Event of 0, 1 or 2 allele drop-outs
C	Event of 0, 1 or 2 allele drop-ins
Γ_{c1}	(First) allele that drops in for 1 or 2 allele drop-ins
Γ_{c2}	Second allele that drops in for 2 allele drop-ins
Γ^o_{c1}	Observed drop-in allele 1
Γ^o_{c2}	Observed drop-in allele 2
x_1, x_2, $\vdots$, x_n	Presence or absence of each of locus $\mathcal{X}$'s alleles

these four nodes is equal to x_k:

$$\forall\ x_k \in \mathcal{X}\text{:}\quad x_k = \begin{cases} true & \text{if} \quad (\Theta^o_m = x_k) \cup (\Theta^o_p = x_k) \cup (\Gamma^o_{c1} = x_k) \cup (\Gamma^o_{c2} = x_k) \\ false & \text{otherwise .} \end{cases}$$

The rest of the network is divided into two parts: the left-hand side models the observed alleles coming from the donor and the right-hand side the observed drop-in alleles (considered to be independent of the donor's alleles). Below we describe each of these parts in turn.

On the left-hand side, the observed donor's alleles depend on the donor's maternal and paternal alleles (nodes Θ_m and Θ_p, respectively) and on the number of alleles that have dropped out (node D). This number may be equal to 0, 1 or 2 for heterozygous genotypes, and to 0 or 2 for homozygous genotypes. Let the states D_0, D_1 and D_2 denote the events of 0, 1 and 2 allele drop-outs, respectively. The states of Θ^o_m and Θ^o_p consist of each of the locus's possible alleles plus the observation of no allele, denoted here by $\emptyset$. The conditional probability table for node Θ^o_m is filled out as follows:

$$Pr(\Theta^o_m = \Theta_m) = \begin{cases} 1 & \text{if} \quad D = D_0 \\ 0.5 & \text{if} \quad D = D_1 \\ 0 & \text{if} \quad D = D_2\ , \end{cases}$$

$$Pr(\Theta^o_m = \emptyset) = \begin{cases} 0 & \text{if} \quad D = D_0 \\ 0.5 & \text{if} \quad D = D_1 \\ 1 & \text{if} \quad D = D_2\ , \end{cases}$$

and 0's for the rest of the possible states. The conditional probability table for node Θ^o_p is filled out in the same way for $D = D_0$ and $D = D_2$, yet $Pr(\Theta^o_p = \Theta_p)$ and $Pr(\Theta^o_p = \emptyset)$ will take values of either 1 or 0 when $D = D_1$, depending on whether $\Theta^o_m = \emptyset$ or $\Theta^o_m \neq \emptyset$. The conditional probability table for node Θ^o_p is therefore filled out as follows:

$$Pr(\Theta^o_p = \Theta_p) = \begin{cases} 1 & \text{if} \quad D = D_0 \\ 1 & \text{if} \quad D = D_1 \quad \text{and} \quad \Theta^o_m = \emptyset \\ 0 & \text{if} \quad D = D_1 \quad \text{and} \quad \Theta^o_m \neq \emptyset \\ 0 & \text{if} \quad D = D_2\ , \end{cases}$$

$$Pr(\Theta_p^o = \emptyset) = \begin{cases} 0 & \text{if} \quad D = D_0 \\ 0 & \text{if} \quad D = D_1 \quad \text{and} \quad \Theta_m^o = \emptyset \\ 1 & \text{if} \quad D = D_1 \quad \text{and} \quad \Theta_m^o \neq \emptyset \\ 1 & \text{if} \quad D = D_2 \ , \end{cases}$$

and 0's for the rest of the possible states.

The probability distribution over D_0, D_1 and D_2 (in node D) is a function of the observed peak heights in relative fluorescence units (rfu)[11] in the EPG (node H) and of the zygosity of the donor's genotype (node Z):

node H: The definition of H depends on the model used for assigning the drop-out probabilities. For example, if one would retain the approach presented in Gill et al. (2009), node H would represent the peak height of the single peak present at the locus under examination. Or, if we adhere to the model presented in Tvedebrink et al. (2009), this would suggest defining H as the mean peak height of a single allele calculated over all of the analyzed loci on the EPG. For the purpose of the discussion in this paper, we apply the model presented in Tvedebrink et al. (2009), so that the states of H represent the mean peak height of an allele in the resulting profile. Using a Bayesian network construction software (*Hugin Researcher* by Hugin Expert A/S), we modeled the target range of possible peak heights as a set of intervals of 5 rfu. Based on the observations made on the EPG, the user instantiates this node (i.e., sets the probability of the observed peak height to 1) in order to carry out the decision-theoretic analysis for the obtained DNA profile.[12] The node's initial probability distribution is therefore irrelevant.

node Z: This node has two states: state Z_o describes a homozygous genotype and state Z_e a heterozygous genotype.

Empirical studies have shown that allele drop-out tends to be encountered more often as the quantity of DNA in the trace material decreases Gill et al. (2000). The peak heights in the resulting DNA profile reflect the quantity of analyzed DNA Tvedebrink et al. (2010). For this reason, we model allele drop-out as a child[13] node of the observed peak height (node H). Several studies have produced models using logistic regression to model the drop-out probability in function of the observed peak height Gill et al. (2009); Tvedebrink et al. (2009). It is important to note that the actual relationship between the peak height and the occurrence of drop-out depends on the laboratory's equipment and protocol, and is therefore established on the basis of each laboratory's experimental data. This relationship is also locus-specific Tvedebrink et al. (2009). In this paper, we apply the model presented in Tvedebrink et al. (2009) to describe the relationship between H and D. According to this model, the drop-out probability, let us call it d, is specified in function of the locus (through a locus-specific parameter β_0) and the mean peak height for one allele in the profile (variable H). Given that a peak in a heterozygous genotype corresponds to one allele, whereas a peak in a homozygous genotype corresponds to two alleles, the drop-out probabilities are different for heterozygous (Z_e) and homozygous (Z_o) genotypes. These drop-out probabilities, denoted here by d_{Z_e} and d_{Z_o}, respectively, are given by the following expressions Tvedebrink et al. (2009):

$$d_{Z_e} = \frac{exp\left[\beta_0 - 4.35ln(H)\right]}{1 + exp\left[\beta_0 - 4.35ln(H)\right]} \tag{A.1}$$

[11] Note that it could just as well be the peak area for a model of the drop-out probability in function of the peak area.

[12] If the obtained EPG does not allow the scientist to determine the value or interval of values of the mean peak height of a single allele, then it is also possible for the user to specify a probability distribution over the possible values of H.

[13] If an arrow goes from node 1 to node 2, so that we have node 1 → node 2, then node 2 is called a *child* of node 1. This means that the probability distribution over the states of node 2 is conditioned by the state of node 1.

Table 12.4: *The probability distribution in node D depends on whether the donor's genotype is homozygote (state Z_o) or heterozygote (state Z_e). Eqs. (A.1) and (A.2) give the expressions for d_{Z_e} and d_{Z_o}.*

$Z:$	Z_o	Z_e
$Pr(D_0)$	$1 - d_{Z_o}$	$1 - 2d_{Z_e}(1 - d_{Z_e}) - (d_{Z_e})^2$
$Pr(D_1)$	0	$2d_{Z_e}(1 - d_{Z_e})$
$Pr(D_2)$	d_{Z_o}	$(d_{Z_e})^2$

for the drop-out of a peak of a heterozygous genotype, and

$$d_{Z_o} = \frac{exp[\beta_0 - 4.35ln(2H)]}{1 + exp[\beta_0 - 4.35ln(2H)]} \tag{A.2}$$

for the drop-out of the peak of a homozygous genotype. Hence, node D is also a child of Z (indicating whether the donor is homozygote or heterozygote). Table 12.4 gives the probability distributions in node D in function of Z. In the case studies in Sections 12.3 and 12.4, we apply this model to loci D21S11 and TH01. We chose these loci because D21S11 presents the minimum drop-out probabilities, with $\beta_0 = 17.45$, and TH01 the maximum drop-out probabilities, with $\beta_0 = 19.40$, according to the data referred to in Tvedebrink et al. (2009).

On the right-hand side, the observed drop-in alleles depend on the number of drop-in alleles (node C) and on the probability distribution over the locus's possible drop-in alleles (nodes Γ_{c1} and Γ_{c2}). C_0 is the event of no allele drop-in at this locus, C_1 the event that one allele has dropped in, and C_2 the event that two alleles have dropped in. We abbreviate the probabilities of these two states with their lowercase equivalents: $Pr(C_0) = c_0$, $Pr(C_1) = c_1$ and $Pr(C_2) = c_2$. Since the drop-ins are considered to be independent appearances of single alleles, we have $c_2 = c_1^2$ and $c_0 = 1 - c_1 - c_1^2$. Thus, all we need to define the probability distribution over C is a probability assignment for c_1. This probability may vary from one laboratory to another, so that each laboratory must determine its own value for c_1.[14] Publications have reported average drop-in probabilities of 0.05 Gill et al. (2007), ≤ 0.05 Taberlet et al. (1996), 0.02 Cowen et al. (2011), 0.018 Mitchell et al. (2012) and 0.0134 Petricevic et al. (2010). In this study, we used $c_1 = 0.05$. The states of nodes Γ^o_{c1} and Γ^o_{c2} consist of each of the locus's possible alleles plus the obseration of no allele, $\emptyset$. If 0 alleles have dropped in, Γ^o_{c1} and Γ^o_{c1} are both equal to $\emptyset$. If a single allele has dropped in, node Γ_{c1} specifies which allele has dropped-in, node Γ^o_{c1} copies the state of Γ_{c1}, and node Γ^o_{c2} is equal to $\emptyset$. If two alleles have dropped in, both nodes Γ_{c1} and Γ_{c2} specify which alleles have dropped in, and node Γ^o_{c1} copies the state of Γ_{c1}, and node Γ^o_{c2} the state of Γ_{c2}. The conditional probability table for node Γ^o_{c1} is therefore filled out as follows:

$$Pr(\Gamma^o_{c1} = \Gamma_{c1}) = \begin{cases} 0 & \text{if} \quad C = C_0 \\ 1 & \text{otherwise} \end{cases},$$

$$Pr(\Gamma^o_{c1} = \emptyset) = \begin{cases} 1 & \text{if} \quad C = C_0 \\ 0 & \text{otherwise} \end{cases},$$

with 0's for the remaining possible states. And the conditional probability table of Γ^o_{c2} is filled out as follows:

$$Pr(\Gamma^o_{c2} = \Gamma_{c2}) = \begin{cases} 1 & \text{if} \quad C = C_2 \\ 0 & \text{otherwise} \end{cases},$$

[14] The literature proposes a way of assigning a value to c_1 on the basis of negative controls Gill and Kirkham (2004), for example by dividing the total number of alleles observed on these controls by the total number of loci tested over all of the negative controls performed Gill et al. (2007).

$$Pr(\Gamma^{o}_{c2} = \emptyset) = \begin{cases} 0 & \text{if} \quad C = C_2 \\ 1 & \text{otherwise} , \end{cases}$$

with 0's for the remaining possible states.

Chapter 13

Decision-theoretic Reflections on Processing a Fingermark

Abstract

A recent publication in this journal (Neumann et al., 2011b) presented the results of a field study that revealed the data provided by the fingermarks not processed in a forensic science laboratory. In their study, the authors were interested in the *usefulness* of this additional data in order to determine whether such fingermarks would have been worth submitting to the fingermark processing workflow. Taking these ideas as a starting point, this communication here places the fingermark in its context of a case brought before a court, and examines the question of processing or not processing a fingermark from a decision-theoretic point of view. The decision-theoretic framework presented provides an answer to this question in the form of a quantified expression of the expected value of information ($EVOI$) associated with the processed fingermark, which can then be compared with the cost of processing the mark.

13.1 Introduction

"To process or not to process a fingermark?" That is the question examined by the field study conducted by Neumann et al. (2011b). The question of processing or not processing a fingermark is one of the decisions confronting a fingerprint examiner for each detected fingermark (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011). It is the decision of whether to submit a recovered trace to a fingermark examination process, which seeks to associate the mark with its source[1].

The study conducted by Neumann et al. (2011b) investigated the usefulness of fingermarks not processed in a forensic laboratory. These unexploited fingermarks consisted of fingermarks that were either not recovered on evidentiary items, or categorized as being of *no value* during an initial analysis phase. The goal of their study was to bring to light the results of comparing these marks with fingerprints of potential sources, and to compare

Reprinted from Forensic Science International, 226, S. Gittelson, S. Bozza, A. Biedermann, and F. Taroni, Decision-theoretic reflections on processing a fingermark, pages e42-e47, Copyright (2013), with permission from Elsevier.

[1] The word *source* refers not only to a particular person, but to the particular area of the person's friction ridge skin that is at the origin of the mark. For simplicity we will only consider friction ridge marks coming from fingers, and therefore speak of a finger as being the source of a fingermark (see Neumann et al. (2011a) for an explanation of the difference between considering a finger and considering a person as being the source of a fingermark). Yet, the ideas presented here for fingermarks may also be extended to other friction ridge marks.

these findings with the additional resources (in USD and in man-hours) required to obtain this additional data. In doing this, they pursued the fundamental issue of whether it is possible for the laboratory to make more beneficial decisions for processing or not processing a fingermark. The results of their study indicated that the average cost for obtaining an additional piece of evidence was about $3000, and they concluded that the laboratory's current procedures "are well optimized for cost-effectiveness" (Neumann et al., 2011b, p. 46).

When is it worth investing a laboratory's resources in processing a fingermark? From a decision-theoretic point of view, this question can be answered on the basis of a quantified expression of the **expected value of information (*EVOI*)** associated with the processed fingermark: a rational decision-maker would decide to process a fingermark when the *EVOI* is greater than or equal to the cost of processing the mark. The objective of the pages that follow is to present a decision-theoretic development for determining the *EVOI* for a fingermark.

This paper is organized as follows: Section 13.2 describes the decision problem, Section 13.3 the decision-theoretic framework for approaching this problem, Section 13.4 the construction of the corresponding influence diagram, and Section 13.5 summarizes the main points with regard to a practical application of the model.

13.2 The question: "To process or not to process a fingermark?"

The fingermark examination process uses the laboratory's resources (quantifiable in monetary units and in man-hours) to compare a fingermark with the fingerprints of potential sources. Submitting the fingermark to this process may produce data in the form of similarities or dissimilarities with a comparison print. This data may lead to the association[2] of the mark with the print of a potential source and/or to exclusions of possible sources. Let us call the fingermark's true source finger F*. Now, if the data narrows down the pool of possible sources to a smaller population containing F* (in the ideal case to a single finger which is F*), it correctly supports a proposition of type "the fingermark comes from finger F*" with regard to the alternative "the fingermark does not come from finger F*, it comes from another finger in the pool of possible sources". However, if the data leads to an erroneous conclusion (i.e., a false association or a false exclusion), or if processing the fingermark does not produce any discriminatory data, the laboratory will have spent its resources in vain, and an erroneous conclusion may lead to an erroneous verdict. A fingerprint examiner deciding to process or not to process a fingermark will attempt to distinguish between these two situations, that is, between: (i) cases where a fingermark's discriminatory data will narrow down the pool of possible sources to a smaller population containing F*, and (ii) cases where such discriminatory data will either be absent or lead to an erroneous conclusion. The examiner will attempt to process only the fingermarks expected to fall into category (i). Hence, a fingermark expected to lead to a strong association (such as an individualization) with a print of the mark's source, and sometimes also fingermarks expected to lead only to correct exclusions, will be considered *suitable for a comparison*, and a fingermark that does not fulfill this expectation will be categorized as being of *no value for a comparison* (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011).

If a fingermark's value reduces to its ability to correctly support a proposition, then the solution to the decision problem of processing or not processing a fingermark is to quantify the expected support for a proposition, and compare it to a pre-defined threshold, above which all marks would be processed. However, a fingermark *suitable for a comparison* does

[2]Note that by *association* we do not imply an individualization. An association is any conclusion tending towards an individualization on a continuous scale of conclusions.

not necessarily represent *useful information*. Neumann et al. (2011b) recognized this as they stated on page 40:

> "... all these extra associations were made in cases where at least one conclusion of identification had already been expressed, hence reducing their potential usefulness."

For a fingermark to be of real value, it must not only support the correct proposition: it must embody useful information. But what is *useful information*? In decision theory, the usefulness of information depends on the decision maker's objective. The definition of useful information therefore depends on *what the information will be used for*. For this, let us take a step back and look at the fingermark in its context in a case brought before a court. In this context, processing a fingermark is one of many forensic analyses. By definition, every forensic analysis brought before court has one and the same ultimate aim: to assist judicial decision-making. Hence, the results of a comparison between a fingermark and a suspect's print is used to help a judicial decision-maker (i.e., a judge or a jury) decide on the verdict of a case. The comparison results of the fingermark are thus useful information if they help the judicial decision-maker reach a correct decision. Therefore, the value of the comparison results for a given fingermark is its contribution to reaching a correct judicial decision.[3]

Thus, we are not interested in one decision, but in two decisions: (i) the decision to process or not to process a fingermark, and (ii) the judicial decision in the fingermark's case. We will call the latter, the **ultimate decision**. This makes the decision of processing or not processing a fingermark a **preliminary decision**: more precisely, it is a preliminary decision of investing in generating data (i.e., by processing a mark) that may help the judicial decision-maker make the ultimate decision. The decision of processing or not processing a fingermark thus requires an evaluation of the expected value of any thus generated data with regard to the ultimate decision. Decision theory provides a framework for this in terms of the *expected value of information* ($EVOI$), which we present in the next section.

13.3 Decision-theoretic framework

A decision-theoretic framework allows us to specify the $EVOI$ of a fingermark in a quantitative way with respect to the ultimate decision (i.e., the judicial decision made by the judge or jury in the context of the fingermark's case). The framework we present here is based on the existing works of Kaplan (1968), Lindley (1977a), Lempert (1977) and Kaye (1999). As the ultimate decision, let us consider the judicial decision of convicting or acquitting a defendant, whom we will call individual D. We define:

- an action space $\mathcal{A}$ consisting of the judicial decision-maker's possible actions a_i, $i = 0, 1$:

  a_0 - acquit D,

  a_1 - convict D;

- a set Θ consisting of the unknown states of nature Θ_j, $j = 0, 1$:

  Θ_0 - D is innocent of the alleged offence,

  Θ_1 - D is guilty of the alleged offence;

- and a set $\mathcal{C} = \mathcal{A} \times \Theta$ of the possible consequences of the actions, where the elements of this set, C_{ij}, denote the consequence of having chosen action a_i when the state of nature Θ_j holds:

[3] This is true for the results of any forensic analysis, not just for the results of fingermark comparisons. Actually, it is the value of any scientific finding brought before court. For explanations on the role of an item of evidence in the logical framework of judicial decision-making, see for example Kaplan (1968); Lindley (1977a); Lempert (1977).

C_{00} - correct acquittal,

C_{01} - false acquittal,

C_{10} - false conviction,

C_{11} - correct conviction.

A utility function describes the satisfaction obtained from each of the consequences. The utility of consequence C_{ij}, denoted $u(C_{ij})$, is a numerical description of how desirable this consequence is (von Neumann and Morgenstern, 1947). According to Lindley (1977a), this utility function is imposed by the objectives and preferences of society.

If the decision maker wants to maximize this satisfaction, then the most rational action is the one with the greatest expected utility. The expected utility of action a_i, denoted $\bar{u}(a_i|\cdot)$, is the weighted average of the utilities of the consequences with respect to the probabilities of each consequence occurring as a result of choosing action a_i, given the information available at the time of making the decision (this information is denoted here by the centered dot after the conditioning bar):

$$\bar{u}(a_i|\cdot) = \sum_{j=0}^{1} u(C_{ij}) Pr(\Theta_j|\cdot) \ .$$

The probability $Pr(\Theta_j|\cdot)$ is the conditional probability of Θ_j being true. This probability is assigned on the basis of the information available to the decision-maker at the time he or she chooses action a_i. What information is available will depend on whether the fingermark was processed or not processed.

For an unprocessed fingermark, the probability of the state Θ_j is conditioned on the background information I, which includes the case circumstances and the other evidence informing the judicial decision-maker's probability distribution over Θ prior to obtaining the results from a comparison of the fingermark with fingerprints of its potential sources. We denote this probability $Pr(\Theta_j|I)$. The maximum expected utility of choosing an action in $\mathcal{A}$ if the fingermark is not processed is therefore:[4]

$$max_i \ \bar{u}(a_i|I) = max_i \left\{ u(C_{i0}) Pr(\Theta_0|I) + u(C_{i1}) Pr(\Theta_1|I) \right\} \ . \tag{13.1}$$

For a processed fingermark, the probability distribution over Θ is conditioned on both the background information I and on the results obtained from processing the fingermark. We denote this probability $Pr(\Theta_j|E, I)$ with E describing the results obtained from a comparison of the fingermark with a fingerprint of finger F, one of the defendant's ten fingers. These results consist of the observations of a set of features on the fingermark and on the fingerprint of finger F. Each of these features may be categorized as belonging to one of three levels: Level 1 (general ridge flow), Level 2 (minutiae and individual ridge paths) and Level 3 (ridge shape and pores) (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011). To present here an illustrative example that keeps the complexity in the equations that follow at a limited level, let us consider a case where the observations are limited to Level 1 and Level 2 features. Further, let us assume that comparing the fingermark with a fingerprint will lead either to the result that the features of Level l, $l = 1, 2$, are similar (i.e., fall within a pre-defined range of variation known as the mark's tolerance (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011)) or dissimilar (i.e., do not fall within this range of variation). We will use the following notation:

E_0^l - dissimilarity of the features of Level l,

E_1^l - similarity of the features of Level l,

[4]This legal application of the formula for the maximum expected utility has been discussed elsewhere in the literature (e.g., Kaplan, 1968; Lindley, 1977a; Lempert, 1977).

so that $E = \{E_k^1, E_m^2\}$, with $k, m \in \{0, 1\}$.

However, when deciding whether or not to process the fingermark, the examiner has not yet compared the fingermark with the fingerprint of finger F. At this moment, one does not know yet what similarities and dissimilarities he or she will observe. The maximum expected utility of choosing an action in $\mathcal{A}$ for a fingermark to be processed is therefore the weighted average of the maximum expected utility obtained after each possible result with respect to the probabilities of each of the possible results:

$$\begin{aligned}
&\sum_{k=0}^{1}\sum_{m=0}^{1} max_i\ \bar{u}(a_i|E_m^2, E_k^1, I)Pr(E_m^2|E_k^1, I)Pr(E_k^1|I) \\
&\qquad = \sum_{k=0}^{1}\sum_{m=0}^{1} max_i \sum_{j=0}^{1} u(C_{ij})Pr(\Theta_j|E_m^2, E_k^1, I)Pr(E_m^2|E_k^1, I)Pr(E_k^1|I)\ , \qquad (13.2)\\
&\qquad \text{which, according to Bayes' theorem, is equal to} \\
&\qquad = \sum_{k=0}^{1}\sum_{m=0}^{1} max_i \sum_{j=0}^{1} u(C_{ij})Pr(E_m^2|E_k^1, \Theta_j, I)Pr(E_k^1|\Theta_j, I)Pr(\Theta_j|I)\ .
\end{aligned}$$

The *expected value of the information* for a fingermark to be processed, denoted $EVOI$, is the difference between Eqs. (13.2) and (13.1):

$$\begin{aligned}
EVOI &= \sum_{k=0}^{1}\sum_{m=0}^{1} max_i\ \bar{u}(a_i|E_m^2, E_k^1, I)Pr(E_m^2|E_k^1, I)Pr(E_k^1|I) - max_i\ \bar{u}(a_i|I) \\
&= \sum_{k=0}^{1}\sum_{m=0}^{1} max_i \sum_{j=0}^{1} u(C_{ij})Pr(E_m^2|E_k^1, \Theta_j, I)Pr(E_k^1|\Theta_j, I)Pr(\Theta_j|I) \\
&\quad - max_i \sum_{j=0}^{1} u(C_{ij})Pr(\Theta_j|I)\ . \qquad (13.3)
\end{aligned}$$

This value is always greater than or equal to zero, reflecting the informative value of observed results for making the ultimate decision (e.g., Raiffa and Schlaifer, 1961; Good, 1967; DeGroot, 1984). We must now compare this informative value with the cost of processing a fingermark. If c is the financial cost of processing the fingermark, then δ is the most rational decision rule for the preliminary decision:

$$\delta = \begin{cases} \text{process the fingermark} & \text{if} \quad EVOI \geq c \\ \text{not process the fingermark} & \text{if} \quad EVOI < c\ . \end{cases} \qquad (13.4)$$

This decision rule requires the $EVOI$ to be expressed in the same units as c. The units of the $EVOI$ are the units of the utilities. The utility function over the space of consequences $\mathcal{C}$ must therefore be defined in monetary units, such as dollars. On this scale, the utility can be positive, negative, or zero. Thus, utilities $u(C_{11})$ and $u(C_{00})$ represent how much society gains from a correct conviction and from a correct acquittal, respectively. Conversely, utilities $u(C_{10})$ and $u(C_{01})$ quantify how much society pays as a result of a false conviction and as a result of a false acquittal, respectively. These latter utilities therefore take negative values. The major challenge in implementing this framework lies in the difficulty of quantifying these utilities. As stated above and emphasized by Lindley (1977a, pp. 209-210):

> "(...) the utility here is imposed by society, certainly not by the defendant and not even by the judge."

In addition to the utility function, the implementation of this decision-theoretic framework requires the prior probability distribution over Θ and the conditional probability distribution over E given Θ. The judicial decision-maker assigns the former based on the background information. This probability distribution is the result of taking into account and combining all of the previously obtained knowledge informing the decision-maker's belief

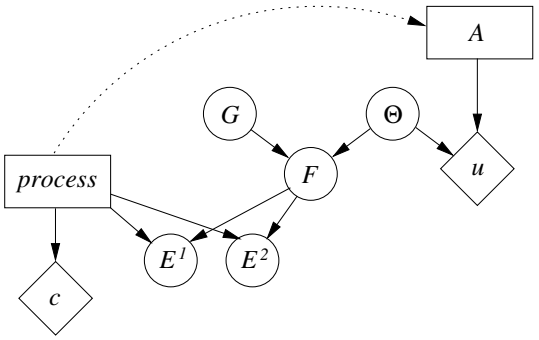


Figure 13.1: *A generic influence diagram for the decision-theoretic framework presented in this paper for the decision of processing or not processing a fingermark (node process). This decision is a preliminary decision for the judicial decision in node A. The continuous arrows represent the network's functional relationships, whereas the dotted arrow stands for a precedence link telling the user that the decision process precedes decision A. The random variables Θ (whether the defendant is innocent or guilty of the alleged offence), G (whether the fingermark comes from an offender's finger), F (whether the fingermark comes from finger F, where F is one of the defendant's fingers), and E^1 and E^2 (the results of comparing the fingermark with a print of finger F) are modeled according to the Bayesian network proposed by Garbolino and Taroni (2002). The utility nodes c and u contain the cost of processing the fingermark, and the utility function over consequences C_{ij}, $i = 0, 1$ and $j = 0, 1$, respectively.*

in the truth of the states of Θ (for example, statements from witnesses and the results of other forensic analyses, such as DNA, fibres, etc.). We do not treat this aspect in further detail in this paper. The latter probability distribution (i.e., the probability distribution over E given Θ) requires the consideration of additional variables. To handle the uncertainty related to these variables in a transparent way, we propose to construct a graphical probability model. We propose an influence diagram, because influence diagrams handle uncertainty and decision theory in a single model (e.g., Shachter, 1988; Howard and Matheson, 2005).

13.4 Influence diagram

An influence diagram is a graphical probability model which, in addition to probabilistic calculations, also calculates expected utilities (e.g., Jensen and Nielsen, 2007; Kjaerulff and Madsen, 2008; Korb and Nicholson, 2011). It is made up of circular nodes representing random variables, rectangular nodes representing action nodes, and diamond-shaped nodes representing utility nodes. Continuous arrows represent the network's functional relationships between the nodes (i.e., the dependence relationships that determine the model's evaluation process), and a dotted arrow stands for a precedence link indicating that one decision precedes another decision (e.g., Korb and Nicholson, 2011). The model calculates the expected utilities for each of the actions based on the probability distributions associated with each of the relevant random variables and the utility values specified in the utility nodes for each of the possible consequences. Its graphical representation of the dependence relationships between the nodes and rigorous computational methods underlying these relationships provides a structured and transparent approach for evaluating these expected utilities and the probabilities required in complex decision problems involving uncertainty. Fig. 13.1 presents the structure of an influence diagram for the decision-theoretic framework presented in this paper. It implements the well known test-action sequence of decisions, because it consists of two decisions, where the first (the preliminary decision) may produce observations that help the decision-maker make the second decision (the ultimate decision). Fig. 13.1 models this decision problem according to the general structure proposed by Korb and Nicholson (2011, page 106), with the action nodes interpreted here as:

process - the preliminary decision of processing or not processing the fingermark,

Table 13.1: *Conditional probability table associated with node F, based on the tables presented by Taroni et al. (2006a). The definitions of parameters p and w are given in the text on page 184.*

	Θ_0		Θ_1	
	G_0	G_1	G_0	G_1
F_0	$1-p$	1	1	$1-w$
F_1	p	0	0	w

A - the ultimate decision of convicting or acquitting defendant D;

the utility nodes as:

c - the cost of processing the fingermark,
u - the utilities assigned to consequences C_{ij};

and the random variables as described earlier in Section 13.3:

Θ - the state of nature,
E^1 - the comparison results of the fingermark's Level 1 features,
E^2 - the comparison results of the fingermark's Level 2 features.

To model the relationship between Θ and $\{E^1, E^2\}$, the influence diagram uses the Bayesian network presented by Garbolino and Taroni (2002), containing the Boolean random variables:

G - whether the fingermark comes from an offender's finger,
F - whether the fingermark comes from finger F.

In node G, the user specifies a probability distribution over the states:

G_0 - the fingermark does not come from an offender's finger,
G_1 - the fingermark comes from an offender's finger;

based on information regarding the fingermark's location and the case circumstances. Forensic science literature (e.g., Stoney, 1991b; Evett, 1993; Stoney, 1994; Evett et al., 1998; Aitken and Taroni, 2004; Taroni et al., 2006a) denotes these probabilities as $Pr(G_1|I) = r$ and $Pr(G_0|I) = 1 - r$.

Node F contains the states:

F_0 - the fingermark does not come from finger F,
F_1 - the fingermark comes from finger F;

with the probability distribution over these states defined as follows (see Table 13.1 for the conditional probability table associated with node F):

- If defendant D is innocent of the offence, and the fingermark does not come from an offender's finger (column 1 of Table 13.1), there is a probability of p that the fingermark comes from finger F. This probability describes the possibility that finger F left the fingermark as a result of an action unrelated to the commission of the offence, as well as the possibility that the fingermark is a fake for which finger F served as a mold.
- If defendant D is innocent of the offence, yet the fingermark comes from one of an offender's fingers (column 2 of Table 13.1), it is impossible for the fingermark to come from finger F.
- If defendant D is guilty of the offence, yet the fingermark does not come from an offender's finger (column 3 of Table 13.1), it is impossible for the fingermark to come from finger F.

Table 13.2: *The conditional probability tables associated with nodes E^1 and E^2. The definitions of parameters d and f are given in the text on page 185. A prime symbol is added to the parameters in the probability table of node E^2 to distinguish these from the parameters in the probability table of node E^1.*

		process		*not process*	
		F_0	F_1	F_0	F_1
E^1:	E^1_0	$1-f$	d	0	0
	E^1_1	f	$1-d$	0	0
	$E^1_\dagger$	0	0	1	1
E^2:	E^2_0	$1-f'$	d'	0	0
	E^2_1	f'	$1-d'$	0	0
	$E^2_\dagger$	0	0	1	1

- If defendant D is guilty of the offence, and the fingermark comes from an offender's finger (column 4 of Table 13.1), there is a probability of w that the fingermark comes from finger F. For example, in a case with a single offender where we assume that each of the offender's ten fingers has an equal probability of being the source of the fingermark, $w = \frac{1}{10}$.[5]

If the fingermark is to be processed, the probabilities of observing E^1_k and E^2_m depend on which state of node F is true. These probabilities are defined as follows (see Table 13.2 for the conditional probability tables associated with nodes E^1 and E^2):

- If the fingermark does not come from finger F (column 1 of Table 2), there is a probability of f of observing a similarity between the fingermark's features and the features of a print of finger F. This probability reflects the rarity of the features in the population of potential sources. To distinguish this probability for the Level 1 features from the corresponding probability for the Level 2 features, we call the former f and the latter f'.

- In a case where the fingermark comes from finger F (column 2 of Table 2), there is a probability of d of observing a dissimilarity between the fingermark's features and the features of a print of finger F. This probability takes into account the possibility that factors, such as a distorted fingermark, a morphological change on finger F during the lapse of time between the creation of the mark and the print (such as scars and warts), the nature of the transfer media, the properties of the substrate, the techniques applied to detect and recover the mark, and the environmental conditions the mark was subject to (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011, Section 2), produce a dissimilarity of the observed features between the mark and its source's print. To distinguish between this probability for Level 1 and for Level 2 features, d refers to the Level 1 features and d' to the Level 2 features.

The conditional probability tables of nodes E^1 and E^2 list a third state: $E^1_\dagger$ and $E^2_\dagger$, respectively. This state describes the absence of data, which is the case when the examiner decides not to process the mark.

According to the influence diagram in Fig. 13.1, the probabilities of observing E^1_k and E^2_m given Θ_j in Eq. (13.3) are a function of the parameters specified in the conditional probability tables of nodes E^1 and E^2, the parameters in the conditional probability table of node F, and the probability distribution over F's parent node G. Section 13.6 presents the algebraic formula produced for evaluating the $EVOI$ for processing a fingermark when all of

[5] Note that in this case of a single offender, w is equal to Neumann et al. (2011a)'s probability $Pr(G = g|I)$ where g denotes the finger number of finger F.

these parameters are introduced into Eq. (13.3) according to the probabilistic relationships modeled by the influence diagram. In this case, the $EVOI$ has been developed for only two levels of features, with the simplifying assumption that processing the fingermark must result in either a similarity or a dissimilarity in a comparison of each level of features with a print of finger F (i.e., processing the fingermark cannot lead to the absence of data in the levels considered). Relaxing this assumption, and extending the model to Level 3 features, would make the computation of the conditional probabilities of E given Θ a lot more involved, and this would considerably increase the complexity of the development of an algebraic expression for the $EVOI$. However, an influence diagram allows its user to handle this complexity by directly computing the expected utilities $\bar{u}(a_i|I)$ and $\bar{u}(a_i|E_m^2, E_k^1, I)$,[6] and the probabilities $Pr(E_k^1|I)$ and $Pr(E_m^2|E_k^1, I)$ for evaluating the $EVOI$ (see the first line in Eq. (13.3)).

13.5 Discussion and conclusions

From a decision-theoretic point of view, the decision "To process or not to process a fingermark?" is a preliminary decision of investing or not investing in information that may help a judicial decision-maker decide on the verdict of the case. If the results produced by processing the fingermark contribute to reaching a correct verdict, then these results are useful information. Quantifying the usefulness of this information on the same scale as the cost of processing the fingermark allows us to formulate a decision rule based on a direct comparison of these two values.

There are four factors that determine how useful the information of a processed fingermark is with regard to the judicial decision of determining the verdict of the case:

(1) the ability of the fingermark's features to help the examiner correctly associate the fingermark with a print of its source and exclude the other potential sources (i.e., the fingermark's suitability for a comparison (Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST), 2011));

(2) the fingermark's relevance in the case;

(3) the other information in the case;

(4) the gain obtained from pronouncing a correct verdict, and the cost of pronouncing a false verdict.

The decision-theoretic model presented in this paper incorporates all of these factors:

(1) parameters d and f, defined for each level of features, quantify each level's ability to help the examiner correctly associate the fingermark with a print of it's source and exclude the other potential sources;

(2) parameter r represents the fingermark's relevance;

(3) the initial probability distribution over Θ describes the judicial decision-maker's prior beliefs in the case, based on the case's other information;

(4) a monetary utility function quantifies the gain, or cost, of each verdict given each of the possible states of nature (i.e., whether defendant D is innocent or guilty of the alleged offence).

Defining the utility function in monetary units produces the $EVOI$ of processing a fingermark also in monetary units. According to this approach, a rational decision maker processes a fingermark when the fingermark's $EVOI$ is greater than or equal to the processing cost.

[6] Note that the influence diagram will actually compute $\bar{u}(a_i|E_m^2, E_k^1, I) - c$ instead of $\bar{u}(a_i|E_m^2, E_k^1, I)$. To obtain the value of $\bar{u}(a_i|E_m^2, E_k^1, I)$ in Eq. (13.3), one must add c to the value indicated by the model.

Neumann et al. (2011b)'s field study actually provides us with this processing cost for a forensic laboratory: $\frac{\$138{,}000}{1619} \approx \$85$ (or ≈ 44 minutes in man-hours). This means that this laboratory's most rational decision rule δ (Eq. (13.4)) for processing or not processing a fingermark is:

$$\delta = \begin{cases} \text{process the fingermark} & \text{if} \quad EVOI \geq \$85 \\ \text{not process the fingermark} & \text{if} \quad EVOI < \$85 \ . \end{cases}$$

However, this decision cannot be made by the forensic laboratory alone. A practical application of this model requires the user to assign values to the parameters, a prior probability distribution over Θ, and a monetary utility function for the possible consequences. The first of these three (assigning the values to the parameters) should be doable based on the case's circumstances and on appropriate data from fingerprint studies. The second point (assigning the prior probability distribution over Θ) refers to the judicial decision-maker's degrees of belief, an aspect inherent in the judicial decision-making process. This probability distribution can therefore only be assigned by the judicial decision-maker. And finally, the third point (the definition of the utility function) remains the major challenge for applying such a decision-theoretic framework in practice (e.g., Taroni et al., 2010): this function quantifies the benefits and costs of correct and false verdicts according to the objectives and preferences of society (Lindley, 1977a). How can we define this function so that it accurately represents the utilities imposed by society? The definition of this function is fundamental for any practical application of the model, and thus produces the most important issue to be solved for applying it in practice. Yet, applicable in practice or not, a normative decision model such as the one presented here remains the rational means for justifying to process (or not to process) a fingermark in the light of the uncertainty and the context of the case that inevitably surround it.

13.6 Appendix

Introducing parameters d, d', f, f', p, r and w into Eq. (13.3), according to the probabilistic relationships modeled by the influence diagram in Fig. 13.1, produces the following expression for the $EVOI$ for processing a fingermark:

$$
\begin{aligned}
EVOI = \; & max\ \{u(C_{00})[(1-f')(1-f)(1-p+pr) + d'dp(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{01})[(1-f')(1-f)(1-rw) + d'dwr]Pr(\Theta_1|I), \\
& \quad u(C_{10})[(1-f')(1-f)(1-p+pr) + d'dp(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{11})[(1-f')(1-f)(1-rw) + d'dwr]Pr(\Theta_1|I)\} \\
& + max\ \{u(C_{00})[f'(1-f)(1-p+pr) + (1-d')dp(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{01})[f'(1-f)(1-rw) + (1-d')dwr]Pr(\Theta_1|I), \\
& \quad u(C_{10})[f'(1-f)(1-p+pr) + (1-d')dp(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{11})[f'(1-f)(1-rw) + (1-d')dwr]Pr(\Theta_1|I)\} \\
& + max\ \{u(C_{00})[(1-f')f(1-p+pr) + d'(1-d)p(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{01})[(1-f')f(1-rw) + d'(1-d)wr]Pr(\Theta_1|I), \\
& \quad u(C_{10})[(1-f')f(1-p+pr) + d'(1-d)p(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{11})[(1-f')f(1-rw) + d'(1-d)wr]Pr(\Theta_1|I)\} \\
& + max\ \{u(C_{00})[f'f(1-p+pr) + (1-d')(1-d)p(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{01})[f'f(1-rw) + (1-d')(1-d)wr]Pr(\Theta_1|I), \\
& \quad u(C_{10})[f'f(1-p+pr) + (1-d')(1-d)p(1-r)]Pr(\Theta_0|I) \\
& \quad + u(C_{11})[f'f(1-rw) + (1-d')(1-d)wr]Pr(\Theta_1|I)\} \\
& - max_i \sum_{j=0}^{1} u(C_{ij})Pr(\Theta_j|I)\ .
\end{aligned}
$$

The algebraic developments in this equation are based on the formulae published by Evett (1993); Garbolino and Taroni (2002); Taroni et al. (2006a).

Bibliography

C.G.G. Aitken. Sampling - How big a sample? *Journal of Forensic Sciences*, 44:750–760, 1999.

C.G.G. Aitken and A.J. Gammerman. Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society*, 29:303–316, 1989.

C.G.G. Aitken and D.A. Stoney. *The Use of Statistics in Forensic Science*. Ellis Horwood, New York, 1991.

C.G.G. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 2nd edition, 2004.

C.G.G. Aitken and F. Taroni. Fundamentals of statistical evidence - a primer for legal professionals. *International Journal of Evidence and Proof*, 12:181–207, 2008.

C.G.G. Aitken, F. Taroni, and P. Garbolino. A graphical model for the evaluation of cross-transfer evidence in DNA profiles. *Theoretical Population Biology*, 63:179–190, 2003.

C.G.G. Aitken, P. Roberts, and G. Jackson. *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. Royal Statistical Society's Working Group on Statistics and the Law, London, 2010.

D.J. Balding. The DNA database search controversy. *Biometrics*, 58:241–244, 2002.

D.J. Balding and J. Buckleton. Interpreting low template DNA profiles. *Forensic Science International: Genetics*, 4:1–10, 2009.

D.J. Balding and P. Donnelly. Inference in forensic identification. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 158:21–53, 1995.

D.J. Balding and P. Donnelly. Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Sciences*, 41:603–607, 1996.

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

C.C.G. Benschop, C.P. van der Beek, H.C. Meiland, A.G.M. van Gorp, A.A. Westen, and T. Sijen. Low template STR typing: Effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics*, 5:316–328, 2011.

J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 2nd edition, 2010.

D. Bernoulli. Exposition of a new theory on measurement of risk. *Econometrica*, 22:23–36, 1954.

A. Biedermann. *Bayesian networks and the evaluation of scientific evidence in forensic science*. PhD thesis, School of Criminal Justice, University of Lausanne, Lausanne, 2007.

A. Biedermann and F. Taroni. A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues. *Forensic Science International*, 163:18–33, 2006.

A. Biedermann and F. Taroni. Bayesian networks for evaluating forensic DNA profiling evidence: A review and guide to literature. *Forensic Science International: Genetics*, 6: 147–157, 2012.

A. Biedermann, F. Taroni, O. Delemont, C. Semadeni, and A.C. Davison. The evaluation of evidence in the forensic investigation of fire incidents (Part I): an approach using Bayesian networks. *Forensic Science International*, 147:49–57, 2005a.

A. Biedermann, F. Taroni, O. Delemont, C. Semadeni, and A.C. Davison. The evaluation of evidence in the forensic investigation of fire incidents (Part II): Practical examples of the use of Bayesian networks. *Forensic Science International*, 147:59–69, 2005b.

A. Biedermann, S. Bozza, and F. Taroni. Decision theoretic properties of forensic identification: Underlying logic and argumentative implications. *Forensic Science International*, 177:120–132, 2008a.

A. Biedermann, F. Taroni, S. Bozza, and C.G.G. Aitken. Analysis of sampling issues using Bayesian networks. *Law, Probability and Risk*, 7:35–60, 2008b.

A. Biedermann, S. Bozza, and F. Taroni. Probabilistic evidential assessment of gunshot residue particle evidence (Part I): Likelihood ratio calculation and case pre-assessment using Bayesian networks. *Forensic Science International*, 191:24–35, 2009a.

A. Biedermann, F. Taroni, and S. Bozza. Implementing statistical learning methods through Bayesian networks. Part 1: A guide to Bayesian parameter estimation using forensic science data. *Forensic Science International*, 193:63–71, 2009b.

A. Biedermann, S. Gittelson, and F. Taroni. Recent misconceptions about the 'database search problem': a probabilistic analysis using Bayesian networks. *Forensic Science International*, 212:51–60, 2011a.

A. Biedermann, F. Taroni, S. Bozza, and W.D. Mazzella. Implementing statistical learning methods through Bayesian networks (Part 2): Bayesian evaluations for results of black toner analyses in forensic document examination. *Forensic Science International*, 204: 58–66, 2011b.

A. Biedermann, S. Bozza, P. Garbolino, and F. Taroni. Decision-theoretic analysis of forensic sampling criteria using Bayesian decision networks. *Forensic Science International*, 223: 217–227, 2012a.

A. Biedermann, S. Bozza, K. Konis, and F. Taroni. Inference about the number of contributors to a DNA mixture: Comparative analyses of a Bayesian network approach and the maximum allele count method. *Forensic Science International: Genetics*, 6:689–696, 2012b.

A. Biedermann, P. Garbolino, and F. Taroni. The subjectivist interpretation of probability and the problem of individualization in forensic science. *Science & Justice*, 53:192–200, 2013.

W. Blackstone. Commentaries on the Laws of England, Book 4, Chapter 27. Made available by the Lonang Institute: *http://www.lonang.com/exlibris/blackstone/bla-427.htm*, (last visited on 23.06.2013), 1765-1769.

D.C. Blakey. Does forensic science give value for money? *Science & Justice*, 35:1–4, 1995.

J.S. Buckleton, C.M. Triggs, and S.J. Walsh. *Forensic DNA Evidence Interpretation*. CRC Press, Boca Raton, 2005.

J.M. Butler. Introduction to interpretation issues. In *DNA Mixture Interpretation Webcast* at *http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation-webcast.cfm* (last visited on 26.06.2013). National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, 12 April 2013.

Centre Universitaire Romand de Médecine Légale (CURML). Fréquences alléliques suisses. Personal communication: August 6, 2008.

C. Champod and F. Taroni. Les préjugés de l'accusation ou de la défense dans l'évaluation de la preuve technique. *Revue Pénale Suisse*, 111:223–235, 1993.

C. Champod and F. Taroni. Probabilités au procès pénal - risques et solutions. *Revue Pénale Suisse*, 112:194–219, 1994.

D. Charlton, P.A.F. Fraser-Mackenzie, and I.E. Dror. Emotional experiences and motivating factors associated with fingerprint analysis. *Journal of Forensic Sciences*, 55:385–393, 2010.

T. Connolly. Decision theory, reasonable doubt, and the utility of erroneous acquittals. *Law and Human Behavior*, 11:101–112, 1987.

R. Cook, I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. A model for case assessment and interpretation. *Science & Justice*, 38:151–156, 1998a.

R. Cook, I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. A hierarchy of propositions: deciding which level to address in casework. *Science & Justice*, 38:231–239, 1998b.

R. Cook, I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. Case pre-assessment and review in a two-way transfer case. *Science & Justice*, 39:103–111, 1999.

J. Cornfield. Bayes theorem. *Review of the International Statistical Institute*, 35:34–49, 1967.

R.G. Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer, New York, 2007a.

R.G. Cowell, S.L. Lauritzen, and J. Mortera. Identification and seperation of DNA mixtures using peak area information. *Forensic Science International*, 166:28–34, 2007b.

R.G. Cowell, S.L. Lauritzen, and J. Mortera. Probabilistic modelling for DNA mixture analysis. *Forensic Science International: Genetics Supplement Series*, 1:640–642, 2008.

S. Cowen, P. Debenham, A. Dixon, S. Kutranov, J. Thomson, and K. Way. An investigation of the robustness of the consensus method of interpreting low-template DNA profiles. *Forensic Science International: Genetics*, 5:400–406, 2011.

J.M. Curran, P. Gill, and M.R. Bill. Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure. *Forensic Science International*, 148:47–53, 2005.

A.P. Dawid. Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search". *Biometrics*, 57:976–978, 2001.

A.P. Dawid. Which likelihood ratio? (comment on 'why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence', by Ronald Meester and Marjan Sjerps). *Law, Probability and Risk*, 3:65–71, 2004.

A.P. Dawid and I.W. Evett. Using a graphical method to assist the evaluation of complicated patterns of evidence. *Journal of Forensic Sciences*, 42:226–231, 1997.

A.P. Dawid, J. Mortera, V.L. Pascali, and D. van Boxel. Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, 29:577–595, 2002.

A.P. Dawid, J. Mortera, and P. Vicard. Object-oriented Bayesian networks for complex forensic DNA profiling problems. *Forensic Science International*, 169:195–205, 2007.

B. de Finetti. Logical foundations and measurement of subjective probability. *Acta Psychologica*, 34:129–145, 1970.

B. de Finetti. Foresight: Its logical laws, its subjective sources (Paper originally published in *Annales de l'Institut Henri Poincaré*, 7:1-68, 1937). In *Studies in Subjective Probability*. H.E. Kyburg and H.E. Smokler (Eds.), Krieger, New York, 2nd edition, pp. 95-158, 1980.

B. de Finetti. The theory of probability in its relationships with the analysis (Paper originally published in "Relazioni della XXVIII Riunione della Società Italiana per il Progresso delle Scienze", Pisa, 11-15 October 1939, 3, Roma, 1940, pp. 27-35). In *Probabilità e induzione*. P. Monari and D. Cocchi (Eds.), CLUEB, Bologna, pp. 365-374, 1993.

M.H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1970.

M.H. DeGroot. Changes in utility as information. *Theory and Decision*, 17:287–303, 1984.

P. Donnelly and R.D. Friedman. DNA database searches and the legal consumption of scientific evidence. *Michigan Law Review*, 97:931–984, 1999.

I.E. Dror and D. Charlton. Why experts make errors. *Journal of Forensic Identification*, 56:600–616, 2006.

I.E. Dror and G. Hampikian. Subjectivity and bias in forensic DNA mixture interpretation. *Science & Justice*, 51:204–208, 2011.

I.E. Dror and R. Rosenthal. Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences*, 53:900–903, 2008.

I.E. Dror, A.E. Péron, S. Hind, and D. Charlton. When emotions get the better of us: The effect of contextual top-down processing on matching fingerprints. *Applied Cognitive Psychology*, 19:799–809, 2005.

I.E. Dror, D. Charlton, and A.E. Péron. Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156:74–78, 2006.

I.E. Dror, C. Champod, G. Langenburg, D. Charlton, H. Hunt, and R. Rosenthal. Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a 'target' comparison. *Forensic Science International*, 208:10–17, 2011.

W. Edwards. The theory of decision making. *Psychological Bulletin*, 51:380–417, 1954.

W. Edwards. *Summing up: The society of Bayesian trial lawyers*. Kluwer Academic Publishers, Dordrecht, 1988.

W. Edwards. Influence diagrams, Bayesian imperialism, and the Collins case: an appeal to reason. *Cardozo Law Review*, 13:1025–1079, 1991 - 1992.

I.W. Evett. What is the probability that this blood came from that person? A meaningful question? *Journal of the Forensic Science Society*, 23:35–39, 1983.

I.W. Evett. A quantitative theory for interpreting transfer evidence in criminal cases. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 33:25–32, 1984.

I.W. Evett. A Bayesian approach to the problem of interpreting glass evidence in forensic science casework. *Journal of the Forensic Science Society*, 26:3–18, 1986.

I.W. Evett. On meaningful questions: a two-trace transfer problem. *Journal of the Forensic Science Society*, 27:375–381, 1987.

I.W. Evett. The theory of interpreting scientific transfer evidence. In *Forensic Science Progress, Vol. 4*. A. Maehly and R.L. Williams (Eds.), Springer-Verlag, Berlin, pp. 141–179, 1990.

I.W. Evett. Establishing the evidential value of a small quantity of material found at a crime scene. *Journal of the Forensic Science Society*, 33:83–86, 1993.

I.W. Evett. Towards a uniform framework for reporting opinions in forensic science casework. *Science & Justice*, 38:198–202, 1998.

I.W. Evett and B.S. Weir. *Interpreting DNA Evidence*. Sinauer Associates, Sunderland, 1998.

I.W. Evett, P.E. Cage, and C.G.G. Aitken. Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 36:174–180, 1987.

I.W. Evett, J.A. Lambert, and J.S. Buckleton. A Bayesian approach to interpreting footwear marks in forensic casework. *Science & Justice*, 38:241–247, 1998.

I.W. Evett, L.A. Foreman, and B.S. Weir. Letter to the editor of biometrics. *Biometrics*, 56:1274–1275, 2000a.

I.W. Evett, G. Jackson, and J.A. Lambert. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science & Justice*, 40:3–10, 2000b.

I.W. Evett, G. Jackson, J.A. Lambert, and S. McCrossan. The impact of the principles of evidence interpretation on the structure and content of statements. *Science & Justice*, 40: 233–239, 2000c.

I.W. Evett, P.D. Gill, G. Jackson, J. Whitaker, and C. Champod. Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks. *Journal of Forensic Sciences*, 47:520–530, 2002.

N. Fenton and M. Neil. Avoiding probabilistic reasoning fallacies in legal practice using Bayesian networks. *Australian Journal of Legal Philosophy*, 36:114–151, 2011.

N. Fenton, M. Neil, and D. Lagnado. Modelling mutually exclusive causes in Bayesian networks. Working paper. *http://www.eecs.qmul.ac.uk/~norman/papers/mutual_IEEE_format_version.pdf*, (last visited on 08.10.2012), 2011.

S.E. Fienberg and M.J. Schervish. The relevance of Bayesian inference for the presentation of statistical evidence and for legal decisionmaking. *Boston University Law Review*, 66: 771–798, 1986.

R. Fimmers, H. Schneider, M.P. Baur, and P.M. Schneider. Erwiderung zum Brief von Taroni et al. an den Herausgeber zum Beitrag von Schneider et al. "Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNA-Datenbank-Treffern". *Rechtsmedizin*, 21:57–60, 2011.

M.O. Finkelstein and W.B. Fairley. A Bayesian approach to identification evidence. *Harvard Law Review*, 83:489–517, 1970.

J. Fortescue. Commendation of the Laws of England, F. Grigor translation, Sweet & Maxwell, London. Digitalized version by Microsoft Corporation in cooperation with Cornell University Libraries, 2007, made available by Cornell University Library: *http://archive.org/stream/cu31924021661909#page/n65/mode/2up*, (last visited on 23.06.2013): p. 45, 1917.

R.S. Frank, S.W. Hinkley, and C.G. Hoffman. Representative sampling of drug seizures in multiple containers. *Journal of Forensic Sciences*, 36:350–357, 1991.

P. Garbolino. Explaining relevance. *Cardozo Law Review*, 22:1503–1521, 2001.

P. Garbolino and F. Taroni. Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*, 125:149–155, 2002.

P. Gill and J.S. Buckleton. Low copy number typing - Where next? *forensic Science International: Genetics Supplement Series*, 2:553–555, 2009.

P. Gill and J.S. Buckleton. A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number. *Forensic Science International: Genetics*, 4: 221–227, 2010.

P. Gill and A. Kirkham. Development of a simulation model to assess the impact of contamination in casework using STRs. *Journal of Forensic Sciences*, 49:485–491, 2004.

P. Gill, R. Sparkes, and C. Kimpton. Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International*, 89:185–197, 1997.

P. Gill, J. Whitaker, C. Flaxman, N. Brown, and J.S. Buckleton. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International*, 112:17–40, 2000.

P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, and B.S. Weir. DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International*, 160:90–101, 2006.

P. Gill, A. Kirkham, and J.M. Curran. LoComatioN: A software tool for the analysis of low copy number DNA profiles. *Forensic Science International*, 166:128–138, 2007.

P. Gill, R. Puch-Solis, and J.M. Curran. The low-template-DNA (stochastic) threshold - Its determination relative to risk analysis for national DNA databases. *Forensic Science International: Genetics*, 3:104–111, 2009.

S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni. Bayesian networks and the value of the evidence for the forensic two-trace transfer problem. *Journal of Forensic Sciences*, 57: 1199–1216, 2012a.

S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni. The database search problem: A question of rational decision making. *Forensic Science International*, 222:186–199, 2012b.

S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni. Modeling the forensic two-trace problem with Bayesian networks. *Artificial Intelligence and Law*, 21:221–252, 2013a.

S. Gittelson, A. Biedermann, S. Bozza, and F. Taroni. Decision analysis for the genotype designation in low-template-DNA profiles. *Forensic Science International: Genetics*, under revision, 2013b.

S. Gittelson, S. Bozza, A. Biedermann, and F. Taroni. Decision-theoretic reflections on processing a fingermark. *Forensic Science International*, 226:e42–e47, 2013c.

I.J. Good. Rational decisions. *Journal of the Royal Statistical Society, Series B (Methodological)*, 14:107–114, 1952.

I.J. Good. On the principle of total evidence. *The British Society for the Philosophy of Science*, 17:319–321, 1967.

M. Goray, E. Eken, R.J. Mitchell, and R.A.H. van Oorschot. Secondary DNA transfer of biological substances under varying test conditions. *Forensic Science International: Genetics*, 4:62–67, 2010.

P. J. Green and J. Mortera. Sensitivity of inferences in forensic genetics to assumptions about founding genes. *The Annals of Applied Statistics*, 3:731–763, 2009.

K.S. Grisedale and A. van Daal. Comparison of STR profiling from low template DNA extracts with and without the consensus profiling method. *Investigative Genetics*, 3, 2012.

M. Hale. Historia placitorium coronae: The history of the pleas of the crown, Volume 2, W. Stokes and E. Ingersoll (Eds.), Robert H. Small, Philadelphia. Made available by Google Books: *http://books.google.ch/books?id=gB4eAAAAMAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false*, (last visited on 23.06.2013): p. 288, 1847.

L.J. Hall and E. Player. Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Science International*, 181:36–39, 2008.

D. Hazard, P. Margot, and O. Ribaux. Pertinence de la trace: Etude théorique et perspectives expérimentales. *Revue internationale de criminologie et de police technique et scientifique*, LXIV:341–374, 2011.

A.S. Heger and J.R. Hill. Probability networks for handling uncertainty in the performance assessment of high-level nuclear waste repositories. *Reliability Engineering and System Safety*, 41:13–20, 1993.

A.S. Heger and J.E. White. Using influence diagrams for data worth analysis. *Reliability Engineering and System Safety*, 55:195–202, 1997.

I. Helsloot and J. Groenendaal. Naturalistic decision making in forensic science: Toward a better understanding of decision making by forensic team leaders. *Journal of Forensic Sciences*, 56:890–897, 2011.

A.B. Hepler and B.S. Weir. Object-oriented Bayesian networks for paternity cases with allelic dependencies. *Forensic Science International: Genetics*, 2:166–175, 2008.

A.B. Hepler, A.P. Dawid, and V. Leucari. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6:275–293, 2007.

R.A. Howard and J.E. Matheson. Influence diagrams. *Decision Analysis*, 2:127–143, 2005.

R.A. Howard, J.E. Matheson, and D.W. North. The decision to seed hurricanes. *Science*, 176:1191–1202, 1972.

G. Jackson. The scientist and the scales of justice. *Science & Justice*, 40:81–85, 2000.

R. Jeffrey. *Subjective probability – a real thing*. Cambridge University Press, Cambridge, 2004.

F. Jensen. *Bayesian Networks and Decision Graphs.* Springer-Verlag, New York, 2001.

F.V. Jensen and T.D. Nielsen. *Bayesian Networks and Decision Graphs.* Springer, New York, 2nd edition, 2007.

P. Juchli, A. Biedermann, and F. Taroni. Graphical probabilistic analysis of the combination of items of evidence. *Law, Probability and Risk*, 11:51–84, 2012.

J.B. Kadane and D.A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence.* John Wiley & Sons, New York, 1996.

J. Kaplan. Decision theory and the factfinding process. *Stanford Law Review*, 20:1065–1092, 1968.

S.M. Kassin, I.E. Dror, and J. Kukucka. The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, 2: 42–52, 2013.

D.H. Kaye. The laws of probability and the law of the land. *The University of Chicago Law Review*, 47:34–56, 1979.

D.H. Kaye. Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review*, 66:657–672, 1986.

D.H. Kaye. Introduction: What is Bayesianism? In *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism.* P. Tillers and E.D. Green (Eds.), Kluwer Academic Publishers, Dordrecht, pp. 1–19, 1988.

D.H. Kaye. Statistical decision theory and the burdens of persuasion: Completeness, generality, and utility. *International Journal of Evidence and Proof*, 1:313–315, 1997.

D.H. Kaye. Clarifying the burden of persuasion: What Bayesian decision rules do and do not do. *International Journal of Evidence and Proof*, 3:1–29, 1999.

D.H. Kaye. Rounding up the usual suspects: A legal and logical analysis of DNA trawling cases. *North Carolina Law Review*, 87:425–503, 2009.

J.H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI-83).* Karlsruhe, West Germany, 8-12 August 1983, Morgan Kaufmann Publishers, San Fransisco, pp. 190-193, 1983.

U.B. Kjaerulff and A.L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis.* Springer, New York, 2008.

D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In *Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-97).* Providence, Rhode Island, 1-3 August 1997, Morgan Kaufmann Publishers, San Francisco, pp. 302–313, 1997.

K.B. Korb and A.E. Nicholson. *Bayesian Artificial Intelligence.* CRC Press, Boca Raton, 2nd edition, 2011.

G. Langenburg, C. Champod, and P. Wertheim. Testing for potential contextual bias effects during the verification stage of the ACE-V methodology when conducting fingerprint comparisons. *Journal of Forensic Sciences*, 54:571–582, 2009.

B.P. Lathi. *Modern Digital and Analog Communication Systems.* Oxford University Press, New York, 3rd edition, 1998: pp. 682–683.

S.L. Lauritzen and A. Mazumder. Informativeness of genetic markers for forensic inference– an information theoretic approach. *Forensic Science International: Genetics Supplement Series*, 1:652–653, 2008.

R.O. Lempert. Modeling relevance. *Michigan Law Review*, 75:1021–1057, 1977.

Y. Lengwiler. The origins of expected utility theory. In *Vinzenz Bronzin's Option Pricing Models: Exposition and Appraisal*. H. Zimmermann and W. Hafner (Eds.), Springer-Verlag, New York, pp. 535–546, 2009.

D.V. Lindley. Probabilities and the law. In *Utility, probability and human decision making*. D. Wendt and C.J. Wlek (Eds.), Reidel, Dordrecht, pp. 223–232, 1975.

D.V. Lindley. Costs and utilities. In *Decision Making and Medical Care*. F.T. de Dombal and F. Gremy (Eds.), North-Holland Publishing Company, Amsterdam, pp. 101–112, 1976.

D.V. Lindley. Probability and the law. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 26:203–220, 1977a.

D.V. Lindley. A problem in forensic science. *Biometrika*, 64:207–213, 1977b.

D.V. Lindley. The Bayesian approach. *Scandinavian Journal of Statistics*, 5:1–26, 1978.

D.V. Lindley. *Making Decisions*. John Wiley & Sons, Chichester, 2nd edition, 1985.

D.V. Lindley. The choice of sample size. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 46:129–138, 1997.

D.V. Lindley. The philosophy of statistics. *The Statistician*, 49:293–337, 2000.

D.V. Lindley. *Understanding Uncertainty*. John Wiley & Sons, Hoboken, 2006.

E. Locard. *L'enquête Criminelle*. Traité de criminalistique. Tome septième, Livre VIII. Desvigne, Lyon, 1940.

P. Lucas, H. Boot, and B. Taal. Computer-based decision support in the management of primary gastric non-hodgkin lymphoma. *Methods of Information in Medicine*, 37:206–219, 1998.

P. Lucas, N. de Bruijn, K. Schurink, and A. Hoepelman. A probabilistic and decision-theoretic approach to the management of infectious disease at the ICU. *Artificial Intelligence in Medicine*, 19:251–279, 2000.

A. Mazumder. *Planning in Forensic DNA Identification Using Probabilistic Expert Systems*. PhD thesis, Department of Statistics, University of Oxford, Oxford, 2010.

R. Meester and M. Sjerps. The evidential value in the DNA database search controversy and the two-stain problem. *Biometrics*, 59:727–732, 2003.

R. Meester and M. Sjerps. Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. *Law, Probability and Risk*, 3:51–62, 2004a.

R. Meester and M. Sjerps. Response to dawid, balding, triggs and buckleton (concerning: Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence. law, prob. and risk 3:51-62). *Law, Probability and Risk*, 3:83–86, 2004b.

A.A. Mitchell, J. Tamariz, K. O'Connell, N. Ducasse, Z. Budimlija, M. Prinz, and T. Caragine. Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics*, 6:749–761, 2012.

J. Mortera, A.P. Dawid, and S.L. Lauritzen. Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, 63:191–205, 2003.

National Research Council (NRC) Committee on DNA Forensic Science: An Update. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington D.C., 1996.

National Research Council (NRC) Committee on DNA Technology in Forensic Science. *DNA Technology in Forensic Science*. National Academy Press, Washington D.C., 1992.

National Research Council (NRC) of the National Academies. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington D.C., 2009.

R.F. Nease and D.K. Owens. Use of influence diagrams to structure medical decisions. *Medical Decision Making*, 17:263–276, 1997.

M. Neil, N. Fenton, and L. Nielson. Building large-scale Bayesian networks. *Knowl. Eng. Rev.*, 15:257–284, 2000. 975727.

C. Neumann, I.W. Evett, J.E. Skerrett, and I. Mateos-Garcia. Quantitative assessment of evidential weight for a fingerprint comparison I. generalisation to the comparison of a mark with set of ten prints from a suspect. *Forensic Science International*, 207:101–105, 2011a.

C. Neumann, I. Mateos-Garcia, G. Langenburg, J. Kostroski, J.E. Skerrett, and M. Koolean. Operational benefits and challenges of the use of fingerprint statistical models: A field study. *Forensic Science International*, 212:32–46, 2011b.

A. Nordgaard and B. Rasmusson. The likelihood ratio as value of evidence – more than a question of numbers. *Law, Probability and Risk*, 11:303–315, 2012.

D.K. Owens, R.D. Shachter, and Jr. R.F. Nease. Representation and analysis of medical decision problems with influence diagrams. *Medical Decision Making*, 17:241–262, 1997.

G. Parmigiani. *Modeling in Medical Decision Making: A Bayesian Approach*. John Wiley & Sons, Chichester, 2002.

G. Parmigiani and L. Inoue. *Decision Theory: Principles and Approaches*. John Wiley & Sons, Chichester, 2009.

M. Peterson. *An Introduction to Decision Theory*. Cambridge University Press, Cambridge, 2009.

S. Petricevic, J. Whitaker, J.S. Buckleton, S. Vintiner, J. Patel, P. Simon, H. Ferraby, W. Hermiz, and A. Russell. Validation and development of interpretation guidelines for low copy number (LCN) DNA profiling in New Zealand using the AmpFlSTR® SGM Plus™ multiplex. *Forensic Science International: Genetics*, 4:305–310, 2010.

V.L. Phillips, M.J. Saks, and J.L. Peterson. The application of signal detection theory to decision-making in forensic science. *Journal of Forensic Sciences*, 46:294–308, 2001.

J.W. Pratt, H. Raiffa, and R. Schlaifer. The foundations of decision under uncertainty: An elementary exposition. *Journal of the American Statistical Association*, 59:353–375, 1964.

S.J. Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. John Wiley & Sons, Hoboken, 2nd edition, 2003.

R. Puch-Solis, P. Roberts, S. Pope, and C.G.G. Aitken. *Assessing the Probative Value of DNA Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*. Royal Statistical Society' Working Group on Statistics and the Law, London, 2012.

H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Harvard University Press, Boston, 1961.

D. Ramos and J. Gonzalez-Rodriguez. Cross-entropy analysis of the information in forensic speaker recognition. In *Proceedings of Odyssey*. Stellenbosch, South Africa, 21-24 January 2008, ISCA Archive, *http://www.isca-speech.org/archive_open/odyssey_2008* (last visited on 25.06.2013), 2008.

D. Ramos Castro. *Forensic Evaluation of the Evidence Using Automatic Speaker Recognition Systems*. PhD thesis, Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, 2007.

F.P. Ramsey. Truth and probability (1926). In *The Foundations of Mathematics and other Logical Essays, Chapter VII*. R.B. Braithwaite (Ed.), Kegan, Paul, Trench, Trubner & Co., London, pp. 156–198 (1999 electronic edition), 1931.

M. Redmayne. *Expert Evidence and Criminal Justice*. Oxford University Press, New York, 2001.

S. Renooij and L.C. van der Gaag. Decision making in qualitative influence diagrams. In *Proceedings of the Eleventh International FLAIRS Conference*. Sanibel Island, Florida, 18-20 May 1998. D.J. Cook (Ed.), AAAI Press, Menlo Park, pp. 410-414, 1998.

O. Ribaux and P. Margot. Case based reasoning in criminal intelligence using forensic case data. *Science & Justice*, 43:135–143, 2003.

B. Robertson and G.A. Vignaux. Probability - The logic of the law. *Oxford Journal of Legal Studies*, 13:457–478, 1993.

B. Robertson and G.A. Vignaux. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*. John Wiley & Sons, Chichester, 1995a.

B. Robertson and G.A. Vignaux. DNA evidence: wrong answers or wrong questions? In *Human Identification: The Use of DNA Markers*. B.S. Weir (Ed.), Kluwer Academic Publishers, Dordrecht, pp. 145–152, 1995b.

M.J. Saks and J.J. Koehler. What DNA "fingerprinting" can teach the law about the rest of forensic science. *Cardozo Law Review*, 13:361–372, 1991.

L.J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46:55–67, 1951.

L.J. Savage. *The Foundations of Statistics*. Dover Publications, New York, 2nd edition, 1972.

P.M. Schneider, H. Schneider, R. Fimmers, W. Keil, G. Molsberger, W. Pflug, T. Rothmel, M. Eckert, H. Pfeiffer, and B. Brinkmann. Allgemeine Empfehlungen der Spurenkommission zur statistischen Bewertung von DNA-Datenbank-Treffern. *Rechtsmedizin*, 20: 111–115, 2010.

D.A. Schum. *The Evidential Foundations of Probabilistic Reasoning*. John Wiley & Sons, New York, 1994.

Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST). Standards for examining friction ridge impressions and resulting conclusions, version 1.0, 13.09.2011. *http://www.swgfast.org/documents/examinations-conclusions/111026_Examinations-Conclusions_1.0.pdf*, (last visited on 10.07.2012), 2011.

R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.

R.D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36: 589–604, 1988.

Q. Shen, J. Keppens, C.G.G. Aitken, B. Schafer, and M. Lee. A scenario-driven decision support system for serious crime investigation. *Law, Probability and Risk*, 5:87–117, 2006.

A. Stockmarr. Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search. *Biometrics*, 55:671–677, 1999.

D.A. Stoney. What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society*, 31:197–199, 1991a.

D.A. Stoney. Transfer evidence. In *The Use of Statistics in Forensic Science*. C.G.G. Aitken and D.A. Stoney (Eds.), Ellis Horwood, New York, pp. 107–138, 1991b.

D.A. Stoney. Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *Journal of the Forensic Science Society*, 34:17–21, 1994.

P. Taberlet, S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L.P. Waits, and J. Bouvet. Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, 24:3189–3194, 1996.

F. Taroni, C. Champod, and P. Margot. Forerunners of Bayesianism in early forensic science. *Jurimetrics Journal*, 38:183–200, 1998.

F. Taroni, C.G.G. Aitken, and P. Garbolino. De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. *Science & Justice*, 41:145–150, 2001.

F. Taroni, J.A. Lambert, L.Fereday, and D.J. Werrett. Evaluation and presentation of forensic DNA evidence in European laboratories. *Science & Justice*, 42:21–28, 2002.

F. Taroni, A. Biedermann, P. Garbolino, and C.G.G. Aitken. A general approach to Bayesian networks for the interpretation of evidence. *Forensic Science International*, 139:5–16, 2004.

F. Taroni, S. Bozza, and C.G.G. Aitken. Decision analysis in forensic science. *Journal of Forensic Sciences*, 50:894–905, 2005.

F. Taroni, C.G.G. Aitken, P. Garbolino, and A. Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, Chichester, 2006a.

F. Taroni, S. Bozza, and A. Biedermann. Two items of evidence, no putative source: an inference problem in forensic intelligence. *Journal of Forensic Sciences*, 51:1350–1361, 2006b.

F. Taroni, S. Bozza, M. Bernard, and C. Champod. Value of DNA tests: A decision perspective. *Journal of Forensic Sciences*, 52:31–39, 2007.

F. Taroni, S. Bozza, A. Biedermann, P. Garbolino, and C.G.G. Aitken. *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. John Wiley & Sons, Chichester, 2010.

F. Taroni, A. Biedermann, R. Coquoz, T. Hicks, and C. Champod. Letter to the editor with reference to schneider et al. "allgemeine empfehlungen der spurenkommission zur statistischen bewertung von DNA-datenbank-treffern" ("recommendations of the german stain commission regarding the statistical evaluation of matches following searches in the national DNA database"). *Rechtsmedizin*, 21:55–57, 2011.

F. Taroni, M. Sjerps, and C. Berger. The strength of evidential interpretation: making connections. In *Sixth European Academy of Forensic Science Conference: Towards Forensic Science 2.0*. The Hague, 20-24 August 2012, 2012.

W.C. Thompson, F. Taroni, and C.G.G. Aitken. How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences*, 48:47–54, 2003.

C.M. Triggs and J.S. Buckleton. The two trace transfer problem re-examined. *Science & Justice*, 43:127–134, 2003.

T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, and N. Morling. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. *Forensic Science International: Genetics*, 3:222–226, 2009.

T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, and N. Morling. Evaluating the weight of evidence by using quantitative short tandem repeat data in DNA mixtures. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 59:855–874, 2010.

D. Tzidony and M. Ravreby. A statistical approach to drug sampling: A case study. *Journal of Forensic Sciences*, 37:1541–1549, 1992.

B.T. Ulery, R. A. Hicklin, J. Buscaglia, and M.A. Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, 7:1–12, 2012.

J. Vatn, P. Hokstad, and L. Bodsberg. An overall model for maintenance optimization. *Reliability Engineering and System Safety*, 51:241–257, 1996.

P. Vicard, A.P. Dawid, J. Mortera, and S.L. Lauritzen. Estimating mutation rates from paternity casework. *Forensic Science International: Genetics*, 2:9–18, 2008.

J. von Neumann and O. Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 2nd edition, 1947.

S.R. Watson and R.V. Brown. The valuation of decision analysis. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 141:69–78, 1978.

B.S. Weir. Statistical analysis. In *Encyclopedia of Forensic Science*. J. Siegel, P. Saukko and G. Knupfer (Eds.), Academic Press, San Diego, pp. 545–550, 2000.

J.P. Whitaker, E.A. Cotton, and P. Gill. A comparison of the characteristics of profiles produced with the AMPFISTER® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International*, 123:215–223, 2001.

G. Whitman and R. Koppl. Rational bias in forensic science. *Law, Probability and Risk*, 9: 69–90, 2010.