Year : 2016

COMBINING EVIDENCE

Juchli Patrick

UNIL | Université de Lausanne
Faculté de droit,
des sciences criminelles
et d'administration publique

COMBINING EVIDENCE

PHD THESIS of **Patrick Osamu Juchli** established under the supervision of **Professor Franco Taroni**

Lausanne, 2016

ISBN 2-940098-74-3

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 M. Patrick Juchli, candidat au doctorat en science forensique, intitulée

« Combining evidence »

Le Président du Jury

Professeur Marcelo Aebi

Lausanne, le 3 juin 2016

ABSTRACT. The goal of the present thesis consists of establishing the normative foundations for reasoning about combined evidence. Unlike the interpretation of single items of evidence, little is known about inference tasks involving multiple items of evidence. In forensic practice, however, experts are regularly confronted with a collection of evidence rather than isolated evidence items. This necessarily raises the question on how to interpret evidence holistically. The study of the relationships between the different evidence items in a collection and between a collection and a common cause (represented as hypotheses), is of central concern for this thesis. Such relationships and causes are almost always unobservable in judicial contexts, and therefore, inherently uncertain. Indeed, uncertainty is a fundamental feature of reasoning about evidence. The framework for handling uncertainty is defined by probability theory. Evidential reasoning is consequentially a form of probabilistic reasoning. The present thesis locates itself in this probabilistic framework and puts a strong emphasis on graphical probabilistic modeling.

The thesis is composed of four cornerstones for each of which a paper was produced. Throughout this thesis, the ordering of the cornerstones is thematic and not chronological. The first paper examines the different types of evidence and their combinations, their generic inference structures, and the relationships between these different inference structures. The examination establishes, thus, a probabilistic ontology of evidence. The following study illustrates the application of generic inference structures in two real forensic cases. One case involves the combination of two features of a single footwear mark. The other involves fingermarks and a footwear mark, thus two distinct marks. The study shows that even apparently simple forms of combinations involve evidential subtleties that require careful analysis. The third study provides novel analysis methods for evidential phenomena exclusively occurring in combined evidence. To date, there are only a few methods for assessing the inferential interactions between items of evidence in a holistic setting. This study addressed this problem. The final project consists of a complex case analysis involving four different DNA specimens collected from a rape case that lead to a wrongful conviction of a young man. The model treats each specimen as a mixture profile, and includes considerations on the relevance of each specimen, the possible number of contributors to each specimen, the inferential relationships between the specimens, as well as between the specimens and the hypothesis about the authorship of the crime. As it turned out, the different specimens were subject to strong inferential interactions – a fact that was completely missed by the expert of the case. This thesis shows: the problems pervading the subject of combined evidence are not academic phantoms; they are measurable, real, and can affect the lives of people for better or worse.

RESUME. L'objectif de cette thèse consiste à établir les fondements normatifs pour raisonner sur la combinaison des éléments de preuves (indices). Contrairement à leur interprétation en isolation, peu est connu sur les inférences impliquant plusieurs indices. Dans la pratique forensique, cependant, les experts sont régulièrement confrontés à une collection d'indices plutôt que isolés. Cela soulève nécessairement la question sur la façon d'interpréter les indices de manière holistique. L'étude des relations entre les différents indices dans une collection et entre une collection et une cause commune (représentée par des hypothèses), est une préoccupation centrale pour cette thèse. Ces relations et ces causes sont largement inobservable dans des contextes judiciaires, et par conséquent, de caractères incertaine. En effet, l'incertitude est une caractéristique fondamentale du raisonnement sur les indices. Le cadre pour gérer l'incertitude est définie par la théorie des probabilités. Le raisonnement sur les indices est alors une forme de raisonnement probabiliste. La présente thèse se situe dans ce cadre probabiliste et met l'accent sur la modélisation probabiliste graphique.

La thèse se compose de quatre jalons majeurs pour chacun desquels un article a été produit. Tout au long de cette thèse, les jalons sont reproduits en ordre thématique et non chronologique. Le premier article examine les différents types d'indices et leurs combinaisons, leurs structures d'inférence génériques, et les relations entre ces différentes structures d'inférence. L'étude établit, par conséquent, une ontologie probabiliste des indices. L'étude qui suit illustre l'application des structures d'inférence génériques dans deux cas forensiques réels. Un cas implique la combinaison de deux caractéristiques d'une trace de semelle. L'autre cas implique des traces digitales et une trace de semelle, c'est-à-dire deux marques distinctes. L'étude montre que même des combinaisons simple produisent des subtilités nécessitant une analyse minutieuse. La troisième étude fournit de nouvelles méthodes d'analyse pour des phénomènes se produisant exclusivement dans les combinaisons d'indices. Jusqu'à ce jour, il n'y a que quelques méthodes d'évaluation des interactions inférentielles entre les indices dans un cadre holistique. L'étude en question a abordé ce problème. Le projet final consiste en une analyse d'un cas complexe impliquant quatre échantillons d'ADN différents provenant d'une affaire de viol qui menait à la condamnation erronée d'un jeune homme. Le modèle traite chaque échantillon comme un profil de mélange, et comprend des considérations sur la pertinence de chaque échantillon, le nombre possible de collaborateurs de chaque échantillon, les relations inférentielles entre les spécimens, ainsi qu'entre les spécimens et l'hypothèse de la paternité de le crime. En fait, les différents échantillons sont soumis à de fortes interactions inférentielles - un fait qui a été complètement manqué par l'expert de l'affaire.

La thèse montre: les problèmes qui envahissent le sujet de la combinaison des éléments de preuves ne sont pas des fantômes académiques; au contraire, ils sont mesurables, réel, et peuvent affecter la vie des gens pour le meilleur ou le pire.

Acknowledgements

The present thesis was established at the School of Criminal Justice of the University of Lausanne under the supervision of Professor Dr. Franco Taroni, professor in forensic science. Further, the dissertation committee consisted of Professor Dr. Alex Biedermann, professor in forensic science at the School of Criminal Justice of the University of Lausanne, Professor Dr. Julia Mortera, professor at the department of economics of the Università degli Studi Roma Tre, Dr. Ian Evett, Expert in the Interpretation of Scientific Evidence at Principal Forensic Services Ltd. The committee was presided by Professor Dr. Marcelo Aebi, professor in criminology at the School of Criminal Justice of the University of Lausanne. The thesis was partly supported by the Swiss National Science Foundation (grant number 100014_150276).

Hereby, I would like to express my deep gratitude to Professor Dr. Franco Taroni, my thesis supervisor, for his valuable guidance and critiques, as well as his unflinching support during our five years of collaboration. I would also like to thank Professor Dr. Biedermann for all his assistance, proofreading, and the innumerable advices. I would also like to extend my thanks to Professor Dr. Julia Mortera and Dr. Ian Evett, for taking part in the expert committee and for their useful suggestions and comments regarding the present thesis. I express my gratitude to Professor Dr. Marcelo Aebi for presiding the expert committee. I wish to thank my colleagues, especially Giulia Cereda and Dr. Simone Gittelson for their lending a helping hand at various instances. I would also like to thank the Swiss National Science Foundation for its financial support during the last two years and the School of Criminal Justice of the University of Lausanne for having provided me with such a welcoming place to study and to do research. Finally, I extend my gratitude to my family and friends, who have supported me all along, and of course, to all the persons important to me.

Contents

List of Figures

List of Tables

Outline of the thesis

What needs to be said about the author's research on the combination of items of evidence from a 'scientific' perspective, can be found in the articles that were compiled during the last five years. You will find the articles in Part III of the present manuscript. The articles produced are listed and presented in thematic order.

- JUCHLI, P. The evidential foundations of probabilistic reasoning: toward a better understanding of evidence and its usage. *Frontiers in Genetics* 4 (2013), doi: 10.3389/fgene.2013.00164.
- JUCHLI, P., BIEDERMANN, A., AND TARONI, F. A probabilistic ontology of evidence and its combinations. Submitted to *Artificial Intelligence and Law* in December 2015.
- JUCHLI, P., BIEDERMANN, A., AND TARONI, F. Graphical probabilistic analysis of the combination of items of evidence. *Law, Probability and Risk* 11 (2012), 51-84.
- JUCHLI, P. AND TARONI, F. Investigating evidential phenomena in combined evidence. Submitted to *Artificial Intelligence* in December 2015.
- JUCHLI, P., AND TARONI, F. Aggregating DNA evidence: A probabilistic analysis of the DNA evidence in *State of Texas vs Josiah Sutton*. To be submitted to *International Journal of Approximate Reasoning*.

Each article listed above is reproduced in the chapters from 1 to 5 respectively.

Part I and II are meant to present you with a personal account and a complementary view on the subject of combining evidence so as to put the articles into the author's perspective. For this purpose, the author takes the liberty to use a personalistic narrative perspective in the form of the first person singular. As you may have noticed, the author will also address you, the reader, directly.

Part I discusses the fundamental concepts of evidential reasoning. The introductory chapter (Chapter 1) outlines the context, objective, and methodological framework of the present thesis. Chapter 2 outlines the basic properties of what we call 'evidence', in particular, the relevance and credibility aspect of evidence. This is accomplished by examining the question of what evidence is (Section 2.1), followed by a discussion on what types of evidence there are when focusing on inferential properties (Section 2.2). Chapter 3 is devoted to the uncertainty aspect of evidence and how it can be measured by probabilities. This discourse starts with the notion of conditional independence and its relationship to personal probabilities (Section 3.1), followed by outlining the importance of the Bayes' theorem for evidential reasoning (Section 3.2), and an explanation on the metrics the inferential force of evidence employed in this thesis (Section 3.3). This thesis uses two metrics for the inferential force. These are the *likelihood ratio* (Subsection 3.3.1) and the *weight of evidence* (Subsection 3.3.2). Chapter 4 is a remark. It discloses my standpoint from which a given text was redacted. You may feel this chapter to be somewhat out of bounds for this thesis. However, please accept this as my personal stance against the 'objective' self-negation of our current scientific tradition. This part closes with Chapter 5 and deals with a generic argument of evidence. Section 5.1 outlines constitutive elements of such an argument. A fundamental property of an argument of evidence is the *drag* *coefficient*, a quantity that encapsulates the uncertainty accumulation in such an argument. A general account of the drag coefficient and the impact of weak and rare events on the drag coefficient is given in Section 5.2.

Part II deals with the study of the combination of evidence items. Each chapter outlines and discusses the results presented in the series of the papers. Namely, Chapter 6 deals with our (i.e., Franco Taroni, Alex Biedermann and I) ontological study of evidence. This study is approached by asking in what way evidence exists (Section 6.1). Section 6.2 presents an excerpt of our ontology for combined evidence. The chapter closes with the discussion on *atomism* and *holism* as it pertains to the study of combined evidence (Section 6.3). Chapter 7 portraits our study of combined evidence in two forensic cases. This involves the examination of the impact of inferential interactions between evidence items on the inferential force (Section 7.1). The relationship between the hypotheses and the evidence items for the combination of evidence is discussed in Section 7.2. Chapter 8 discusses the methods developed for the examination of large bodies of evidence. In a first step we discuss how a measure for inferential interactions can be derived from the multiplication law of probability (Section 8.1) and how it can be extended to masses of evidence. The measure for the degree of *dissonance* in evidence is explained in Section 8.2, as well as the questions arising from dissonant evidence. Section 8.3 outlines the relative contribution measure for evidence and its general tendencies in large bodies of evidence. Section 8.4 closes the chapter with general questions about large bodies of evidence that remain yet to be addressed. Chapter 9 discusses the case assessment and interpretation on *State of Texas vs Josiah Sutton*. It involves the aggregation of multiple DNA typing results of potential mixture profiles. Section 9.1 discusses the main procedures we followed for the aggregation. A combination of evidence relies heavily on additional information, such as circumstantial information or expert knowledge. This has important implications for the evaluator and his models (Section 9.2). The inferences suggested by the model are discussed in Section 9.3. The main conclusions are given in Chapter 10. This chapter closes the second part of the thesis by contrasting the study of combined evidence with the general state of forensic practice.

Part I

EVIDENCE

1 Introduction

> 'Indeed this is the nature of all arguments, for what is certain cannot be proved by what is uncertain.'
>
> Quintilian, Institutio Oratoria, 5.10.8.

The Science of Evidence, a notion coined (or promoted?) by D. A. Schum [126], considers evidence to be a subject of systematic research. The endeavor spans from the examination of evidential properties, the discovery of evidence, its interpretation and uses, to the communication of evidence-related findings, all the while embracing a generalized approach. The expression 'generalized approach' means also that the Science of Evidence is a multidisciplinary activity, which is characterized by an inclusive nature. The aim is not to accumulate methods and techniques from different disciplines in order to build a large toolbox for solving problems. It is to build an integrated understanding of evidence by looking at different disciplines and not the other way around. The spirit of this science is, therefore, essentially the same as Forensic Science,which only stresses its close ties to the justice and security establishment.

The study of combined evidence is a particular excercitation within the vast discipline of the Science of Evidence or Forensic Science. Its object is the extension of the evidence evaluation from an atomistic to a holistic interpretative framework. This was, to my best knowledge, first explicitly stated by J. H. Wigmore in 1937:

> 'What is wanted is simple enough in purpose, – namely, some method which will enable us to lift into consciousness and to state in words the reasons why a total mass of evidence does or should persuade us to a given conclusion, and why our conclusion would or should have been different or identical if some part of that total mass of evidence had been different. The mind *is* moved; then can we not explain why it is moved? If we can set down and work out a mathematical equation, why can we not set down and work out a mental probative equation?' [146, p. 4]

Almost 80 years later, however, the problem largely remains as indicated by P. Dawid in 2011:

> 'Modern technology provides for the collection and manipulation of vast quantities of data of many different kinds, and a new and all-pervading field of intellectual and practical activity, 'Information Technology', has sprung up to support these data-handling requirements. In contrast, relatively little attention has been payed to the issues of combining, comparing, linking and – most important – interpreting all these data, so turning them from *information* into *evidence*.' [29, p.1]

The present thesis explores the normative foundations of evidence combination in Forensic Science.

I pursued this object within a probabilistic framework. It is widely recognized in law [e.g., 93, 100, 121] as well as in forensic science [e.g., 4, 44, 118], that the interpretation of evidence is best operated within this framework. Today, there are no other normative systems that can measure up to the richness and versatility of probability theory to capture evidential subtleties and complexities [126]. Since the seminal work of J. Pearl and others [e.g., 92, 108], graphical probabilistic models (i.e. Bayesian networks) have assisted human inference in various domains (see [86, 112] for an overview). Bayesian networks have also proven to be greatly useful in forensic and judicial contexts [e.g., 11, 133]. I relied heavily on the tool of graphical probabilistic models to explore the normative foundation of combined evidence. Note that the thesis provides information about probability theory and Bayesian networks only on a need-to-know basis. If you wish to know more about these subjects, you may consult the general literature on this topic. Namely, on the subject of probability theory you may find the sources [e.g., 35, 97, 122] helpful, and on the subject of Bayesian networks [e.g. 77, 83, 133]. Further sources are cited at different locations throughout this thesis.

2 Relevance and credibility

In the end, every observation that prompts us to contemplate about its significance is essentially an item of evidence (for something). The use of evidence is deeply engrained in humans, and even more so, in a trained forensic scientist, such as myself. This favors the content of a particular perception to be closely tied to a particular conclusion. That is, the understanding of evidence grows based on particular instances, but not the other way around. Therefore, it comes more naturally to us to examine the relationship between, say, 'the green color of a tomato' and 'the ripeness of the fruit', or 'the correspondence between a fingermark and fingerprint' and 'a common individual', than to examine the relationship between 'the green tomato and the ripeness of the fruit' and 'the correspondence between mark and print, and a common individual'. However, a serious contemplation about evidence must bring to light the latter more obscure relationships. A good starting point is to begin by examining naïve, general questions and to further the study depending on the answers obtained by such questions. It seems almost natural to initiate the discussion by asking what 'evidence' actually is (Section 2.1). The answers provided by such a questions will allow us to draw further distinctions among what we call 'evidence' (Section 2.2).

2.1 What is evidence?

An insight that turned out to be very helpful in understanding the meaning of evidence is presented in the present section. This insight can be found in different forms in [1, 69, 125].

Let us start by setting up the following two assertions, namely,

First

Every item of evidence is a thing[1]; 'The thing' is the generic term for 'the evidence'.

Second

There exist things that are 'not evidence'.

From the first assertion it follows that items of evidence regularly possess at least one property in common with all the things that are not items of evidence. From the second statement it follows that items of evidence regularly possess at least one property that the other things do not have. For the former we shall use the expression *the general property of things*, and for the latter *the particular property of evidence*. Thus, every item of evidence possesses a general property of things and a particular property of evidence. The general property of things is the reference to itself. Ideally, this reference is tautological since the thing introduces itself to our mind on the basis of an identity statement. That is, thing A states 'thing A', or in short $A : A$. For example, a package containing a substance resembling narcotics actually contains narcotics, a fingermark is actually a fingermark, and so on. However, this reference is not always tautological, that is, either thing B states 'thing A' ($B : A$), or thing A states

[1] The 'thing' as it is understood here, can also be immaterial, such as a verbal statement from a person, a noise, or even the absence of something.

'thing B' ($A : B$). For example, a suspicious person is selling a substance as narcotics, but it turns out to be flour; or a pattern resembling a fingermark is found at the crime scene, however it turned out to be an artifact ($B : A$), and so on. In another case, one may find a batch of toy figures that contain narcotics, or a counterfeit fingermark ($A : B$). An enquiry of the first reference involves the assessment of the *credibility* of the thing, or more precisely, the credibility of the identity statement of the thing. Later we will see that the credibility of a thing involves different intricate aspects.

However, an item of evidence possesses a second reference to something else. That is, they 'point beyond themselves' as I. Hacking says [69]. This reference is the particular property of evidence. As can be seen, this reference is precisely *not* tautological but expresses a relationship of a thing to something else, that is to some $\overline{A}$.[2] Based on such a relationship we are inclined to say A points at C, or A indicates C, where C is a $\overline{A}$. This indicative property is also what qualifies the evidence as a *sign*. Consider the following excerpt from Quintilian's *Institutio Oratoria*:

> 'Such signs or indications enable us to infer that something else has happened; blood for instance may lead us to infer that a murder has taken place. But bloodstains on a garment may be the result of the slaying of a victim at a sacrifice or of bleeding at the nose. Everyone who has a bloodstain on his clothes is not necessarily a murderer.' [115, 5.9.9]

In other words, it may happen that we take a thing to be an indication (sign) for a C, whereas actually it is an indication for something else D. Hence, the enquiry on whether C is indicated or D pertains to the *relevance* of the thing, but not to its credibility (e.g., bloodstain) for an issue of interest (e.g., murder).

Let C stand for the issue of interest (e.g., the violation of a narcotics act). Generally, you would aspire to the ideal situation, where the item is entirely credible and relevant: $A : A$ (identity) and A implies C. In this situation the evidence is *conclusive*. However, if the thing is entirely relevant but not credible, then the evidence is worthless for our enquiry. More precisely, you would have $B : A$ and A implies C, but since you assume that A, which is in fact B, your thing is actually saying nothing about A implying C (e.g., what you thought was narcotics was actually flour. However, the possession of flour is not regulated by the narcotics act). In turn, if the thing is entirely credible but not relevant, then the evidence is also worthless. In such situations you would have, $A : A$ but A implies D instead of C (e.g., you find narcotics delivery car, but as it turns out, the delivery is destined for a legal research project of a student at the School of Criminal Justice at the University of Lausanne). In short:

- An item of evidence possesses a reference to itself and to something else.
- An examination of the former reference amounts to assessing the *credibility of the evidence*.
- An examination of the latter reference amounts to assessing the *relevance of the evidence*.
- A meaningful item of evidence possesses at least some credibility and some relevance, otherwise it bears no meaning at all.
- Hence, both the credibility of the evidence as well as its relevance need to be assessed for any item of evidence.

You can find additional information regarding the credibility and the relevance of the evidence from a slightly different perspective in a book review (see Part III Chapter 1). Let me emphasize at this enjuncture that I did not intend to provide a general and clearcut definition of evidence, but only what I termed as a 'helpful insight'. The question of what evidence is, was, and continues to be, bestowed with various views in different disciplines by different scholars (see for example [8]). To my knowledge a general and clear-cut definition of evidence does not exist. However, all views seem to gravitate around the issues relative to the credibility and/or the relevance of evidence.

[2] The overscore or overline stands for a negation of the term it scores over. Thus, $\overline{A}$ designates a negated A ('not A').

2.2 A substance-blind approach to evidence

The first paragraph of Section 2.1 pointed out the difficulty in focusing on relationships between items of evidence that completely differ in their perceptual content or substance (i.e., 'green color of a tomato' and 'correspondence between fingermark and fingerprint'). Later, I identified two references that every item of evidence possesses. Each reference is associated with the credibility and the relevance of the evidence respectively. The ponderosity of this insight arises from its applicability to all items of evidence irrespective of their substances.

Given that every item of evidence possesses an aspect of relevance and credibility, these so-called *credentials of evidence* provide ideal dimensions, along which the investigation and classification of evidence can be achieved unhinged from its substance. Based on the credibility and relevance of evidence, Schum [125, 126] established a classification of evidence. It is in the context of his classification of evidence that a particular intellectual attitude becomes most evident. That is, the investigation of evidential properties on an inferential basis rather than evidential substance. Schum calls this intellectual attitude 'a substance-blind approach to evidence':

> 'Investigators in the fields of law, history, and intelligence analysis must be prepared to evaluate evidence having any conceivable substance. The problem then is: How can we ever say anything general about evidence if it comes in so many substantive varieties? One answer is provided if we choose to ignore its substance and focus instead on its inferential properties.' [6, p. 72]

In order to explain the substance-blind approach Schum [126] refers to a passage in J. H. Poincaré's 'La valeur de la science':

> 'Maintenant qu'est-ce que la science? [...], c'est avant tout une classification, une façon de rapprocher des faits que les apparences séparaient, bien qu'ils fussent liés par quelque parenté naturelle et caché. La science, en d'autres termes, est un système de relations.' [111, p. 265]

> 'Now what is science? [...] it is before all a classification, a manner of bringing together facts which appearances separate, though they were bound together by some natural and hidden kinship. Science, in other words, is a system of relations.' [110, p. 137]

It was through the focus on relations among items of evidence that 'are bound together by some natural and hidden kinship', that allowed Schum to gain a deeper insight into different forms and combinations of evidence [126]. By doing so, Schum created a substance-blind classification[3] of evidence, which is shown in Figure 2.1 [6, 125, 126]. Schum's classification shows different combinations between dimensions of credibility and relevance.

The relevance dimension contains two categories: directly and indirectly relevant evidence. An item of evidence is said to be *directly relevant*, if it is conclusive on some issue of interest when the item is completely credible. If an item remains inconclusive on the issue of interest, even if it is completely credible, then the item is said to be *indirectly relevant*. An evidence item is said to be *ancillary*, if it serves to assess the relevance of another item of evidence. Stated otherwise, ancillary evidence is evidence about some other evidence.

The credibility dimension contains four major categories: tangible, testimonial, and missing evidence, as well as accepted facts. *Tangible evidence* refers to physical objects and their derivations, for example, documents, sensor records, images, charts, measurements, and so on. An item of evidence is *testimonial*, if it is provided by a person. It is *unequivocal* in cases, in which a person asserts that some event occurred (e.g., an eyewitness states: 'I clearly saw Mr. X breaking into my neighbour's flat'). In contrast, a testimony is *equivocal*, if a witness is not certain whether some event occurred (e.g., an eyewitness states 'I am not quite sure, but I believe rather strongly to have seen Mr. X breaking into my neighbour's flat'). In some cases a testimony might be even *completely equivocal* (e.g., an eyewitness states 'I don't know at all whether Mr. X did break into this flat' or 'I can't remember what happened'). Both, tangible evidence and testimonial evidence can be *positive* (+) or *negative* (-). An evidence item is positive if the occurrence of an event is reported (e.g., 'a fingermark was found on the door knob'), and negative

[3] The words 'categorization', 'taxonomy', or 'ontology' are used interchangeably with 'classification'.

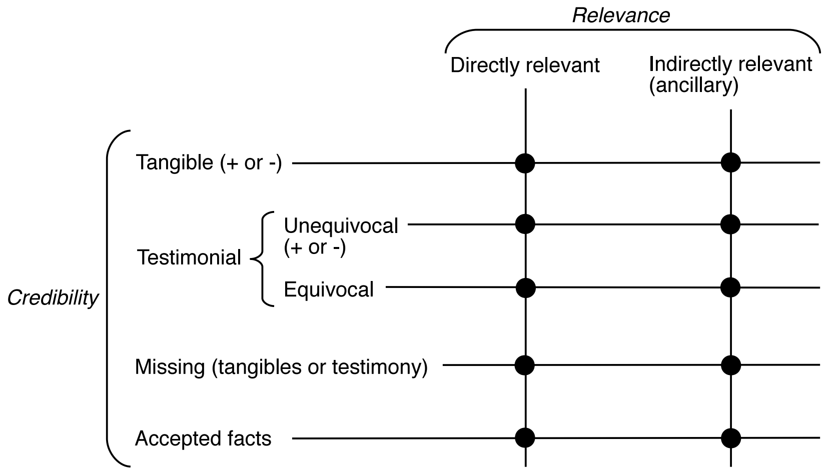


Figure. 2.1 – Classification of evidence according to Schum's substance blind approach

otherwise (e.g. 'no fingermark was found on the door knob'). An item of evidence can also be *missing*. That is, a particular item of evidence was expected to be found, but was not produced (e.g. 'Mr. X must have grabbed the door knob, however there was no fingermark'). Note that negative and missing evidence are not the same. The former refers to the absence of evidence, and the latter to the evidence of absence. Finally, *accepted facts* are evidential instances, for which the credibility is considered to be given (e.g. 'Lausanne is situated at the northern shore of the Lac Léman (i.e., Lake Geneva)').

How an item of evidence is best classified depends on the situation at hand and the reasoner's perspective. The classification of an item of evidence is, therefore, *context-dependent* [6]. For a forensic scientist, for instance, the correspondence between a fingermark and a reference print may be an item of tangible evidence. For a lawyer, who receives an expert testimony by the forensic scientist, however, the same correspondence may represent testimonial evidence. In another case, especially in cases of evidence combination, the same evidence may be used twice for different inferences and, thus possibly, represent two types of evidence. For example, an item of tangible evidence can serve, at the same time, as an item of ancillary evidence for an other item (e.g., the bloodstains on garment for inferring the identity of a murder victim) and as an item that is directly relevant for an issue of interest (e.g., the bloodstains for inferring a stabbing of the victim by the suspect).

3 Uncertainty about evidence

In Section 2.1 I have explained that an item of evidence is not necessarily credible or relevant. On this occasion you might have asked yourself: How do I know whether an item of evidence is credible or relevant? The answer is: Most of the time you don't; most of the time you will be *uncertain* about an item's credibility and relevance. There is, thus, no choice left but to exploit the experienced uncertainty as well as possible. A widely accepted method to exploit uncertainty is the theory of probability, where probabilities are considered as a measure of the uncertainty you experience relative to some event, or otherwise stated as your degree of belief that some event applies. Any given probability is, therefore, an expression of how you relate to the world: 'It is not solely a feature of your mind, it is not a value possessed by an event but expresses a relationship between you and the event and is a basic tool in your understanding of the world.' [97, p. 38] If the measure of probability is understood in such a way, then the probability is said to be 'subjective' [37], or 'personal' [97, 122].

3.1 Conditional independence, and personal probabilities

In order to better understand what 'subjectivity' means regarding the theory of probability, it is important to become acquainted with the fundamental concept *conditional independence*. A probability function denoted by $Pr(\cdot)$ is required to be coherent, in the sense that they adhere to the laws of probability theory [97]. Let e_1 and e_2 denote two events. Assume further that we had knowledge about a third event e_3. If we believe that event e_2 does not affect our belief about e_1 given our knowledge of e_3, then the events e_1 and e_2 are said to be *conditionally independent* given e_3. Stated more formally, one has $Pr(e_1 \mid e_2, e_3) = Pr(e_1 \mid e_3)$ and conversely $Pr(e_2 \mid e_1, e_3) = Pr(e_2 \mid e_3)$. Thus, we can write $Pr(e_1, e_2, e_3) = Pr(e_1 \mid e_3)Pr(e_2 \mid e_3)Pr(e_3)$. If, however, we believe that e_1 affects the probability of e_2 given e_3, then one has $Pr(e_1 \mid e_2, e_3)$ and conversely $Pr(e_2 \mid e_1, e_3)$. At this point it is important to note that the conditional (in)dependence relationship a symmetric relationship. That is, if e_1 is conditionally (in)dependent of e_2 given e_3, then e_2 is independent of e_1 given e_3.

Now, we have asserted that a probability expresses the relationship between us and the world. We are never in a state of complete ignorance; there is always something we know at a given time. Thus, everyone of us has an individual 'knowledge base' or 'current state of knowledge'. This is denoted K and represents, thus, everything a reasoner knows [37].[1] From a personalistic point of view any given probability can only exist relative to some K. Thus, following a strict notation, probabilities such as $Pr(e_2)$ or $Pr(e_1 \mid e_2)$ are not entirely correct, because this would mean that some e_1 or e_2 possess a probability on their own. However, since the probability expresses a relationship between us and the world, some e_1 or e_2 obviously cannot possess a probability. Instead we should write $Pr(e_2 \mid K)$ or $Pr(e_1 \mid e_2, K)$, because 'the' probability is always my or your probability of something, *given* my or your current state of knowledge K. That is, any given probability is necessarily conditioned by some K and represents therefore a *personal probability*. [2]

[1] This knowledge is often called 'background knowledge'.

[2] For clarity, I use $Pr(e \mid K)$ as a short for $Pr(E = e \mid K)$. Also for clarity purposes the K is sometimes omitted all together. Note also that in

3.2 Bayes' theorem

Let e denote some evidence. Further, let $H = \{h_p, h_d\}$ stand for a set of hypotheses (propositions). In most forensic applications we distinguish between a prosecution hypothesis h_p and a defense hypothesis h_d.

At the very beginning of any evidential investigation a reasoner, such as you or I, *is* in a given state of knowledge. When the reasoner has fixed some hypotheses H and starts to contemplate about them, he will eventually relate the hypotheses to his current state of knowledge. At this point he forms his initial beliefs regarding the hypotheses. His belief is then translated into a probability for each hypothesis so that $Pr(H \mid K) = Pr(h_p \mid K) + Pr(h_d \mid K) = 1$ applies. These probabilities are called '*prior* probabilities' because we assign them to h_p and h_d prior to the observation of an evidential fact. Next, imagine that the reasoner obtains an item of evidence, say, e for the moment. This means that everything that is not e (i.e. $\bar{e}$) is eliminated from our consideration so that now $Pr(H \mid e, K) = Pr(h_p \mid e, K) + Pr(h_d \mid e, K) = 1$ holds. At this point let us focus on h_p. The joint probability of e, h_p is given by

$$Pr(e, h_p \mid K) = Pr(e \mid K)Pr(h_p \mid e, K) = Pr(h_p \mid K)Pr(e \mid h_p, K). \quad (3.1)$$

Next, assume that the reasoner were interested in the probability of h_p, given that e occurred (i.e., $Pr(h_p \mid e, K)$). How should he accomplish the transformation from the *prior* probability $Pr(h_p \mid K)$ to the *posterior* probability $Pr(h_p \mid e, K)$? Bayes' theorem or Bayes' rule, which follows from Equation (3.1), suggests an answer to this question, notably

$$Pr(h_p \mid e, K) = \frac{Pr(e, h_p \mid K)}{Pr(e \mid K)} = \frac{Pr(e \mid h_p, K)Pr(h_p \mid K)}{Pr(e \mid K)}, \quad (3.2)$$

where $Pr(e \mid K) \neq 0$. This probability, also called 'normalization constant' in this context, serves to redress the proportions of $Pr(e, h_p \mid K)$ relative to $Pr(e \mid K)$ in order to assure that $Pr(H \mid e, K) = Pr(h_p \mid e, K) + Pr(h_d \mid e, K) = 1$ holds. The probability $Pr(e \mid h_p, K)$ is sometimes referred to as the (Bayesian) *likelihood*[3]. This probability is an important constituent for the measures of the inferential force.

3.3 Inferential force

The notion of *inferential force* is a crucial feature of evidence-based reasoning. It stems from the question of how strong one or more items of evidence are.

> 'In evaluating evidence or 'weighing' evidence in an inferential task, one recognizes that items of evidence differ in strength; for various reasons some items are persuasive and allow for substantial revision in our opinions, while other items seem to justify little or no opinion revision. Thus, a major task in inductive inference consists of evaluating the inferential or probative strength of evidence.' [127, p.107]

In fact, together with the properties of credibility and relevance, these three elements of evidence are called *the three credentials of evidence* by D. Schum [125]. The most common measures for the inferential force are the *likelihood ratio* and *the weight of evidence*. These measures not only allow us to compare different items of evidence regarding their respective strength but also to evaluate the combined strength of multiple items. These measures are presented in the following two subsections (3.3.1 and 3.3.2).

some papers of this thesis the knowledge base is denoted as I.

[3] Namely, it is a probability regarding its dependence on e but a likelihood regarding its dependence on h_p [97]. Thus, it is the probability of e given h_p but the likelihood of h_p given e.

3.3.1 Likelihood ratio

There is now wide agreement in forensic science to measure the inferential force of evidence by the likelihood ratio (LR) [4, 118]. Say we were interested in the inferential force exerted by event e on our hypotheses H. When taking the ratio of Bayes' theorem for h_p and for h_d (see Equation 3.2) we obtain the odds form of Bayes' theorem. The LR appears in the odds form of the Bayes' theorem:

$$\underbrace{\frac{Pr(h_p \mid e,K)}{Pr(h_d \mid e,K)}}_{\text{posterior odds}} = \underbrace{\frac{Pr(e \mid h_p,K)}{Pr(e \mid h_d,K)}}_{\text{LR}} \underbrace{\frac{Pr(h_p \mid K)}{Pr(h_d \mid K)}}_{\text{prior odds}}. \tag{3.3}$$

The prior odds refer to the odds of the hypotheses prior to our observing e, and the posterior odds those after receiving e. The ratio at the center of Equation (3.3) is the likelihood ratio of h_p given e.[4] It conveys succinctly if, and to what extent, e generates inferential force on H. Thus, the LR tells us how much we can learn from e regarding our hypotheses, or stated otherwise, the impact of e on the relative magnitudes of the prior odds and the posterior odds. This view can be emphasized by writing the LR in terms of the prior odds and the posterior odds. For this purpose, let $O(h_p \mid e,K)$ denote the posterior odds and $O(h_p \mid K)$ the prior odds. By reformulating Equation 3.3 we have

$$\text{LR} = \frac{Pr(e \mid h_p,K)}{Pr(e \mid h_d,K)} = \frac{O(h_p \mid e,K)}{O(h_p \mid K)}. \tag{3.4}$$

The values of the LR range from 0 to ∞. The value 0 implies that e is impossible given h_p and K. Conversely, the value ∞ implies that e is impossible given h_d and K. If the LR is larger than one, then e supports the hypothesis in the numerator (here h_p) over the one in the denominator (h_d). If the LR is exactly one, then e has no inferential force and we learn that e is equally probable under h_p and h_d. In such cases the posterior odds and the prior odds possess exactly the same value. If the LR is smaller than one, then e supports the hypothesis in the denominator over the one in the numerator. Note that the LR is sometimes called the Bayes factor (BF), or the Bayes-Turing-Jeffreys-factor. More precisely, if the evaluation involves only simple hypotheses[5], the BF reduces to the LR. In contrast, if the evaluation involves composite hypotheses, then the inferential force of the item depends on the assessments you made on the priors and is, therefore, measured by the BF (for further explanations see for instance [136]). Composite hypotheses occur typically in contexts of multiple hypotheses. In such cases, it is advisable to revert to the original form of Bayes' theorem shown in Equation (3.2). However, for the remaining part of this thesis you will not need to know how to work with multiple hypotheses. For further information on this subject you are referred to [9, 58] for a general account and to [18, 133, 135] for forensic applications.

Finally, suppose we were confronted with a set of items, say $\mathbf{E} = \{e_1, e_2, ..., e_n\}$, where a given item is denoted as $e_i \in \mathbf{E}$. In the general form, the LR is given by

$$\frac{Pr(\mathbf{E} \mid h_p,K)}{Pr(\mathbf{E} \mid h_d,K)} = \left[\prod_{i=2}^{n} \frac{Pr(e_i \mid e_1,e_2,...,e_{i-1},h_p,K)}{Pr(e_i \mid e_1,e_2,...,e_{i-1},h_d,K)}\right] \frac{Pr(e_1 \mid h_p,K)}{Pr(e_1 \mid h_d,K)}. \tag{3.5}$$

If the items are further independent given H (i.e. e_i is independent e_j for all $i \neq j$), then this reduces to

$$\frac{Pr(\mathbf{E} \mid h_p,K)}{Pr(\mathbf{E} \mid h_d,K)} = \prod_{i=1}^{n} \frac{Pr(e_i \mid h_p,K)}{Pr(e_i \mid h_d,K)}, \tag{3.6}$$

[4] The articles reproduced in Part III use different notations for the LR depending on the general framework in which the paper was written. The following notations for the LRs are all equivalent: $Pr(e \mid h_p,K)/Pr(e \mid h_d,K) \equiv V_{e|H,K} \equiv LR_{e|H,K} \equiv F(h_p : e \mid K)$.

[5] A hypothesis is said to be simple, if all the parameters of the distribution are specified. Otherwise, the hypothesis is said to be complex.

and if each item possesses the same inferential force, then we can simplify even further

$$\frac{Pr(\mathbf{E} \mid h_p, K)}{Pr(\mathbf{E} \mid h_d, K)} = \left[\frac{Pr(e_i \mid h_p, K)}{Pr(e_i \mid h_d, K)}\right]^n. \tag{3.7}$$

3.3.2 Weight of evidence

The weight of evidence (WoE) is a concept that I. J. Good seems to have appreciated because of its intuitive appeal and versatility. This is reflected in his many works [64]. The WoE corresponds to the logarithm of the BF, or the LR respectively. This makes WoE an additive and symmetrical measure of the inferential force. Following the naming of the events in the previous subsection, the WoE of the LR in Equation 3.3 is denoted $W(h_p : e \mid K)$, and read as 'the weight of evidence in favor of h_p as compared to h_d provided by e and given K' [e.g., 58, 60, 65]. If the logarithm is at the base of 10, then WoE is measured in units of *ban*s (abbreviated by b), where a tenth of a ban is a *deciban* (abbreviated by db)[6]. One deciban is the smallest unit of WoE perceptible to human judgement and corresponds to a BF of roughly 5/4 (i.e., $\log_{10}(5/4) \approx 0.097$) [65]. If the logarithmic base is 2, then the unit is in *bit*s, in which case the same unit is used for information and for the WoE [63].

Let $O(h_p \mid e, K)$ denote the posterior odds and $O(h_p \mid K)$ the prior odds. By taking the logarithm of Equation 3.3, the odds and the LR become additive

$$\log O(h_p \mid e, K) = W(h_p : e \mid K) + \log O(h_p \mid K), \tag{3.8}$$

Thus, the weight of evidence is essentially the difference between the log posterior odds and the log prior odds ($W(h_p : e \mid K) = \log O(h_p \mid e, K) - \log O(h_p \mid K)$). In general, the WoE can take a value between $-\infty$ and ∞. A value of 0 implies that e provides no weight at all (which corresponds to LR $= 1$). If the WoE takes a value smaller than 0, then e provides weight in favor of h_d (i.e., LR < 1). If its value is larger than 0, then e provides weight in favor of h_p (i.e., LR > 1). You can see that, as opposed to the LR, the values of the WoE are symmetric regarding 0.

When dealing with multiple items of evidence $\mathbf{E} = \{e_1, e_2, ..., e_n,\}$, the Equations (3.5) to (3.7) are also additive. Namely, in general we have

$$W(h_p : \mathbf{E} \mid K) = W(h_p : e_1 \mid K) + \sum_{i=2}^{n} W(h_p : e_i \mid e_1, e_2, ..., e_{i-1}, K). \tag{3.9}$$

If the items are independent given H, then (3.9) simplifies to

$$W(h_p : \mathbf{E} \mid K) = \sum_{i=1}^{n} W(h_p : e_i \mid K), \tag{3.10}$$

and if all the items possess the same weight, then we can simplify even further

$$W(h_p : \mathbf{E} \mid K) = n \times W(h_p : e_i \mid K). \tag{3.11}$$

The properties that are exploited in this thesis are the additivity and the symmetry of the WoE. However, I invite you to keep in mind that the WoE has many other interesting properties lending themselves to a more concise interpretation and, therefore, to a better understanding of what this measure of inferential force means. For example, the WoE can also be written in terms of information (usually measured in bit) and can be translated into

[6] The name stems from the English town of Banbury, where I. J. Good was part of A. Turing cryptanalysis team during the Second World War. According to I. J. Good, it was Turing who proposed the name ban for the unit of WoE in dependence on Banbury [63].

communication theory [66]. The *mutual information* between h_p and e (or the *logarithm of the association factor* between H and e) is given by

$$I(h_p : e \mid K) = \log \frac{Pr(e, h_p \mid K)}{Pr(e \mid K) Pr(h_p \mid K)} = \log \frac{Pr(e \mid h_p, K)}{Pr(e \mid K)} = \log \frac{Pr(h_p \mid e, K)}{Pr(h_p \mid K)} = I(e : h_p \mid K). \tag{3.12}$$

Thus, the WoE can also be written as the difference between the mutual information between h_p and e and between h_d and e

$$W(h_p : e \mid K) = I(h_p : e \mid K) - I(h_d : e \mid K) = I(e : h_p \mid K) - I(e : h_d \mid K). \tag{3.13}$$

I. J. Good showed that the mutual information between (the proposition expressing) the evidence and (the one expressing) the hypothesis corresponds to the explanatory power of the hypothesis for the evidence, whereas the weight of evidence corresponds to the degree of corroboration [60]. Hence, according to Equation (3.13), the WoE expresses essentially the difference between the explanatory power of two competing hypotheses regarding some evidence. For further information on this subject regarding its theoretical aspects you are referred to [e.g., 61, 95] and for different practical applications to [66].

4 Notes on subjectivity and objectivity

The dualism of subjectivity and objectivity is one that caused me a great deal of confusion that lasted for quite some time. I personally find this dual distinction arduous when used in epistemic (e.g., scientific) contexts, especially in expressions such as 'objective knowledge' or 'objective fact'. By reading scientific literature it was often not clear to me, in what sense these words were used, and what intention an author pursued by using them. This distinction has ample connotations from apologists of different philosophical drifts, most of whom hold some absolute claim towards a given scientific practice. For example, 'the scientist must exclusively focus on value-free, and hence, objective facts!'[1], or 'everything is subjective, and thus, relative!'. This renders the discourse on uncertainty unnecessarily sluggish and dogmatic. Do not get me wrong, I agree with de Finetti's view on science, or knowledge in general [2]:

> 'Il y a toujours une infinité d'explications possibles pour un même groupe d'observations: si nous en choisissons une, et si nous énonçons une loi, ce ne pourra être que pour des raisons subjectives qui nous la font considérer digne de confiance.' [33, p. 64]

> 'There are always an infinite number of possible explanations for a given group of observations: if we chose one (explanation) and assert a law, then this is only due to subjective reasons, which make them seem trustworthy.' (free translation)

Thus, the following problem arises: subjectivity is defined in opposition to objectivity, and vice versa. If we fully accept subjectivity, which seems reasonable to me, it makes no sense to think in categories of subjectivity versus objectivity, since the latter does not exist as such, and consequently neither does the former. This renders every discourses held within the boundaries of subjectivity and objectivity largely meaningless. It is just as de Finetti stated a 'total inadequacy of the present language' [34, p.139]. In this context he provided the quote by Sir Harold Jeffreys, which is repeated here:

> 'ordinary language has been created by realists, and mostly very naïve ones... We have enormous possibilities of describing the inferred properties of objects, but very meagre ones of describing the directly known ones of sensation... The idealist must either do his best with realist language or make a new one, and not much has been done in the latter direction' [74, p.124]

The question is, thus, whether it is possible to find words that fit such a 'subjective' world view. For this purpose, consider again de Finetti's understanding of science. This argument can be very confusing, because at first glance, this argumentation truly seems to imply the subjectivity of these so-called 'positive facts', viz. of everything I become aware of.[3] However, it makes less sense at second glance, and even less at the third. Following a rigorous reading, de Finetti's statement is itself voiced from a subjective position (which is again voiced from a subjective

[1] The NAS report of 2009 makes recurrent use of such statements, for example, 'The goal is to make scientific investigations as objective as possible so the results do not depend on the investigator.' [105, ch.4 p.10]

[2] See also I. W. Evett: '... a so-called 'objective test' can only exist within a framework of assumptions and the validity of those assumptions in an individual case is a matter for *subjective* judgement. The objectivity is an illusion.' [42, p.119]

[3] A similar argumentation can be found in de Finetti's philosophical manifest Probabilism [36], where he cites Adriano Thilgher: 'All objects,

position, ...). We seem to be hung up in an infinite regress[4]. Clearly, the act of thinking produces the *judgement* of subjectivity (of the reasons that make a thing trustworthy). Indeed, what else could? However, does that mean that the act of thinking that produces a judgement can simultaneously presuppose this very judgement[5]?
J. G. Fichte showed how such an impasse is caused by an erroneous understanding of our self-awareness [50]. For one: 'think about a dice!'. Notice how you become immediately aware of your thinking. Thus, this awareness is neither accidental nor accessory, but it is something intrinsically tied to your act of thought. If you are immediately aware of yourself, then this self-awareness can only be described in the way that

'... deine innere Tätigkeit, die auf etwas ausser ihr (auf das Object des Denkens) geht, geht zugleich in sich selbst, und auf sich selbst. Aber durch in sich zurückgehende Thätigkeit entsteht uns [...] das Ich. Du warst sonach in deinem Denken deiner selbst dir bewusst, und dieses Selbstbewusstseyn eben war jenes unmittelbare Bewusstseyn deines Denkens; sey es, dass ein Object, oder dass du selbst gedacht wurdest. – Also das Selbstbetwusstsein ist unmittelbar; in ihm ist Subjectives und Objectives unzertrennlich vereinigt und absolut Eins.' [50, p. 528]

'..your inner activity that points at something beyond itself (upon the object of thought), points at the same time into itself and at itself. The I arises from the activity returning into itself. Hence, in your act of thinking you were aware of yourself, and this self-awareness is precisely that immediate awareness of your act of thinking; (irrespective of) whether it is an object or yourself that was thought of. – Thus, the self-awareness is immediate; in it, the subjective and the objective are inseparably unified and absolute one.' (Free translation)

This view suggests that the distinction between subject and object (and thus, subjectivity and objectivity) is imposed by the particular condition of our mind during the act of thinking (i.e., 'your inner activity'). This distinction, however, cannot refer to an actual division of our minds from the world. Hence, this necessarily raises the question of whether such a distinction has any relevance for science in general, and for forensic science and law in particular. I argue that it is not. Consider again H. Poincaré's words on science quoted in the previous section, namely, that science 'is a manner of bringing together facts which appearances separate, though they were bound together by some natural and hidden kinship'. For H. Poincaré objectivity – if such can be found – can only be found in these relationships, because only these are transmissible or transferable by discourse from one mind to another. Stated otherwise, these relationships are *intelligible*. Thus, H. Poincaré later writes:

'Les objets extétieurs, par exemple, pour le lesquels le mot *objet* a été inventé, sont justement des *objets* et non des apparences fuyantes et insaisissables parce que ce ne sont pas seulement des groupes de sensations, mais des groupes cimentés par un lien constant. C'est ce lien, et ce lien seul qui est objet en eux, et ce lien c'est un rapport.' [111, p. 266]

'External objects, for instance, for which the word *object* was invented, are really *objects* and not fleeting and fugitive appearances, because they are not only groups of sensations, but groups cemented by a constant bond. It is this bond, and this bond alone, which is the object in itself, and this bond is a relation.' [110, p. 137]

men, and things of which I speak are, in the last analysis, only the content of my present act of thought: the very statement that they exist outside and independently of me is an act of my thought: I CAN ONLY THINK OF THEM AS INDEPENDENT OF ME BY THINKING THEM, *I.E.*, MAKING THEM DEPENDENT ON ME.' [36, p. 171]

[4] In a similar context J. G. Fichte states: 'Aber um deines Denkens dir bewusst zu seyn, musst du deiner selbst dir bewusst seyn. –*Du bist– deiner* dir bewusst, sagst du; du unterscheidest sonach nothwendig dein *denkendes* Ich von dem im Denken desselben *gedachten* Ich. Aber damit du dies könnest, muss abermals das Denkende in jenem Denken *Object* eines höheren Denkens seyn, um Object des Bewusstseyns seyn zu können; und du erhältst zugleich ein neues *Subject*, welches dessen, das vorhin das Sellbstbewusstlseyn war, sich wieder bewusst sey. [...] und nachem wir einmal nach diesem Gesetze fortzuschliessen angefangen haben, kannst du mir nirgends eine Stelle nachweisen, wo wir aufhören sollten;...' [50, p. 526] In my free translation: 'However, for you to be aware of your thinking, you have to be aware of yourself. You say, – *you are* – aware *of you*; in this act of thought, therefore, you distinguish necessarily between your *thinking* the I and the I *that is thought*. For you to be able to do so, however, that which is thinking in your (present) thought must be anew an *object* of some higher act of thinking in order to become an object of awareness; at the same time you obtain a new *subject* so that you are again aware of that, which formerly was the self-awareness. [...] once engaged in such a scheme of reasoning, you cannot provide me with a point, at which we should halt.'

[5] Note that neither de Finetti's texts 'La prévision' or 'Probabilism' explicitly imply such a presupposition [33, 36]. However, they neither prevent it.

That is to say, these relationships are *intelligible* (i.e., they can be understood), because they can be thought. Only what is intelligible is potentially transferable, and only what is transferable can be reproduced by another mind. This is the reason why H. Poincaré writes that only what can be reproduced by another mind holds the prerequisite for objectiveness. Thus, from my pragmatic perspective, it is secondary whether something is subjective or objective. What we should care about is whether something can be reproduced in a mind other than my own and adopted. Stated otherwise, the real question is not about whether something is objective or subjective, absolutely true or relatively true; it is about *persuasiveness*. More precisely, whether it is generally persuasive, or only for a particular individual. In the latter case, that something is intimate. If, however, it is something that everyone can check for himself, or that can be transferred to a conscious and 'sane' mind by discourse, then it is common. Everything else, is somewhere in between these two extremes at varying degrees of what is personal.

5 Argument of evidence

Section 2.1 explained the distinction between the credibility of the evidence and its relevance. Section 3.3 discussed the inferential force of evidence. These three features of evidence constitute the 'three credentials of evidence' [125]. For any given item of evidence not only do we have to assess these three features, but also we also have to bring them together coherently (i.e., by obeying the laws of probability). How this is accomplished is the subject of the present section following the teachings of D. A. Schum and A. W. Martin [125, 127].

Most of the time, it seems overly simplistic to reach the hypotheses from some evidential fact (e.g., observation or testimony) in a single inference step. Usually, we will find several sources of uncertainty introducing themselves between some hypotheses and fact. This means that inferences have to be made step by step from one source of uncertainty to the other until we reach the hypotheses. These kinds of inferences are called *simple cascaded inferences* [125] or *catenated inferences* [146]. The structures of such inferences can be represented by *chains of reasoning* of different lengths as shown in Figure 5.1.

5.1 Argument structure

The fundamental inference structure of an item of evidence must address (at least) the credibility and the relevance of the item. If the credibility and the relevance contain a single source of uncertainty each, then we obtain a two-staged chain of reasoning. This reasoning chain corresponds to the graphical probabilistic model (i.e. Bayesian network, or BN) in Figure 5.1 a. Let me quickly explain what these structures are exactly. Each node represents a variable such as in our case H, E, or R, where $H = \{h_p, h_d\}$ denotes our hypotheses, $E = \{e, \bar{e}\}$ some event possibly relevant for our hypotheses, and $R = \{r, \bar{r}\}$ the report of the event (which can be an observation or a testimony). The arrows (or arcs) codify a conditioning order, and express therefore the conditional dependence relationship between these variables (see Section 3.1). In the present case, we have R conditioned by E, which in turn is conditioned by H. In order to choose a given conditioning order for the construction of an inference structure, it is helpful to reason from cause to effect [83, 120] [1]; that is, from the hypotheses to what is finally reported. This is the information conveyed by the graphical structure of these models.

The probabilistic structure of these models is not directly visible, with the exception of the conditional dependence relationship between the variables. However, to each of these variables we assign a specific probability table (so called *node probability tables*, or NPT) that codify the probabilities given to each variable. These are shown in the Tables 5.1 (a) to (c). The graphical aspect codifies the dependence relationship. The node probability table

[1] Note, however, that causality is not the same concept as conditional dependence. Consider this beautiful example by P. Lipton: 'Numerous sticks are in free fall, twisting and tumbling. A snapshot is taken before any of them hits the ground, and what is found is that at that moment more of the sticks were near the horizontal than near the vertical. [...] Think of a single stick with a fixed midpoint. There are only two ways it could be vertical − pointing up or pointing down − but many ways it could be horizontal − any orientation in the horizontal plane. More sticks were near the horizontal because of the geometrical fact that there are more ways for sticks to be near the horizontal. But geometrical facts are not causes.' [98, p. 21] However, the probability of your finding an object in free fall in the horizontal position rather than in the vertical position is clearly conditioned by the geometric fact presented in the object.

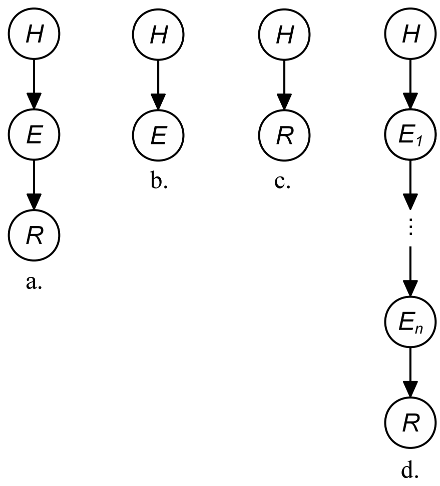


Figure. 5.1 – Simple cascaded inferences. a. Generic argument of evidence (two-stage); b. Argument of evidence with no uncertainty regarding the credibility of the item (single-stage); c. Argument of evidence with no uncertainty regarding the relevance of the item (single-stage) d. argument of evidence for multiple inference stages (n-Stage).

codifies the probabilities for each variable. Together they form a graphical probabilistic model, or in the present thesis, a Bayesian network.

Table. 5.1 – Probability tables for the nodes H, E, and R for the reasoning chain of Figure 5.1 a.

(a)

H:	h_p	$Pr(h_p \mid K)$
	h_d	$Pr(h_p \mid K)$

(b)

H:		h_p	h_d
E:	e	a_1	b_1
	$\bar{e}$	$1-a_1$	$1-b_1$

(c)

E:		e	$\bar{e}$
R:	r	a_2	b_2
	$\bar{r}$	$1-a_2$	$1-b_2$

Let me note a few things in preparation for the following sections. The link in the chain representing the inference step, also called 'reasoning stage' or simply 'stage', from R to E is the argument of credibility and that from E to H is the argument of relevance. Together they form the argument of evidence. The inferential force of a report r in the form of an LR is given by

$$\frac{Pr(r \mid h_p, K)}{Pr(r \mid h_d, K)} = \frac{Pr(r \mid e, K)Pr(e \mid h_p, K) + Pr(r \mid \bar{e}, K)Pr(\bar{e} \mid h_p, K)}{Pr(r \mid e, K)Pr(e \mid h_d, K) + Pr(r \mid \bar{e}, K)Pr(\bar{e} \mid h_d, K)} = \frac{a_2 a_1 + b_2(1-a_1)}{a_2 b_1 + b_2(1-b_1)}.$$

Assuming that the reports appear perfectly credible to us, that is, they represent no source of uncertainty to us, then we have $a_2 = 1$ and $b_2 = 0$ and the LR reduces to a_1/b_1. This situation is depicted in Figure 5.1 b. Conversely, if the event seems perfectly relevant to us, that is $a_1 = 1$ and $b_1 = 0$, then the LR reduces to a_2/b_2.

5.2 Drag coefficient

Often you will encounter items of evidence, for which you identify n sources of uncertainty between a report and the hypotheses. The chain of reasoning depicted in Figure 5.1 d. corresponds to this situation. A forensic example of such a multistage cascaded inference can be found in [15, 16]. In such situations we have $a_1, a_2, ..., a_n$ and $b_1, b_2, ..., b_n$ (and the corresponding complement) probabilities involved, namely, for each of the n stages. Let a_i and b_i denote the probabilities of the ith stage out of n stages. The general form of the LR is given by

$$\frac{Pr(r \mid h_p, K)}{Pr(r \mid h_d, K)} = \frac{a_1 + D_n}{b_1 + D_n} = \frac{a_1 + D_{n-1} + \dfrac{b_n}{\prod_{i=2}^{n}(a_i - b_i)}}{b_1 + D_{n-1} + \dfrac{b_n}{\prod_{i=2}^{n}(a_i - b_i)}}, \tag{5.1}$$

where D_n is the *drag coefficient* (see also Section 3.2.4 of III.3). It encapsulates the inferential drag produced by additional stages (of uncertainties). In general, the larger the value of D_n, irrespective of its sign, the weaker the inferential force. Moreover, positive values of D_n imply that r favors h_p, whereas for negative values r favors h_d. From Equation 5.1 you can gather that the uncertainties are recursively accumulated along the stages of the chain. It further shows that the amount of drag accumulated at a given stage is determined by two things: the differences of likelihoods a_i and b_i *and* the size of b_i. This is significant: the ratio between likelihoods is not sufficient; the values of the likelihoods themselves are equally important. In other words, this Bayesian mechanism incorporates the 'rareness or improbability' of the events of each stage [125]. Note also that the difference between the likelihoods at any stage decreases and the drag, in turn, increases. This means that the inferential force decreases with every additional reasoning stage. Thus, the longer a reasoning chain, the weaker the inferential force.

Weak events and rare events. Let us take a closer look at how weak and rare events influence the drag. A weak event exists if the likelihoods a_i and b_i take values of similar magnitude. A chain of reasoning involving an event, for which the corresponding likelihoods take the same value ($a_i = b_i$), generates no inferential force, because this event produces the maximal drag of ∞. For example, at the second stage such an event produces $D_2 = b_2/0 = \infty$, and at the third $D_3 = b_2/(a_2 - b_2) + b_3/[(a_2 - b_2) \times 0] = \infty$, and so on.[2] Hence, the closer the values of a_i and b_i, the smaller their difference $(a_i - b_i)$, and thus, the larger the drag and the smaller the inferential force. Moreover, there is no mechanism in a reasoning chain to compensate a drag once accumulated. On the contrary, the difference between the likelihoods at any stage decreases as mentioned before (cf. the end of (the previous) Section5.2). Thus, for any simple chain of reasoning the following rule holds: irrespective of the weak link's location, a chain cannot be stronger than its weakest link [125].

A rare event exists if the likelihoods a_i and b_i are both small. This means that the difference $(a_i - b_i)$ is also small. Does that mean that a rare event is a weak link? The answer is no, because for rare events it matters where they are located in the chain. As mentioned in the previous section, the drag depends on the difference between the likelihoods as well as on their ratio. Rare and weak events might produce similar effects regarding the difference between the likelihoods, however, the effect is quite different for their ratios. In fact, as D. A. Schum showed in [125], this becomes clear when we factor out the last term involving a_n and b_n:

$$D_n = D_{n-1} + \frac{b_n}{\prod_{i=2}^{n}(a_i - b_i)} = D_{n-1} + \frac{b_n}{\prod_{i=2}^{n-1}(a_i - b_i)} \left[\frac{b_n}{a_n - b_n}\right] = D_{n-1} + \frac{b_n}{\prod_{i=2}^{n-1}(a_i - b_i)} \left[\frac{a_n}{b_n} - 1\right]^{-1}.$$

We see that for the nth stage (the one on the bottom of the chain) the difference between the likelihoods plays no role. Only the ratio matters. In other words, at the nth stage it does not matter whether we have, say $a_n = 0.9$ and

[2] Throughout the thesis $\lim_{x \to 0} 1/x = 0$ for a division by zero is adopted.

$b_n = 0.01$ or some rare event with $a_n = 0.009$ and $b_n = 0.0001$. In both cases the ratio is 90. However, in any other reasoning stage the difference comes into play. In which case the first pair of likelihoods yields 0.89, but the second 0.0089. Thus, if a rare event is not located at the nth stage, then it has the same effect as a weak event.

Part II

COMBINING EVIDENCE

'Celui qui a entendu la même chose de 12'000 témoins oculaires a seulement 12'000 probabilités, ce qui équivaut à une forte probabilité, ce qui est loin d'être certain.'

Voltaire, 1694 - 1778

In Part I we have seen what an argument of evidence is, and how it serves to connect a particular report to the hypotheses of interest. During this discussion it became clear that between a report and the hypotheses we may feel compelled to introduce several sources of uncertainty creating a chain of reasoning for simple cascaded inferences. Graphically speaking, this operation amounts to a vertical expansion of our inference. In this part, however, we examine how arguments of evidence are combined. That is, we examine the relationships between different lines of arguments in order to realize a horizontal expansion of the argument. Such argument structures codify so called *complex cascaded inferences* as opposed to *simple cascaded inferences* [127], which I discussed in the previous chapter.

Note further that evidence can be mixed or not. Evidence is said to be *mixed* if it contains evidence items of different types [146]. A collection of multiple items of evidence associated with particular problem or a case is called a 'body of evidence'. Large bodies of evidence are also named masses of evidence. As long as it is not explicitly indicated in this text, bodies of evidence and masses of evidence can be mixed or not.

6 An ontological study of evidence

The notion of ontology is often delimited by, or seen in opposition to, the notion of phenomenology, which is the systematic study of experiences or phenomena. An *ontology*, on the other hand, is the systematic study of things that exist, including their corresponding relationships. Thus, the paper does not investigate how we experience evidence, but what kinds of evidence there exist and how they relate to each other. You may say at this point: Since we study evidence from a perspective of subjective probability, every item of evidence is nothing but an intellectual experience. Given this fact, I agree that the distinction between ontology and phenomenology is not quite clear. Note, however, that it is also true that in our daily lives we experience evidence as facts that exist. Moreover, we also experience differences between items of evidence, and make distinctions based on the differences we experienced. From a pragmatic perspective it seems natural to accept these experiences as facts. More precisely, we simply assumed that evidence exists. Now, given this assumption, the next question would be: in what ways or how does it exist? What I found by following this question is discussed in this chapter.

6.1 In what ways does evidence exist?

Remember in Section 2.1 we said there exist things and that some of those things can be called ‘evidence’. We have also seen that contrary to mere things, items of evidence possess a reference to something else. Thus, we drew a line through the concept ‘thing’ and created the concept of evidence, based on whether such a reference is present or absent. Things that possess such a reference were named evidence. I am not quite clear how we should call things that do not possess such a reference. An expression like ‘not evidence’ seems correct but also uninformative. I will name them ‘facts of subsistence’, because we haven’t assigned any form to these things yet (and I am not going to in this thesis). That is, for the time being, we have acknowledged that these things simply *are*, without further specifying in what way they are. In short, we have divided the concept of things into two, in the course of which two categories were established: things named *evidence*, and another herewith named *facts of subsistence*. Note that nothing was added by such an operation as the concept of evidence already existed, although latently, in its generic concept ‘thing’. Stated otherwise, ‘evidence’ is an intrinsic property of ‘thing’. We merely emphasized in what way evidence and facts of subsistence are opposite. This intellectual operation is called *analysis* [50, 80].[1]

Next, I had to ask myself how to continue our ontological examination of evidence. In this respect D. A. Schum stated:

> ‘In my studies of evidence, I have troubled about how evidence might be usefully classified. This would be an utterly impossible task if we just considered the substance or the content of evidence. [...] It seems safe to say that the substance of content of evidence is unlimited in its variety. If all evidence items are absolutely unique, how are we ever to say anything general about evidence? [...] It occurred to me that evidence might be usefully classified on ‘inferential grounds’ rather than upon any grounds regarding its substance or content.’ [126, p. 206]

[1] The French term for analyzing in the context of law is ‘trancher’, which can be translated as ‘to dissect’.

I decided to follow a similar substance-blind approach. However, the deviation from D. A. Schum's approach consisted in focusing on the graphical probabilistic aspect of evidence. I wanted to find out whether such a perspective could add anything regarding the substance-blind classification already established by D. A. Schum (see Section 2.2). For the study we established three guiding questions, namely:

1. What criterion do we apply to distinguish one manifestation of evidence from another?
2. What are the implications entailed by such a criterion for a probabilistic argument we might want to undertake?
3. Can we establish a structure of kinship between such manifestations based on the probabilistic argument we identified?

The first question pertains to the drawing of further lines within the concept of evidence − just as we did when we distinguished between evidence and facts of subsistence. Thus, the criterion mentioned in the first question are ultimately analytical criteria. The second question aims at finding out how a given criterion affects the way in which we reason. The last question deals with commonalities and presuppositions among the implications generated by the criterion. That is, two or more criteria may have something in common, they can thus be regrouped[2]. In other cases, a given criterion may presuppose another criterion in order to make sense. By following these questions I created an ontology that resembles a genealogical tree. Such a tree is shown later in Figure 6.2 for recurrent combinations of evidence with the corresponding generic Bayesian networks. This excerpt of my ontology is discussed in the following section.

6.2 Ontology for combined evidence

The finding that the criteria themselves can be classified, is one result of my study. In particular, some criteria seem to prescribe a particular inference structure without giving any description of the inference; others, however, describe a particular inference within a given structure but provide no structural norm. I called the first type of criteria *inferentially prescriptive* and the latter *inferentially descriptive* criteria. To the present day, the natural redundance criterion discovered by D. A. Schum is the only prescriptive criterion for recurrent forms of combined evidence.

Imagine two reports R_1 and R_2 on two events, say E_1 and E_2. Each event is relevant to some degree for the hypotheses of interest H. If $E_1 = E_2$, then these two events are naturally redundant because they pertain to the same inferential object (but not necessarily to the same content or substance). On a structural level, this implies that there can only be one argument of relevance based on a single event $E = E_1 = E_2$. This situation is depicted as a BN in Figure 6.1 b. However, there remains one argument of credibility for each of the two reports. In turn, if the events E_1 and E_2 do not refer to the same inferential object, then each event necessarily possesses an argument of relevance on its own (see Figure 6.1). Again there is one argument of credibility for each report R_1 and R_2 (see Figure 6.1).

The inferentially descriptive criteria could further be classified into two classes. Some descriptive criteria were describing a particular probability assignment while others a particular form of observation (or variable instantiation in the context of Bayesian networks). The latter class of criteria is discussed in the paper reproduced in Chapter 2 of Part III. The probability assignment in an inference structure determines the inferential force a given item develops, that is, which hypothesis a given item favors. In combined evidence the items can either favor the same hypothesis or different hypotheses. If they favor the same hypothesis, then the evidence is said to be *harmonious*, and *dissonant* otherwise.

[2] This type of reasoning is called *synthesis*. As opposed to the analysis, a synthetic reasoning determines what is common in two different, or opposite things. Thus in essence, the analysis refers to a reasoning leading to a judgement of negation, whereas the synthesis leads to a judgement of affirmation [50]. The difference between these two modes of conceptual reasoning, will become important later.

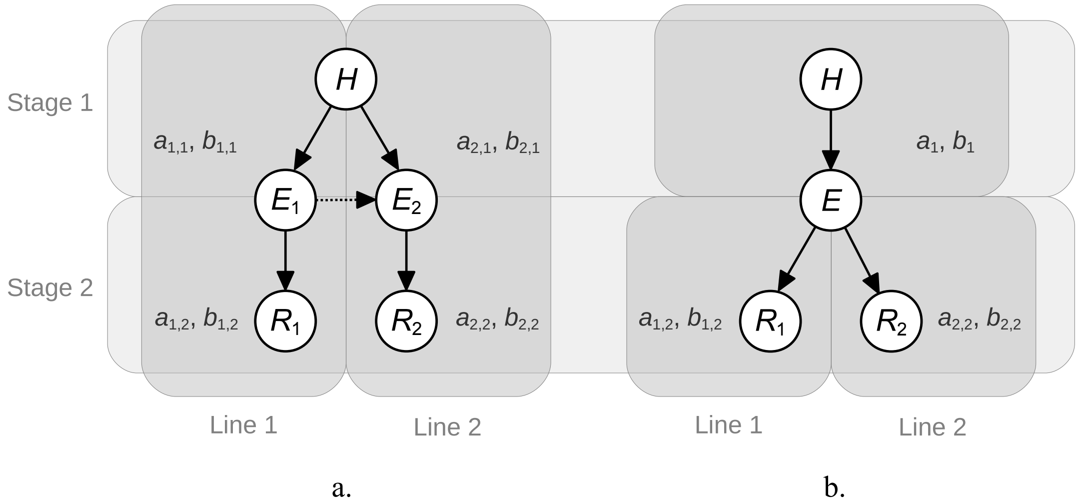


Figure. 6.1 – BNs for generic combinations of evidence with two lines of reasoning and two reasoning stages. a. Reports on naturally nonredundant events (i.e., $E_1 \neq E_2$). b. Reports on naturally redundant events (i.e., $E_1 = E_2 = E$). The probabilities of each reasoning stage and line of reasoning are denoted by the probabilities of $a_{i,j}$ and $b_{i,j}$ (or a_i and b_i), where i refers to the line of reasoning and j to the reasoning stage.

In Section 5 of Part I, the probabilities of the jth reasoning stage were denoted a_j and b_j. Following the same logic, the probabilities $a_{i,j}$ and $b_{i,j}$ refer to the jth stage of the ith line of reasoning. Thus, $a_{1,2}$ and $b_{1,2}$ refer to the conditional probabilities of the first report $R_1 = r_1$ (see Figure 6.1). Conversely, let $a_{2,2}$ and $b_{2,2}$ denote the conditional probabilities for the second report $R_2 = r_2$. If the inequality signs for $a_{1,2}$ and $b_{1,2}$, and $a_{2,2}$ and $b_{2,2}$ are the same (e.g., $a_{1,2} > b_{1,2}$ and $a_{2,2} > b_{2,2}$), then this is indicated by two arrows pointing into the same direction: $\upuparrows$. I called two reasoning stages, for which the relationship $\upuparrows$ applies, *inferentially congruent* reasoning stages. In turn, if the inequality signs are opposite (e.g., $a_{1,2} > b_{1,2}$ and $a_{2,2} < b_{2,2}$), then this is indicated by arrows pointing into different directions: $\uparrow\downarrow$. Two reasoning stages, for which the relationship $\uparrow\downarrow$ holds, were named *inferentially inverted* reasoning stages.

Harmonious evidence involving reports on naturally redundant events ($E = E_1 = E_2$) are said to be *corroborative*. Dissonant evidence involving reports on naturally redundant events are said to be *contradicting*. More precisely, they are corroborative if the reasoning stages encapsulating the arguments of credibility are inferentially congruent ($\upuparrows$), because in such cases the reports favor the same event. In contrast, they are contradictory if they are inferentially inverted ($\uparrow\downarrow$), that is, if they favor different events. As can be seen from Figure 6.2, there is exactly one possible situation in terms of congruence and inversion for corroborative and contradictory evidence respectively. Harmonious evidence involving reports on naturally nonredundant events ($E_1 \neq E_2$) are said to be *convergent*. Dissonant evidence involving reports on naturally nonredundant events are said to be *conflicting*. As you can see from Figure 6.2, as opposed to naturally redundant events, there are two possible situations in terms of congruence and inversion for convergent and conflicting evidence each.

Finally, the events in convergent evidence may be subject to different types of inferential interaction depending on probabilities that were assigned. Thus, if such an interaction is absent, then this means that the events are conditionally independent given the hypotheses. If they are present, then the distinction is threefold: directional change, redundance, and synergy. Inferential interactions will be discussed in later chapters and play a major role in combined evidence.

In any case, from Figure 6.2 you can see that the different criteria and their classes form the underlying skeleton for the different inference patterns and the characteristics of inferences these patterns may incorporate. Based on this ontology, there are two major reasoning types we can undertake. If you assess some evidence at hand by proceeding from the top to the bottom of Figure 6.2, the directionality of your reasoning process is from a generic towards a more specific concept of evidence. This means that your examination of the evidence is analytical in nature. In contrast, say you have already identified the exact inference pattern and the characteristic inference in your pattern, then you can follow the ontology from bottom to top. This allows you to check whether you have assessed all the criteria and their classes adequately, and whether your pattern and inference you chose are really the best for your evidence at hand. In such cases, the reasoning process passes from a specific towards a more generic concept of evidence, and your examination is synthetic. A similar skeleton that also involves single items of evidence is presented in the paper reproduced in Part III Chapter 3. It can be used in the same manner.

6.3 Atomism vs holism?

A question that arises in the context of combined evidence (complex cascaded inferences) is the degree of granularity we should choose, that is, how intricate and detailed an examination of evidence ought to be. Should we consider the evidence 'as a whole' and try to model it as such, or is it better to examine each item with its peculiar reasoning stages separately and aggregate them 'holistically' in a single model? I believe that the position held by D. A. Schum and P. Tillers is the only reasonable choice:

> 'It is hard even to imagine what it means to take evidence "as a whole". We perceive slices and various features in almost everything we see – and if we don't, perhaps we can't see anything at all. Moreover, it is hard to imagine how we can imbibe the evidence we "see" without performing *some* sort of mental analysis, which by definition seems to involve some sort of dissection. [...] The admonition not to analyze and dissect almost seems tantamount to advice not to think too carefully about the way you think. [...] Any theory that assumes an absolute dichotomy between holistic thinking and nonholistic thinking is thoroughly implausible and any theory that admonishes people to think globally rather than locally is vacuous.' [141, p.1252]

This means that any serious examination of evidence, and *a fortiori* combined evidence, requires both at varying degrees: a careful analysis of its atomistic parts (i.e. reasoning stages) and a logical aggregation of these parts into a whole, the latter being essentially driven by a synthetic process. Moreover, the evidential subtleties that we are going to examine in the following chapters, are subtle precisely because they cannot be accurately understood or even recognized if we one-sidedly adopt a coarse or narrow vantage point from which to consider evidence. In order to recognize and examine subtleties, we need similarly subtle successions of vantage points from coarse to narrow, and back. This requires an aptitude to analyze as well as to synthesize.

This requirement is equally relevant to the model construction for combined evidence. Problems involving combined evidence are usually too large to examine (see for example Chapter 9 of the present part). Additionally, as indicated above, a solely holistic approach is vacuous. Caesar's strategy of 'divide et impera' is a good advice for handling such large problems. It is also quite common in model construction [88, 101, 125]. The strategy stipulates the decomposition of the unmanageable global problem into closed, local problems. For each local problem a local model is created, which are then aggregated into a global model. As you may have anticipated, the problem decomposition is based on analysis. It allows us to obtain a collection of local problems in the form of semantical units. An aggregation of these semantical units for a reconstitution of the global problem is based on synthesis. That is, common features of these distinct semantical units are exploited for their combination.

My ontology of evidence in combination with D. A. Schum's ontology serves as a stepping stone for the in-depth examination of evidential subtleties, as well as the model construction of large problems. It allows us to identify and understand different forms of evidence, as well as their characteristic inferences arising from their atomistic parts. Moreover, the approach I chose, makes explicit reference to the necessary criteria for establishing distinctions between different forms of evidence. On the basis of such an ontology we can know exactly what criteria we presupposed in the light of some evidence. Stated otherwise, the different forms of evidence, are forms of evidence precisely because one can identify and apprehend the origin of their forms in such criteria. By doing so the concept of evidence is not one large impenetrable object of our mind; instead we have transformed it into slices of intellectually palpable aspects of evidential inference.

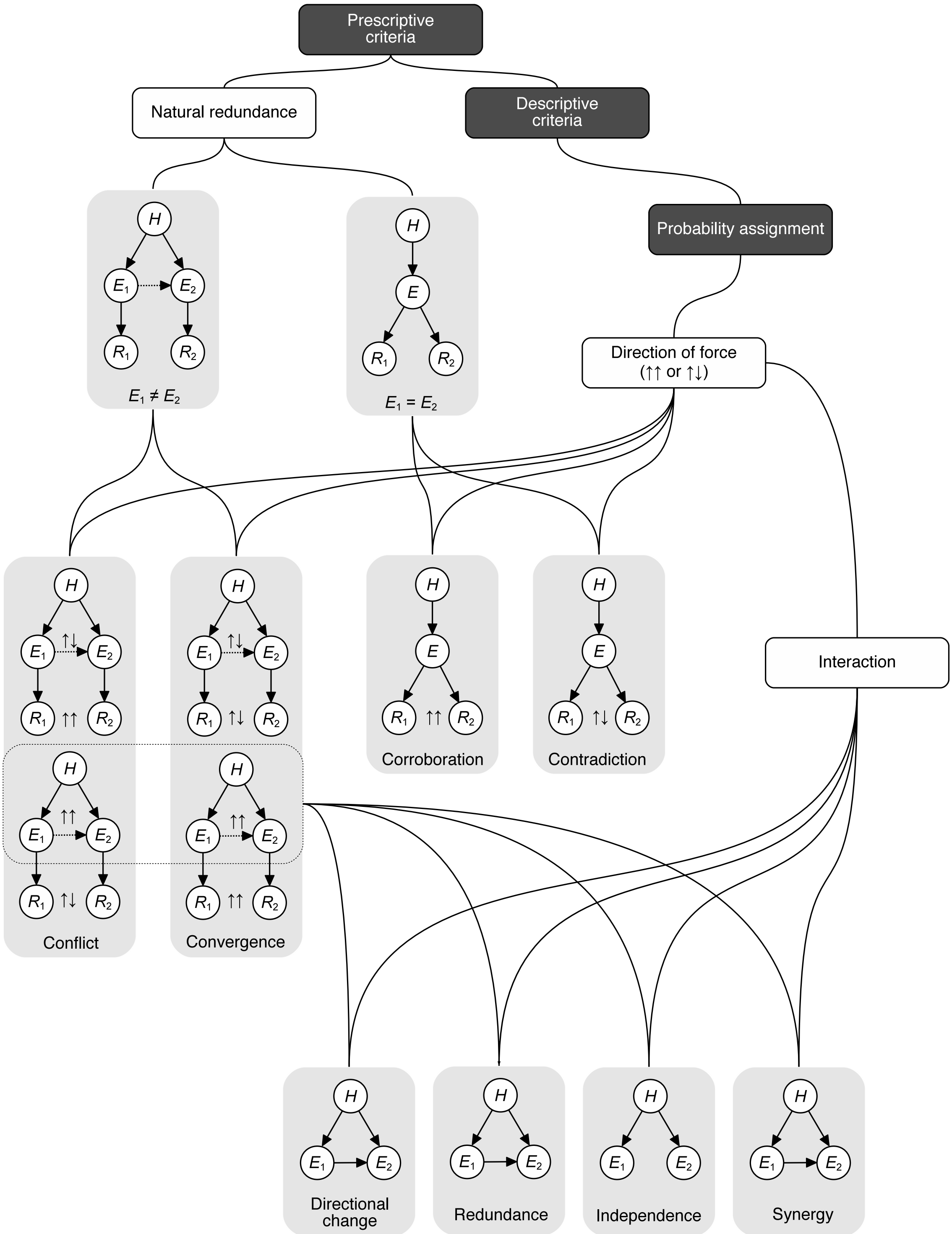


Figure. 6.2 – Ontology of recurrent combinations of evidence. Boxes in grey and white letters designate the classifications of the criteria. Boxes in white with black letters refer to a given criterion. An evidence examination from the top to the bottom of the ontology is analytic, and from the bottom to top synthetic.

7 A graphical probabilistic analysis of the combination of items of evidence

The next study deals with the examination of two real cases that were originally studied by Prof. Chrisophe Champod (similar case examinations are also described in [132, 133]). One case involves the evaluation of two features of a single shoe mark evidence. The other case deals with two distinct types of traces, namely a fingermark and a shoe mark. The study shows that even in the smallest forms of evidence combination, that is evidence involving two items, much can be learned about the nature and the probabilistic subtleties of combined evidence. The paper derived from this study is reproduced in Chapter 3 of Part III. The main questions of interest for this project were:

1. What are the relationships among a set of (usually unobservable) propositions and a set of (observable) items of scientific evidence?

2. What are the inferential interactions among the evidence items?

3. To what extent do these inferential interactions in combined evidence cause a deviation from the assessment of isolated items of evidence in terms of inferential force?

The first question refers to the relationship between the hypotheses and the items of evidence. The second question pertains to the inferential interactions we have seen in the previous chapter (the present cases involve redundance, synergy, and conditional independence). The third question relates to the examination of how these inferential interactions impact the inferential force of evidence, and how this may lead to the over- or understatement of the inferential force of evidence.

7.1 Inferential force and inferential interaction

Figure 7.1 a. depicts the Bayesian network used in the first case study, and Figure b., a simplified version of the one used in the second case study. As can be seen in these figures, the credibility aspect of the items were not considered because my main interest pertained to the arguments of relevance and their possible interactions rather than to the impact of observational errors. As a consequence, the report node R is absent in the present discussion.

As mentioned above, the first case involves a shoe mark. The general sole pattern (E_1) and the shoe size (E_2) were jointly evaluated in the light of the hypotheses (F), that is whether or not the suspect's pair of shoes is the source of the marks (i.e. source-level hypotheses[1]). Both items favored the hypothesis that the suspect's shoe is the source of the mark rather than an unknown shoe. That is, the evidence is an instance of convergent evidence. The study showed, that the shoe size was redundant to a certain degree given the general pattern. Stated otherwise, the

[1] I refer here to the hierarchy of propositions proposed by R. Cook et al. in their seminal paper of 1998 [23]. For more recent developments on this topic see [47, 43].

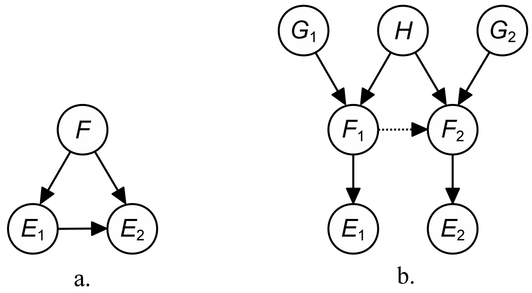


Figure. 7.1 – Bayesian networks for the combined assessment of two features and items of evidence respectively; a. the combination of the general pattern of a shoe mark (E_1) and the shoe size (E_2) given source-level propositions (F). b. the combination of a fingermark (E_1) and a shoe mark (E_2) given crime-level propositions (H). The nodes F_1 and F_2 refer to the respective source-level propositions for the two marks respectively. The nodes G_1 and G_2 represent the events of whether or not the marks were left by the offender (i.e., relevance for the crime in question).

general pattern already incorporated some information about the shoe size. It turned out, the LR took a value of roughly 160'000 when accounting for the inferential interaction. However, if these two features had been considered to be conditionally independent given the hypotheses (F), then the LR would have taken a value of over a million; a considerable difference of inferential force.

The second case involved the combination of fingermark (E_1) and a shoe mark (E_2) evidence stemming from a burglary case. More precisely, E_1 and E_2 refer to correspondences between a given mark and reference prints. The hypotheses, in turn, were formulated as crime-level propositions (i.e., $H = \{h_p, h_d\}$). Both evidence items favored h_p over h_d. It is, thus, again a case of convergent evidence. However, both marks were poor in recognizable details and generated, therefore, limited inferential force when considered in isolation. That is, the fingermark yielded an LR of roughly 138, whereas the shoe marks yielded 33. However, a joint evaluation of the evidence items produced an LR well over 4000, which is considerable.

Note also the dashed arc in Figure 7.1 b. It shows that an interaction is only conceivable between the source nodes F_1 and F_2 but not between E_1 and E_2. Indeed, it seems unreasonable to assume that the degree of correspondence between the fingermark and a reference print can directly influence the correspondence between a shoe mark and a shoe print – even if both were left by the same individual. Now, I said before that an interaction is conceivable (but not necessary). In fact, the variables represented by the source nodes are conditionally dependent given H only if we believe that the probability that the suspect left the shoe mark for reasons unrelated to the crime depends on whether the suspect left the fingermarks or not. However, by testing the impact of a conditional dependency relationship we noticed, that the inferential force hardly changed.

In essence, thus, inferential interactions can have a massive impact on the joint inferential force of evidence items, such as in the first case, but also almost no impact, such as in the second case. This result shows that inferential interactions are treacherous and their effects must be carefully examined from case to case.

7.2 Within a mark and between marks

Compare the reasoning patterns of Figure 7.1 a. and b.. As explained in the previous section, the situation in a. corresponds to the evaluation of two features of a single shoe mark, whereas case b. corresponds to the evaluation of a fingermark and a shoe mark. In the latter case, the combination is realized in the reasoning stage from the source-level propositions to the crime-level propositions. Structurally speaking, the two chains of reasoning associated with each mark converge in the hypotheses node. In contrast, the features of the shoe mark converge

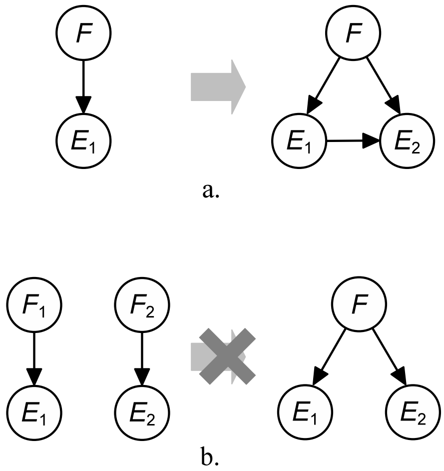


Figure. 7.2 – Bayesian networks for within-mark and between-mark evaluation given source-level propositions. a. within-mark inference: a mark is decomposed into its observational components. b. between-mark inferences.

in the source node. An interesting question at this point is whether we can combine distinct evidence items based on source-level propositions (i.e., by using a reasoning structure shown in a.). By applying this question to the second case, we can imagine a proposition of the sort ‘the suspect left the marks’. Notice, we can reformulate this proposition without any loss or addition of information as ‘the suspect left both marks’; and finally as ‘the suspect left the finger *and* the shoe marks’. However, the use of such propositions is tantamount to evaluating each mark in isolation. Nothing is gained from such a combination. Why then, does it make sense in the first case?

The reason is this: in source-level evaluations we establish an inferential connection between a particular object to a particular source. As a consequence, the evidence item is bound to its content. The two features in the shoe mark, namely the general pattern and the shoe size, are observational slices of the same item, and as a consequence, also of the identical content. Such slices are analytical products and therefore distinct, *intrinsic features* of the evidence. This implies that there was only a single inference to begin with: the connection of a particular shoe mark to a particular shoe (see Figure 7.2 a.). The synthesis of the observational slices leads back to the generic concept (e.g. the shoe mark).

In contrast, distinct marks, such as a fingermark and a shoe mark, are bound to distinct contents. Thus, we can only define the source as a simple sum of multiple contents (e.g., ‘the suspect left the finger *and* the shoe marks’). The same argument convinces equally in cases where the distinct marks are of the same type, such as multiple fingermarks, or multiple shoe marks. In such cases, we are compelled to define the source as a simple sum of multiple redundant contents. In general, an attempt to combine such distinct items at a source-level does not amount to a synthesis. There is no added semantical value because we produce only an umbrella term for the sum of multiple contents instead of a real generic concept that goes beyond its content.

This is represented in Figure 7.2 b. Thus, a meaningful combination of distinct marks can only be achieved when evidence items are examined in a more general framework. More precisely, we have to consider generic concepts that go beyond their contents and extend to *extrinsic features* of the evidence items. This can be achieved when activity- or crime-level propositions are considered.

8 An investigation of evidential phenomena in large bodies of evidence

Section 6.3 explained that we need subtle successions of vantage points from coarse to narrow in order to examine and understand combined evidence or even masses of evidence. Methods to examine evidential phenomena do not exist aside from sensitivity analyses [e.g., 83, 142] and conflict analyses [e.g., 76, 87]. Yet, these methods examine probabilities rather than the inferential force. The only method that uses the inferential force as a metric for examining evidential phenomena is D. A. Schum's redundance measure. It measures the inferential interaction in convergent evidence. However, there are two shortcomings with D. A. Schum's redundance measure. First, its application is limited to two events. Second, it is not clear how this measure was derived.

There were two further concerns, aside from the ones regarding the redundance measure. On the one hand: although dissonance was described by D. A. Schum, there were no methods to define and quantify evidential dissonance. On the other hand: in practice we are often interested in knowing which evidence item provides more inferential force relative to others. A quick outline of these subjects is presented below:

Redundance. In [125] D. A. Schum uses the redundance of statistical communication theory to measure redundance in probabilistic inference. The explanation starts with the average uncertainty[1] of an outcome $x_i \in X$, where $X = \{x_1, x_2, ..., x_n\}$, which is given by

$$H(X) = -\sum_{i=1}^{n} Pr(x_i) \log_2 Pr(x_i),$$

where $H(X) \geq 0$ applies, with equality, if and only if $Pr(x_i) = 1$ for an i. The average uncertainty is maximal if the distribution of $Pr(x_i)$ is uniform (i.e. $Pr(x_i) = 1/n$). The information theoretic redundance measure is given by the difference between the maximal average uncertainty $H_{\max}(X)$ and $H(X)$

$$R = \frac{H_{\max}(X) - H(X)}{H_{\max}(X)} = 1 - \frac{H(X)}{H_{\max}(X)}. \tag{8.1}$$

The ratio $H(X)/H_{\max}(X)$ is the relative uncertainty in a message array. It is maximal if $H(X) = H_{\max}(X)$, in which case we have $R = 0$. In contrast, if $H(X)$ approaches zero, then R approaches 1. The redundancy in this case is very high. Thus, we have the bounds $0 \leq R < 1$. An example is provided in Appendix A.

At this point, D. A. Schum makes a leap. That is, he establishes the following Equation

$$R_{e_2|e_1} = \frac{\log LR_{e_2|H} - \log LR_{e_2|e_1,H}}{\log LR_{e_2|H}} = 1 - \frac{\log LR_{e_2|e_1,H}}{\log LR_{e_2|H}}, \tag{8.2}$$

[1] This is alternatively called the average Shannon information content or entropy.

where $\log LR_{e_2|H} \neq 0$ (i.e. $LR_{e_2|H} \neq 1$). Note that Equation (8.1) uses the expected log-probabilities, whereas Equation (8.2) uses the logarithm of LRs. It remains, therefore, unclear why Equation (8.2) seems to make sense and allows us to capture the inferential interactions between two events, such as between e_1 and e_2. We felt unsatisfied by such a leap, as well as this measure's restriction to two events. We asked ourselves:

1. Can a redundance measure be established without making such a leap?
2. Can the redundance measure be generalized to an arbitrary number of events?

Dissonance. D. A. Schum has introduced the notion of dissonant evidence, that is, a body of evidence in which the items do not favor the same hypothesis. However, there was no method to assess and define the dissonance quantitatively. This raised the following questions

1. Is it possible to establish a mathematical definition of dissonance?
2. Can we effectively quantify the degree of dissonance in an entire body of evidence or to an arbitrary subset of items of evidence?

Relative contribution. When examining a body of evidence, we are often interested in knowing which evidence item(s) contribute most inferential force relative to others. For this reason we established the relative contribution measure, which is the ratio between inferential force of a subset of evidence items and some other set (i.e. another subset or the entire body of evidence). In this regard, I investigated whether there are general tendencies we can observe when accumulating evidence. More precisely, I asked

1. What is the most effective way to express the relative contribution of a set of evidence items?
2. What properties does the relative contribution exhibit in extreme cases of evidence accumulation?

All these questions were investigated by using WoE measure of the inferential force. The WoE measure is the most prolific metric to examine large bodies of evidence given its additive and symmetric properties (see Section 3.3.2). Moreover, D. A. Schum himself used the WoE for his redundance measure. The solutions to these questions provide a subtle succession of vantage points from which to examine combined evidence or masses of evidence. These measures are discussed in the present chapter. The paper itself is reproduced in Chapter 4 of Part III. Note also that the current state of knowledge K is omitted from our notation in order to increase clarity.

8.1 The redundance measure: A consequence of the multiplication law

As explained above, the transposition of the redundance idea from statistical communication to the context of evidential reasoning is unsatisfying. In order to address this deficit I derived the redundance measure from the context of evidential reasoning based on the metric of the WoE. I could show that the redundance measure is a direct consequence of the multiplication law of probability applied to events that are conditionally dependent given some hypotheses $H = \{h, \overline{h}\}$. Let e_1 and e_2 denote two events that are relevant for the hypotheses and that both favor h over $\overline{h}$. In particular we can establish the following equation

$$W(h : e_1) + W(h : e_2) - W(h : e_1, e_2) = W(h : e_2) - W(h : e_2 \mid e_1). \tag{8.3}$$

The lefthand side of this equation corresponds to the numerator in D. A. Schum's redundance measure (see Equation (8.2)). However, the righthand side of Equation 8.3 is the crucial expression: Not only does it provide a precise description of inferential interactions, but it also represents the key to the generalization of the redundance measure to an arbitrary number of events as we will see later.

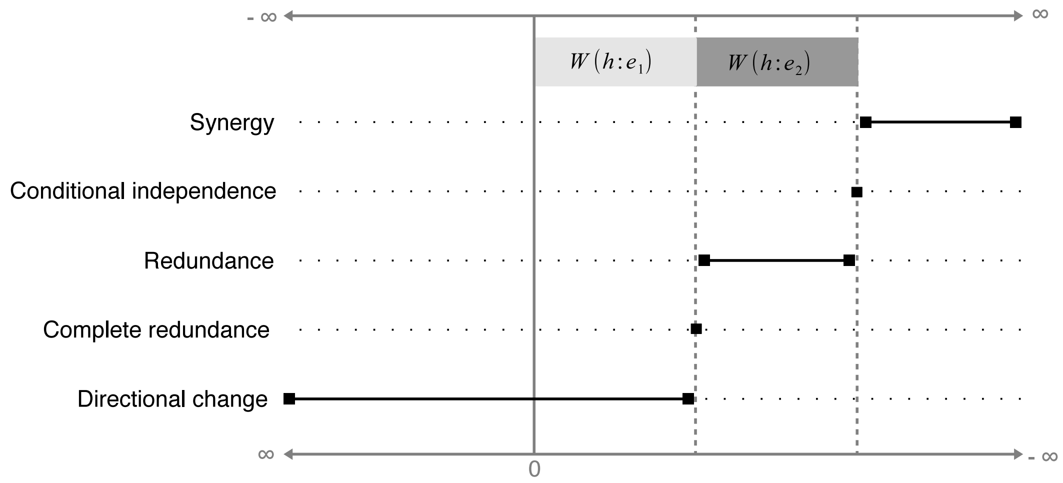


Figure. 8.1 – Schematic depiction of the domain of values for the weight of $W(h:e_1,e_2)$ associated with each type of inferential interaction between the events e_1 and e_2 given H. The length of the grey bars indicates the amount of weight produced by the simple sum of the weights of each event in isolation. The upper abscissa, ranging from $-\infty$ to ∞, applies if both events favor h over $\bar{h}$ (i.e. $0 < W(h:e_1)$ and $0 < W(h:e_2)$). The lower abscissa, ranging from ∞ to $-\infty$, applies if both events favor $\bar{h}$ over h (i.e. $W(h:e_1) < 0$ and $W(h:e_2) < 0$).

Consider what happens if both events are conditionally independent given H. In such cases we necessarily have $W(h:e_1)+W(h:e_2) = W(h:e_1,e_2)$ and $W(h:e_2) = W(h:e_2 \mid e_1)$. In this case both sides of Equation (8.3) become zero. What happens if $W(h:e_2 \mid e_1)$ is larger than $W(h:e_2)$? In such cases we have $W(h:e_1)+W(h:e_2) < W(h:e_1,e_2)$, that is, the joint weight of the events is larger than the simple sum of the weight of each event. The events are therefore *synergistic* given H. In contrast, if $0 < W(h:e_2 \mid e_1)$ is smaller than $W(h:e_2)$, then we obtain $W(h:e_1) < W(h:e_1,e_2) < W(h:e_1)+W(h:e_2)$. In other words, the joint weight of the events is smaller than the simple sum of their weights, but still larger than $W(h:e_1)$ alone. The events are thus *partially redundant* given H. However, it may happen that $0 = W(h:e_2 \mid e_1) < W(h:e_2)$. In such cases we obtain $W(h:e_1,e_2) = W(h:e_1)$. Since event e_2 possesses no weight at all once we know about e_1, it is *completely redundant* given H. Finally, you may be confronted with a situation, in which you find $W(h:e_2 \mid e_1) < 0$, although $0 < W(h:e_2)$. This implies that $W(h:e_1,e_2) < W(h:e_1)$. That is that the joint weight of the events is smaller than the weight of e_1 alone. You are dealing, therefore, with events that are *directionally changed* given H. Figure 8.1 summarizes the relationship between the different inferential interactions and the weights.

From two items to masses Equation (8.3) could be generalized to an arbitrary number of events irrespective of where they are located in a chain of reasoning. Consider again the expression $W(h:e_1)+W(h:e_2)-W(h:e_1,e_2)$. Let us rewrite $W(h:e_1)+W(h:e_2) = W(h:e_1,e_2 \mid \{E_1 \perp E_2 \mid H\}) = W_{\perp}(h:e_1,e_2)$. Hence, we can rewrite $W(h:e_1)+W(h:e_2)-W(h:e_1,e_2) = W_{\perp}(h:e_1,e_2)-W(h:e_1,e_2)$. The extent to which the events are engaged in an inferential interaction can, therefore, be expressed by the extent to which the value of $W(h:e_1,e_2)$ deviates from $W_{\perp}(h:e_1,e_2)$. This deviation is captured by the difference $W_{\perp}(h:e_1,e_2)-W(h:e_1,e_2)$.

Next, let $\mathbf{E} = \{e_1,e_2,...,e_n\}$ denote the set of events that are conditionally dependent given H, and which converge in H. Thus, we can generalize the redundance measure. Note that at this point I decided to call it the impact measure for inferential interactions (in $\mathbf{E}$ given H). This is because this method only allows us to measure the overall impact of the inferential interactions, but it cannot identify which inferential interactions are present[2]. The impact measure $ia(h:\mathbf{E})$ is given by

$$ia(h:\mathbf{E}) = \frac{W_{\perp}(h:\mathbf{E}) - W(h:\mathbf{E})}{W_{\perp}(h:\mathbf{E})} = 1 - \frac{W(h:\mathbf{E})}{W_{\perp}(h:\mathbf{E})}, \tag{8.4}$$

where $0 \neq W_{\perp}(h:\mathbf{E})$. This measure allows us also to examine an arbitrary subset of events such as $\mathbf{E}' \subseteq \mathbf{E}$ down to a pair of events in which case we return to the initial redundance measure (Equation 8.2).

8.2 Dissonance

The WoE is negative if the event favors $\overline{h}$ over h, and a positive sign otherwise. At the same time, the WoE measure is symmetric as mentioned in Section 3.3.2. The WoE is therefore an ideal metric for a measure of dissonance. For example, say we have two events e_1 and e_2 relevant for H. For clarity purposes let us assume that these are independent given H (the following discussion is also valid for events that are conditionally dependent given H). If e_1 disposes $W(h:e_1) = 2$ bans in favor of h, then an event e_2 that produces exactly the same amount of inferential force, but in the opposite direction, possesses $W(h:e_2) = -2$ bans. My dissonance measure recruits this property of the WoE.

I defined the *weight potential* of a set of events $\{e_1, e_2\}$ as $W_{\text{pot}}(h:e_1,e_2) = |W(h:e_1)| + |W(h:e_2)|$. In the present example we have $W_{\text{pot}}(h:e_1,e_2) = 4$. Next, I defined the *expressed weight*, which takes the form $W_{\text{ex}}(h:e_1,e_2) = |W(h:e_1) + W(h:e_2)|$. In the present case we have $W_{\text{ex}}(h:e_1,e_2) = 2 + (-2) = 0$. It seems natural at this point, to measure the dissonance in terms of the weight lost due to dissonance, or in other words, the difference between the weight potential and the expressed weight

$$W_{\text{diss}}(h:e_1,e_2) = W_{\text{pot}}(h:e_1,e_2) - W_{\text{ex}}(h:e_1,e_2). \tag{8.5}$$

Equation (8.5) possesses properties that are instructive for the degree of dissonance in a body of evidence. In general, the expressed weight can never be larger than the weight potential, that is $W_{\text{pot}}(h:e_1,e_2) \geq W_{\text{ex}}(h:e_1,e_2)$. This implies that $W_{\text{diss}}(h:e_1,e_2) \geq 0$, with zero only if $W_{\text{pot}}(h:e_1,e_2) = W_{\text{ex}}(h:e_1,e_2)$. In such cases, the degree of dissonance is minimal and the events are harmonious. The dissonance is maximal if $W_{\text{ex}}(h:e_1,e_2) = 0$ (and $W_{\text{pot}}(h:e_1,e_2) > 0$), that is, all the weight produced is lost through. We obtain, therefore, $W_{\text{diss}}(h:e_1,e_2) = W_{\text{pot}}(h:e_1,e_2)$. Note that for contradicting and conflicting evidence as well as evidence involving directional change, the weight lost through is always non-zero ($W_{\text{diss}}(h:e_1,e_2) > 0$). Figure 8.2 shows a schematic depiction of the different weights.

The dissonance measure can be applied to any number of events that are relevant to our hypotheses. Let $\mathbf{E}$ denote a set of events we would like to examine for dissonance so that $W_{\text{diss}}(h:\mathbf{E}) = W_{\text{pot}}(h:\mathbf{E}) - W_{\text{ex}}(h:\mathbf{E})$. Similarly we can focus on any desired subset of $\mathbf{E}$ such as $\mathbf{E}' \subseteq \mathbf{E}$ in order to obtain $W_{\text{diss}}(h:\mathbf{E}') = W_{\text{pot}}(h:\mathbf{E}') - W_{\text{ex}}(h:\mathbf{E}')$. A recurrent application of the dissonance measure to different subsets of the body of evidence can help us to trace the origin of a dissonance in our model. Stated otherwise, depending on the ordinality of the chosen subset, the dissonance measure can become a coarse chopper just as well as a delicate scalpel for our dissonance analysis.

How should we deal with dissonant evidence? Imagine you were examining a body of evidence, say $\mathbf{E}$, in the light of some hypotheses H. You apply the dissonance measure and find that your body of evidence is dissonant (i.e., $0 < W_{\text{diss}}(h:\mathbf{E})$). What are you supposed to do with this observation? Well, that depends. If you believe that it is in

[2]The medical notion of comorbidity designates presence of multiple illnesses in an organism. A comorbidity complicates the ascription of observed symptoms to a particular illness. In analogy, we could borrow this medical term for the presence of multiple interactions in a body of evidence. In short: the comorbidity of multiple interactions in a body of evidence complicates the diagnosis of particular types of interactions from the observation of their cumulative impact on the inferential force of the body of evidence.

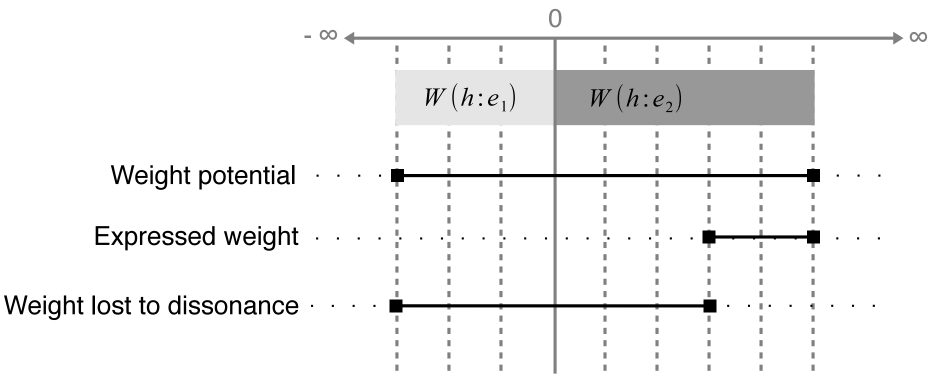


Figure. 8.2 – Schematic depiction of the weight potential, the expressed weight, and the weight lost through. The grey bars in the top represent the weights provided by two events e_1 and e_2.

our world's nature to produce dissonances, and that your body of evidence is just another caprice of such a world, then you might not attribute much significance to this observation. Stated otherwise, only harmonious evidence can surprise you. This extreme view has an opposite stance: you believe that our world does not produce dissonances. Our world's true nature is perfect harmony. Thus, if you observe a dissonance in your body of evidence, then this is because there is something wrong with your evaluation. Which of the two views is correct? Should we adopt a less extreme view, some middle ground between these views? To tell you the truth: I don't know (yet). To my best knowledge none of these questions has been tackled. I can only propose my intuitions on this subject. To me, dissonances represent warning signs, indicating that we may have made some erroneous assumptions or inferences. This is suggested by the fact, that we find contradicting, or conflicting arguments generally unpersuasive. That is, we are reluctant to accept such arguments and feel compelled to reexamine them. Conversely, I also believe that the larger the body of evidence, the more likely[3] it is for a dissonance to appear, simply because: 'It is not easy to tear any event out of the context of the universe in which it occurred without detaching it from some factor that has influenced it.' [113, p. 35]. Stated otherwise, any given model is at best an approximation of an extremely restricted part of the real world [87]. Therefore, we cannot (and should not) expect a model to provide a perfect fit for any body of evidence. There will always remain some elements of the real world that escape our senses and mind. Hence, the relevant question is not, whether we observe a dissonance among evidence items, but rather whether this dissonance is critical or harmless for the evaluation task.

8.3 Relative contribution

The relative contribution in terms of WoE is a straightforward concept. Let $\mathbf{E} = \{e_1, e_2, ..., e_n\}$ denote a body of convergent evidence and $\mathbf{E}' \subseteq \mathbf{E}$, where $\mathbf{E}' = e_1, e_2, ..., e_k$, the subset of evidence that is of particular interest. Assume further that none of the items provides conclusive evidence for h. Then the contribution of weight by $\mathbf{E}'$ relative to $\mathbf{E}$ is a ratio of their weights: $W(h : \mathbf{E}')/W(h : \mathbf{E})$. Now clearly, if $k = n$, then this ratio is 1, and as k tends towards 0, the smaller its weight contribution is relative to $\mathbf{E}$ until it reaches zero. We can also reason in the opposite direction and apply a slight modification. Say k starts at zero and increases towards n, where n tends towards infinity (i.e., $\mathbf{E}$ contains an infinite number of items of evidence). In such cases, the weight contribution

[3]Note that throughout this thesis, the words 'likely' and 'probable' correspond to the everyday use of the terms and are non-technical. They are used interchangeably.

of $\mathbf{E}'$ relative to $\mathbf{E}$ tends also towards zero. Finally, imagine all the items disposed the same amount of weight $W(h : e_i)$, where e_i is some interchangeable (regarding its WoE) item of $\mathbf{E}$. According to Equation (3.11) we have $W(h : \mathbf{E}')/W(h : \mathbf{E}) = (k \times W(h : e_i))/(n \times W(h : e_i)) = k/n$, that is the simple ratio of the number of items in the corresponding sets.

8.4 Analyzing large bodies of evidence

The different measures and methods to examine large bodies of evidence can be adjusted to any desired granularity, irrespective of the location of the events in a chain of reasoning. They provide, therefore, subtle successions of vantage points, which I have mentioned at the beginning of this chapter. In turn, these measures also raise questions. For example, how often do we encounter particular types of inferential interaction? Are there types of evidence that are particularly susceptible to inferential interactions, and if so to what types of interaction? Is dissonance a recurrent feature of combined evidence? How should we deal with dissonances and how are they dealt with by fact-finders? None of these questions has been raised in forensic or judicial literature, let alone studied. I believe that these questions must be followed up by further research.

9 *State of Texas vs Josiah Sutton*: A case analysis

In the previous chapter we have encountered the interaction type of directional change. This particular type of inferential interaction is, in my opinion, the most treacherous of all interactions we have seen, because it can lead to grave misinterpretations of cases. As you will see in the present analysis, the DNA expert in the case *State of Texas vs Josiah Sutton*[1] overlooked the fact that the DNA profiles were subjected to directional change. This led the DNA expert to believe that Josiah Sutton was a contributor to the DNA specimens from which the profiles were derived. However, a proper study of the DNA profiles in the framework of the case information suggests the opposite conclusion.

Josiah Sutton was accused of having participated, together with an unidentified assailant, in the abduction, rape, and sexual assault of a female. The crime was executed by two assailants and took place in the victim's own car. The DNA evidence consisted of four typing results that were established from four different specimens: the sperm fraction of the vaginal swab taken from the victim, the debris of the pubic hair combings from the victim, a stain from the victims jeans, and a stain from the victim's car. All these items were tested positive for semen. At the trial, the expert concluded that the typing results were a 'mixture of DNA types consistent with J. Sutton, the victim, and at least one other unknown donor...' [72]. Professor W. C. Thompson reviewed the case. By assuming that all the specimens were directly connected to the crime, and by considering all the crime profiles together rather than in isolation, he realized that it was extremely implausible for Sutton to be a DNA contributor to any of the specimens [139].

My research project consisted in the probabilistic joint evaluation of the four specimens given crime-level propositions. The main questions were:

1. How a Bayesian network could be created to model the DNA evidence, and

2. whether a Bayesian network, could reproduce the findings and conclusions obtained by W. C. Thompson.

The paper is reproduced in Chapter 5 of Part III.

9.1 Aggregating multiple DNA typing results

The modeling approach can be roughly divided into five stages: (i) the definition of the hypotheses, (ii) the creation of possible worlds in which the specimens can exist, (iii) the assessment of the number of contributors, (iv) the establishment of the possible contributor scenarios, and (v) the implementation of the typing results in the model. The present discussion focuses on the creation of the possible worlds and contributor scenarios (points (ii) and (iv)). The remaining points are explained only as much as needed for the understanding of points (ii) and (iv).

[1] State of Texas vs Josiah Sutton, District Court of Harris County (1999), Case No. 800450

(i) Hypotheses. At the outset, the prosecution considered two suspects. These were Sutton and an acquaintance of his named Adams. Thus, we established four propositions, each identifying a pair of individuals as the victim's assailants. Namely, we envisaged every possible pairing among Sutton, Adams, and two unknown persons.

(ii) Possible worlds. The enumeration of the possible worlds consists of determining the necessary conditions for the specimens to exist in the way they did. As a result, each world can be described by a set of necessary conditions. This involved the consideration for the creation of a given semen stain (i.e. ejaculation, non use of condoms), the relevance of the semen regarding the crime, or that the semen came from a person unrelated to the crime. We further concluded that the sperm fraction of the vaginal swab and the pubic hair combings had to be considered as naturally redundant given the hypotheses in order to avoid fallacious double counting. We calculated the most likely worlds based on the probabilities assessed for all the necessary conditions.

At the outset, we were confronted with 256 possible worlds. Given the practical impossibility of examining the typing results in each world, it is crucial to reduce the number of worlds. This reduction is achieved following three strategies: (a) by exploiting circumstantial information, (b) by weeding out worlds that are incompatible with the probability assignments previously made, and (c) by using items of evidence (i.e., the fact that all specimens yielded a DNA profile different from the victim). From the 256 initial worlds, nine survived the reduction process, from these nine worlds we retained the three likeliest worlds to be incorporated into the final model.

(iii) Number of contributors. The assessment of the number of DNA contributors to a given specimen consisted of computing the posterior probabilities of the number of contributors based on the typing result of a given specimen. This was accomplished by using the method proposed in [12]. The most likely number of contributors for the vaginal swab and the pubic hair combings were two and three, and for the remaining stains one and two.

(iv) Contributor scenarios. The hypotheses, the likeliest worlds, and the likeliest number of contributors, governed the logical framework that enumerates all the possible contributor scenarios. A contributor scenario defined as the actual DNA contributors in a given specimen, such as, 'Sutton, Adams, and the victim', 'two unknown persons', and so on. We ended up with eight distinct contributor scenarios for the vaginal swab and the pubic hair combings (considered as one specimen), and sixteen scenarios for each stain. Further, we found that in certain worlds, the contributor scenarios of the vaginal swab and the pubic hair combings logically determined the possible contributor scenarios of the stains. Based on the logical framework and the interdependencies between the scenarios of the different specimens we created a model that reproduces the scenarios. The scenarios themselves serve as a gateway between the hypotheses and the typing results.

(v) Typing results. For the evaluation of the typing results we implemented the modeling approach described in [104]. The genotype of the each possible contributor was modeled once and is the same for all the specimens. The typing results themselves enable the discrimination among the possible contributor scenarios.

9.2 The importance of circumstantial information

The reduction of possible worlds and of the number of contributors is crucial. This cannot be stressed enough. The more worlds and the more potential contributors we have to consider, the larger the number of possible contributor scenarios. In theory there are an 'infinite' number of possible contributor scenarios. This is because, on the one hand, we cannot determine a maximum number of contributors in a specimen with certainty; and on the other hand, we can imagine an infinite number of possible worlds that could have produced the specimens. However, a detailed examination of the framework of circumstances, expert knowledge, and evidence items can squeeze out crucial information for restricting the number of contributors and worlds. Conversely, without any additional information we are likely to end up with an astronomical number of contributor scenarios: a situation unfit for any interpretative effort. It appears that such information does not only serve to inform probabilities on parameters (such the relevant population for the random match probability), but also to organize evidence items for a joint evaluation; such information affects, therefore, the inference structure itself.

Notwithstanding, I believe that the combination through well defined contributor scenarios is a generally applicable method for evaluating multiple DNA specimens. In return, the heavy reliance on such information implies three things that must be kept in mind. First, our model is highly context-specific and likely to be unsuited in other cases; they are 'custom products' for a given case. Second, it is crucial for the evaluator of combined evidence to have access to such additional information. Otherwise, a combination becomes unfeasible. Third, given that such models are context-specific, the standpoint from which the model was created must be made transparent to fact-finders. In essence, the additional information that influenced the model must be properly conveyed to the fact-finders so as to render the conclusions derived from the model *intelligible* for them.

9.3 Inferences from the model

Figure 9.1 shows a schematic of the case examination. Given the size of the final model, I had to outsource the BNs for the computation of the posterior probabilities of the number of contributors and the likelihoods of the worlds. These probability values were then imported into the final model. This is shown by the boxes and arrows drawn with dashed lines. The final model has a distinct architecture produced by the organization of different problem areas, such as the scenarios for each specimen, the typing results for each specimen, and the observed typing results (i.e., credibility of the results). The different problem areas are depicted by boxes drawn with continuous lines. The corresponding arrow shows the influence of one problem area on another. The arcs in grey show further the conditional dependencies between the typing results given the hypotheses. The inferential interaction of directional change between the different typing results stems from these connections.

The inferences from our model corroborated W. C. Thompson's analysis [139]. That is, when the items were evaluated in isolation, they then supported the Sutton's involvement in the crime (LR ≈ 10). Note that the inferential force we obtain is very weak. In contrast, when the items are evaluated jointly, then the items favored two unknown assailants (LR ≈ 0.19). The application of our inferential interaction measure, which was discussed in the previous chapter, clearly indicated the presence of a directional change in the evidence.

In general, the joint evaluation of multiple DNA profiles is not a trivial task and requires an examination at a high degree of granularity. This is also indispensable for the detection of inferential interactions.

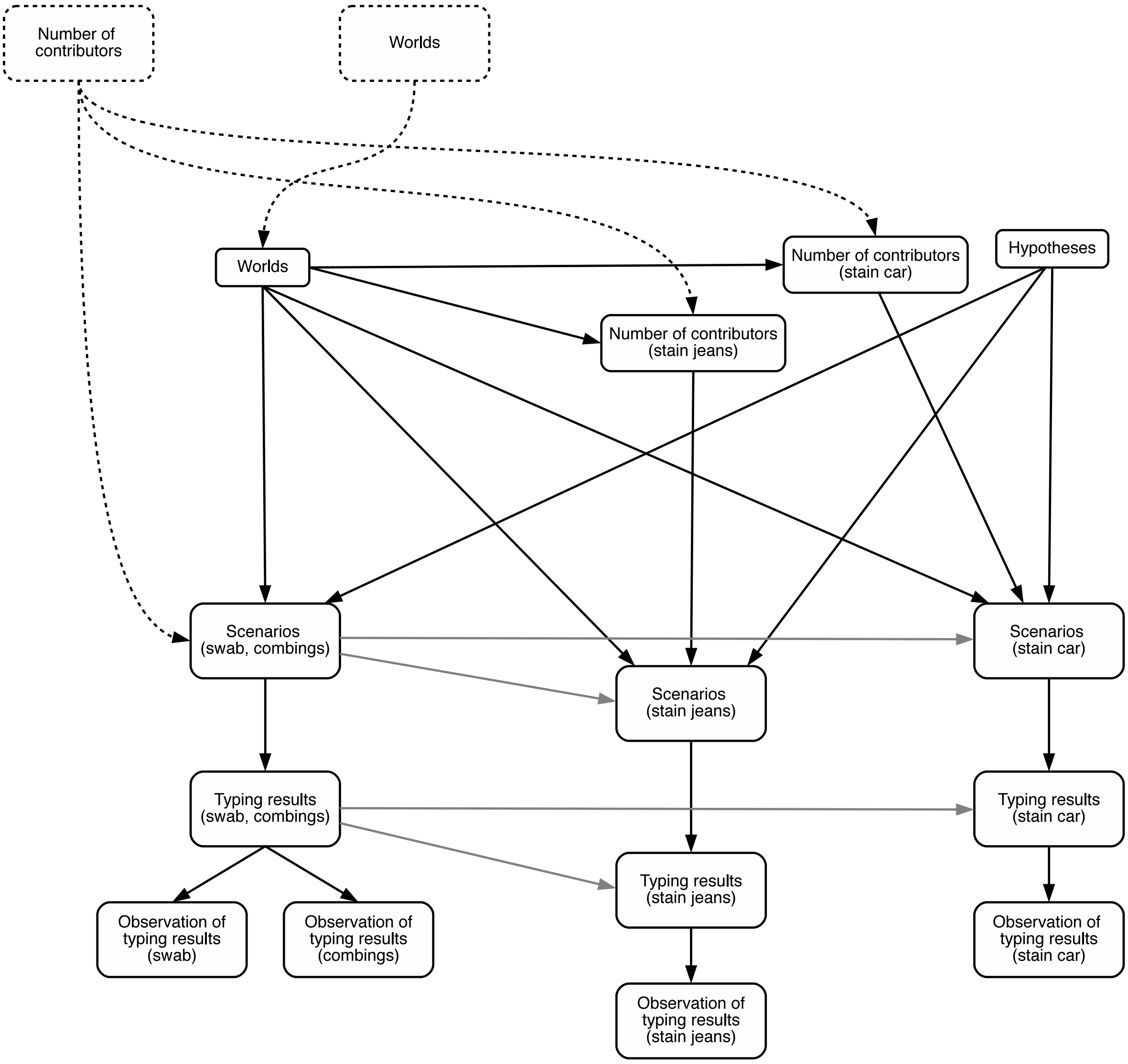


Figure. 9.1 – Architecture of the case examination. The boxes with dashed lines represent BNs that were outsourced from the final BN. The dashed arrows indicate the importation of probability values from these BNs to the final BN. The boxes with continuous lines represent thematically ordered clusters of uncertainties. The arrows represent influences between the clusters. The arcs in grey further highlight the influences responsible for the inferential interactions between the specimens.

10 Conclusion

This objective of the present thesis is to investigate the normative foundations of combined items of evidence. This was pursued by the method of probabilistic reasoning aided by Bayesian networks. The distillates of the thesis are wrapped in four catch phrases: 1. Reasoning about combined evidence is bi-directional; 2. Combinations ought to be relational; 3. Conclusions from combined evidence tend to be 'ethnocentric'; 4. Arguments of combined evidence tend to be ephemeral. Each of these statements is discussed in the following paragraphs.

1. Reasoning about combined evidence is bi-directional. This thesis shows the importance of an in-depth analysis and precise synthesis for the study of combined evidence. That is, the thinking is inherently bi-directional: top down and bottom up.

On the one hand, we are required to understand arguments of evidence down to its atoms (i.e. its reasoning stages), because evidential subtleties quickly escape coarse structures of evidential arguments. As a result, reductions (e.g., bypassing) of arguments are preferably (i.e., if feasible) realized after assessing the impact of evidential subtleties in more intricate and detailed argument structures rather than before.

On the other hand, bodies of evidence stemming from real cases are often very large. An effective evaluation of such bodies requires its decomposition into smaller units. A thorough understanding of generic arguments of evidence as well as the relationships between generic arguments not only supports the identification of suitable units, but also provides indications on how to reconstitute these units into a 'single final idea'.

2. Combinations ought to be relational. Combined evidence items do not simply represent sums of evidence items, such as item *A and* item *B*. The opposite holds: evidence items enter into a specific relationship, which is either dissonant or harmonious. In particular instances an item *A corroborates, contradicts, converges with, or conflicts with* an item *B*. These relationships can be defined and quantified within in the framework of probability theory by the dissonance measure. Moreover, certain relationships can further exhibit different forms of inferential interactions. These interactions cause the inferential force of combined evidence to deviate greatly from the inferential force of the sum of isolated items. This deviation can be exploited for the quantification of inferential interactions such as the impact measure. That is, we can compare the inferential force taking into account inferential interactions against the inferential force assuming no interaction.

Stated otherwise, combining evidence not only asks us to look at the evidence items, but also, and more importantly, what lies hidden between the items.

3. Conclusions from combined evidence tend to be 'ethnocentric'. The combination of evidence items requires an extensive use of circumstantial information and expert knowledge, which in turn impacts not only how we inform probabilities but also the argument structure itself. In addition, these structures are highly sensitive to our choice of hypotheses (e.g., source-, activity-, or crime-level propositions). The argument structure reflects, therefore, our point of view shaped by our hypotheses of interest, circumstantial information, and expert knowledge.

Stated otherwise, '... in making our decisions on both hypotheses and basic assumptions, ethnocentricity and even egocentricity play significant roles.' [114, p. 519]

In practice this means that a Bayesian network, such as the one created for the DNA evidence in *State of Texas vs Josiah Sutton*, is unintelligible to someone who is ignorant of our point of view: it is, in P. Feyerabend's sense, *incommensurable* [49]. The group of people who share our point of view will tend to agree with our model and its suggested conclusions (see point 4). Simultaneously, the group of people, who do not share our point of view, will find our model and its conclusions incomprehensible. We have to keep in mind that reasoning about combined evidence is 'ethnocentric' with respect to the former group, but marginalizing, in turn, for the latter group. An expertise on combined items of evidence requires, therefore, a particularly high standard for the transparency and the clear communication of the results.

4. Arguments of combined evidence tend to be ephemeral. The more evidence we integrate into our inferences the better. Stated negatively, the omission of available evidence in our inference is – normatively speaking – unreasonable [62]. Conversely, increasing evidence for our inference, means that the inference will be decreaslingly influenced by initial personal biases. Thus, divergent prior opinions tend to converge [e.g., 39, 54]. This '... ultimately leads to a consensus which will not be challenged again until the basic assumptions on which it rests are questioned.' [114, p. 519]

The consequence of questioning the basic assumption for combined evidence is this: the evidence items in Bayesian networks tackling large bodies of (mixed) evidence are organized based on the hypotheses of interest, circumstantial information, and expert knowledge. These are the basic assumptions on which our inference rests. This is also in agreement with the first principle of evidence interpretation, which states that

> 'Interpretation of scientific evidence is carried out within a framework of circumstances. The interpretation depends on the structure and content of the framework.' [45, p.235]

Therefore, reasoning about such bodies of evidence has inevitably a strong inclination towards contextualization and reluctance to generalization. More precisely, Bayesian networks involving large bodies of evidence are situation-specific and not generic (see for example the DNA evidence in *State of Texas vs Josiah Sutton*). This implies that not only our probability assessments, but the entire argument structure of the model becomes increasingly sensitive to changes regarding the hypotheses of interest, circumstantial information, or expert knowledge. Note that this sensitivity is not superficial; it is fundamental. We are likely held to overhaul the entire argument in order to cope with a change in our basic assumptions. As a result, the stronger the contextualization of the model, the less flexible and the more ephemeral it becomes.

Epilogue

'La vérité est que nul ne peut agir avec l'intensité que suppose l'action criminelle sans laisser des marques multiples de son passage.' [99]

'The truth is that no one can operate with an intensity required for a criminal act without leaving multiple traces of his motion.' (Free translation)

Stated in more general terms, an event, be it a crime, an accident, or a natural phenomenon, affects the environment on a broad front. Hence, it is more likely that we find a body of (mixed) evidence items rather than a single item purporting an event. These collections are 'cemented by a constant bond' [111, p. 266] and not loose parts, or a bunch of isolated entities. Conversely, we perceive items of evidence as isolated entities, because our attention towards things is inherently selective: we always focus on 'certain aspects of the stimulus situation to the exclusion of other aspects' [19, p.105]. As a consequence 'we perceive slices and various features in everything we see' [141, p.1252].

Our thinking is subjected to the same selective modality. Notice how we reason: we proceed by analyzing one evidence item after another, one evidence feature after another, and one piece of circumstantial information after another. The reconstitution of the collection is established afterwards, and is conveyed to us as a product of an intellectual effort, a synthesis. In contrast, we do not recognize a complete body of evidence that contains all the items, features, and pieces of circumstantial information at the outset. This suggests that we primarily perceive and think of evidence items as isolated entities, not because that is what they are in nature, but because it is in our nature to perceive and think of isolated aspects, or slices of reality. We are well advised to appreciate this inherent mental bias.

This bias is exposed to further exacerbation in forensic contexts due to the historical character of forensic investigations. Namely, the event that produced a collection of items is itself unobserved in the vast majority of the cases. We can only observe the relicts of past events (which may present themselves more or less revealing). Thus, the event itself is an issue of inference and so is the membership of an item to a certain body of evidence. In practice, we have to assess the membership one item after another given a set of possible events (e.g. hypotheses, or intermediate hypotheses). As a result, the consideration of evidence items as isolated entities seems more immediate than collections of evidence; the latter appear more indirect and artificial.

However, just because bodies of evidence feel indirect and artificial does not mean, that we should give in to our inherent bias and just attend to some part of the evidence, or to each item in isolation. On the contrary, Wigmore clearly states the imperative to consider all evidence items, and that all of them be properly organized relative to each other [146]. More precisely, in the formation of our belief, we must recruit all the evidence at once, and not in succession. Otherwise, we might neglect some items, devote asymmetrical attention to different items, and risk being misled by a part of the evidence.

> 'Our object, specifically, is in essence: *To perform the logical (or psychological) process of a conscious juxtaposition of detailed ideas, for the purpose of producing rationally a single final idea.* Hence, *to the extent that the mind is unable to juxtapose consciously a larger number of ideas, each coherent group of detailed constiutent ideas must be reduced in consciousness to a single idea; until the mind can consciously juxtapose them with due attention to each, so as to produce its single final idea.*' [146, p. 748]

That said, the reality of forensic practice is still charging in the (exact) opposite direction. Different types of forensic evidence tend to be delegated to specialized laboratories, where they are subjected to standardized routine-examinations and default-computations [e.g., 13]. It seems, that our bias is institutionalized rather than being rectified; instead of cultivating a multidisciplinary and inclusive approach to evidence as a remedy, a further fragmentation into specialism seems to be the dominating current in forensic practice [26]. Roux et al. gave a succinct characterization of this unfortunate drift:

> '... forensic science seems to be engaged in an out of control spiral forcing the discipline to reduce its scope to a series of service laboratories with limited strict analytical functions, rather than a set of interrelated processes that meet the needs of the security and criminal justice systems more holistically...' [119, p. 2]

Such a highly standardized and fragmented milieu is not only detrimental to forensic science in general, but it also lacks everything that is vital for the study of combined evidence. The consternation we experience when the 'out of control spiral' reaches its closing scene was poetically captured two hundred years ago by Johan Wolfgang von Goethe:

'Wer will was Lebendigs erkennen und beschreiben,
Sucht erst den Geist heraus zu treiben,
Dann hat er die Teile in seiner Hand,
Fehlt, leider! nur das geistige Band.' [144]

'He who would study organic existence,
First drives out the soul with rigid persistence;
Then the parts in his hand he may hold and class,
But the spiritual link is lost, alas! [143, p. 77]

The 'soul' of forensic science – as a science on its own – and the study of combined evidence – as a particular exercitation of this science – is driven out by a narrow attention slanted towards the dissection of the study of evidence into well defined fragments and specialisms. In order to maintain 'the spiritual link' we are bound to apply equal devotion and rigor in bridging the divide between these fragments. The study of combined evidence requires an extensive and inclusive mindset with a pronounced sensorium for hidden relationships. Such a mindset is completely incompatible with the dominating current previously outlined.

Part III

ARTICLES

1 The Evidential Foundations of Probabilistic Reasoning: towards a better understanding of evidence and its usage

“The Evidential Foundations of Probabilistic Reasoning” by David A. Schum “...contains a collection of thoughts..." (p. 1) on issues related to evidence and to inference tasks based on evidence. The study of such issues is best summarized by an expression introduced in Chapter 1: ‘Science of Evidence’. The Science of Evidence tries “...to treat the study of evidence as having a life of its own..." (p. 8). This perspective of examining evidence and inference with an interdisciplinary, generalist approach, is also reflected by the author David Schum himself: he is a professor of law and information technology and engineering at George Mason University. The fundamental insights he shares in this book are - unfortunately - all too often overlooked and unknown in forensic and judicial practice and research.

An important feature of evidential inference is its involvement with uncertainty, and consequently its probabilistic nature. This view is held also by Schum. He acknowledges that uncertainty is a prevalent feature of reasoning tasks based on evidence, and that it attends situations of daily life but also and most prominently, legal applications: “...in any inference task our evidence is always incomplete, rarely conclusive, and often imprecise or vague; it comes from sources having any gradation of credibility. As a result, conclusions reached from evidence [...] can only be probabilistic in nature." (p. xiii) Unfortunately, forensic practice regularly distrusts the notion of probability because people focus on precise numbers (derived from a generous data pool). However, assigning numbers for probabilistic evidence evaluation is neither a prerequisite nor an end for analyzes of evidential inference. Schum’s work is directly relevant to this aspect by demonstrating that (i) purely structural considerations on evidence and (ii) adopting probabilities as numerically variable ingredients of inferences, enable us to approach numerous problems, and to explore evidential subtleties or complexities. Let us first consider (i) and then (ii).

(i) Every item of evidence fans out into two primary dimensions: *relevance* and *credibility*. A relevance relationship between an event (for the purpose of this review let us say, ‘DNA matches with suspect’s DNA’) and a hypothesis (‘suspect is the assailant’) can involve a multistage reasoning (chain of reasoning). A given linkage pattern between elements of a chain of reasoning is called ‘argument’. Elements regarding the credibility of evidence (e.g. ‘how reliable is the expert reporting the DNA typing results?’) are located upstream in such a chain of reasoning. Depending on the type of evidence and the desired level of detail, it may also involve a multistage reasoning process and produce an argument. Thus, a probabilistic assessment of evidence requires an argument structured in terms of relevance and credibility. The argument structure becomes even more complex when multiple items of evidence are involved. In spite of this fact, basic configurations of evidence combination can be identified and analyzed probabilistically. Schum shows in his studies that such basic configurations of evidence combination result in specific inference structures and well defined inferential mechanisms.

(ii) Every item of evidence is characterized by an *inferential force*. It expresses if and to what extent evidence supports a hypothesis. Its quantity depends on the argument structure we choose for the evidence and on the probabilistic assessment we attach to the argument. The likelihood ratio is commonly used in Bayesian analysis to measure the inferential force of evidence. The study of likelihood ratios under varying probabilities is an important aspect of Schum's work: '[m]y essential research strategy was to perform sensitivity analyzes on the likelihood ratios I identified." ([124], p. 576) By doing so, Schum shows how certain argument structures give rise to peculiar inferential phenomenons such as in this non-exhaustive enumeration: inferential drag, redundancy, and synergism. Each additional reasoning stage in a chain of reasoning generally weakens the inferential force of an item of evidence: an inferential drag is accumulated. The likelihood ratio analysis on the inferential drag shows how such an accumulation is generated. Redundancy and synergism occur in specific configurations of evidence combination. The presence of the former implies that knowledge of one item of evidence can diminish or even nullify the inferential force of another. Ignoring redundancies can lead to overstatements of the joint inferential force of the items of evidence. Synergy relates to the opposite situation: the knowledge on one item of evidence increases the inferential force of another. Ignoring synergies leads to understatements of the joint inferential force.

Now, how is such knowledge useful in practice? First, it does not matter from which domain the evidence comes from, nor do we need to be familiar with its domain-specific methods and techniques to enhance our reasoning with these insights. Second, by identifying generic inference structures we know which inferential mechanisms we are exposed to and which we are not. Hence, we are less likely to be subjected to flawed reasoning leading to over- and understatements when assessing the inferential force of evidence. Imagine, for example, a DNA trace is analyzed by two laboratories. Now we have two results, but is our evidence also twice as strong? Third, knowledge on basic inference structures creates gateways to contextualized evidence interpretation, and even more so when we deal with masses of evidence (see for the analysis of a judicial case [79] and for a forensic case [78]). This is a particularly strong point since an item of evidence is typically found in conjunction with other evidence.

The book discusses a vast array of evidence-related subjects from different standpoints and across different disciplines. It demands time due to its broad scope; careful reading, and mental flexibility due to its interdisciplinary character. Sometimes it might even ask for the reader to be patient, as some subjects are developed incrementally making a few passages appear repetitive. In turn, many topics and problems that have appeared opaque and uneasy before may become clear and intellectually palpable afterwards. For readers who are interested in better understanding the properties of evidence and how to embrace evidence by systematic and logic reasoning, this a book that deserves serious consideration.

2 A probabilistic ontology of evidence and its combinations

ABSTRACT. This paper investigates and presents different manifestations of evidence from a forensic and judicial perspective to establish an ontology of evidence and its combinations. In general, the discussion will emphasize the following questions: (i) what criteria do we apply to distinguish one manifestation of evidence from another? (ii) what are the implications entailed by such a criteria for a probabilistic argument we might want to undertake?, and (iii) can we establish a structure of kinship between such manifestations based on the probabilistic argument we identified? Rather than focusing on a codification of evidence as in the Anglo-American law of evidence, this paper aims to discuss fundamental inference patterns and structures, and to elicit their probabilistic properties. Hence, distinctions that may be important from a non-probabilistic point of view may not be of the same importance in the present paper and vice versa.

Keywords. Ontology of evidence · combined evidence · uncertainty · probabilistic reasoning

2.1 Introduction

2.1.1 Evidence and uncertainties

In ancient Greek and Rome the study of evidence or signs[1] was an important discipline of rhetorics. This discipline is understood as the study to persuade other people, through deliberate argumentation, especially on political or judicial matters [52]. A telling example of what was understood as a sign or evidence can be found in Quintilian's Institutio Oratoria: "The Latin equivalent of the Greek σέμειον is signum, a sign, though some have called it indicium, an indication, or vestigium, a trace. Such signs or indications enable us to infer that something else has happened; blood for instance may lead us to infer that a murder has taken place."[2] [115, 5.9.9] Notice the twofold reference that evidence incorporates: to the thing itself and to the another thing ('something else has happened'). Modern scientific jargon often uses the word 'hypothesis' to describe the latter thing. For example, smoke is evidence for the hypothesis of a fire and a bloodstain on a garment is evidence for the hypothesis of someone's murder [3].

For the sake of argument we might agree that 'smoke' is conclusive evidence for the hypothesis of 'fire' that is, by seeing smoke we are certain that there is a fire. On the other hand, "(...) blood stains on a garment may be the

[1] Evidence is, first and foremost, a sign.

[2] In German we use the word 'Indiz' and in French we use both words 'indice' and more rarely 'vestige' when we mean evidence. It appears that these words meant exactly the same thing then and now including all the subtleties, except for the mathematical connotation that appeared much later in history [55, 52].

[3] Other terms that are commonly used for the 'thing itself' and the 'other thing' are the *factum probandum* (the fact that is to be proven) and the *factum probans* (the fact that proves) [146].

result of the slaying of a victim at a sacrifice or of bleeding at the nose. Everyone who has a bloodstain on his clothes is not necessarily a murderer." [115, 5.9.9] In this case our observation is inconclusive evidence for our hypothesis and almost all evidence is inconclusive in real life. As a result, we believe that someone is murdered to be true only to some degree.[4] It is here that we stumble upon the fundamental feature of evidential reasoning: uncertainty. Imagine now, that the bloodstain in question is very large and turns out to be of human origin. Knowledge or ignorance of this fact changes the degree to which we believe that someone was murdered. This suggests that uncertainty is conditioned on our state of knowledge at the time we form our belief. Therefore, uncertainty, as we understand it, should not be divorced from the person that perceives and assesses it: "(...) uncertainty is a personal matter; it is not 'the uncertainty but 'your' uncertainty." [97, p. 1]

By acknowledging the involvement of (your) uncertainty we can make further refinements. For example, we notice that a given item of evidence *e* does not always occur or happen. Therefore, the occurrence of an event claimed by *e* can be uncertain for us. In other words, at the initial stage, when we are considering ways to gather data, the item of evidence *e* conditional on the chosen way, is uncertain for us [96]. This kind of uncertainty is also sometimes referred to as the 'credibility' of the source that provides a given item of evidence *e* [125]. We will show in this paper that, depending on the situation and the nature of the source we might be held to make further distinctions within a source's credibility.

A further aspect of evidence is that conclusions drawn from evidence *e* upon a hypothesis *h* may leave us with a state of uncertainty, since *e*, if obtained, does not necessarily mean that *h* is true for us. This merely means that *e* is inconclusive, which is a feature of evidence already discussed above. In either case, it is of general interest to enquire about the inferential force and relevance that evidence *e* bears upon hypothesis *h*, and of the argument (reasoning pattern) on which we base the assessment of the credibility of the source and of the relevance relationship that links *e* to *h* [125]. In summary, thus, evidence can be conclusive or inconclusive, with the latter being associated with uncertainty. According to Schum, this raises 'three credentials of evidence' that need to be assessed: *credibility*, *relevance*, and *inferential force*.

2.1.2 Uncertainties and Probabilities

We have pointed out that uncertainty is a fundamental feature of the situation in which a person finds himself in a state of incomplete knowledge, resulting from inconclusive evidence. We have seen that relevance and credibility refer to types of uncertainty, but in order to assess the inferential force of evidence, we somehow have to combine these different types of uncertainties. The problem, therefore, is: how can we combine uncertainties? An effective solution to this problem consists in turning uncertainties into numbers which are easy to combine. In other words, we want to measure uncertainties which, following much forceful argument that is now being increasingly accepted, '(...) must obey the rules of probability calculus."[96, p. 298]

The relationship between law and probability theory is one of mutual affinity and maybe even necessity[5]. The benefit of probability theory for reasoning in judicial and forensic context remains immense even today [e.g., 4, 79, 118], essentially because it has become a standard. *Probability calculus* is our method of choice for handling uncertainties[6]. Although, probability theory allows us to build coherent and logical reasoning patterns, or probabilistic arguments[7] it may quickly become cumbersome, the more uncertainties we have to address. Not

[4] One early source where the distinction between conclusive and inconclusive evidence is made, can be found in Rhetorics to Alexander, attributed to Aristotle : "One sign causes belief, another knowledge." [7, 1430b] Signs that cause belief are called eikota (εικοτά), which literally means likelihoods [52]. Signs that are infallible are called tekmerion (τεχμεριον) and are based on syllogisms [7, 1357b].

[5] See for instance on forerunners of quantified probability calculus in the service of reasoned belief: "Significantly, most of these pioneers either had legal training (Fermat, Huygens, Leibniz) or came from a family that dealt in law (Cardano, Pascal, Arnauld)." [57, p. 229]

[6] We entertain the concept of subjective or epistemological probabilities here (as opposed to objective or frequentist probabilities). Probabilities in this sense represent personal beliefs, an expression of the relationship between a person's mind and the (uncertain) event the person reasons about as mentioned previously in Section 2.1.1 [96, 97].

[7] When we use the word 'argument' from here on, we mean a probabilistic argument.

only does it become increasingly difficult to examine the role that individual uncertainties play in our argument, it also becomes more difficult to establish an argument that embraces all the uncertainties in the first place. This is unfortunate, since our mind only has a limited capacity to deal with multiple ideas at a given moment. We can think about a great number of things but not about all of them at the same time. This, however, is precisely what is required in thinking about evidence, (and especially combinations of masses of evidence). Thus, "[o]ur object then (...) is in essence: To perform the logical (or psychological) process of a conscious juxtaposition of detailed ideas, for the purpose of producing rationally a single final idea." [146, p. 748]

To successively subsume different uncertainties so that we can rationally produce a 'single final idea' in the form of a graphical probabilistic model, Bayesian networks (BNs) can be used [28, 75, 108]. They offer a graphical representation scheme of uncertainties including their relationships of probabilistic independence given the current state of knowledge.

The application of BNs to problems that preoccupy people has been very successful in diverse fields of research (see [86, 112] for an overview). Forensic and judicial applications are no exceptions in this respect [e.g., 133]. BNs thus represent our method of choice for our study of the uncertainty regarding evidence and its combinations.

2.1.3 Ontology of evidence and its combinations from a probabilistic perspective

Differentiating manifestations of evidence and its combinations is essentially a task of classification or categorization. It allows us to say, roughly, that this form of evidence is one thing and that form of evidence another. Such formation of categories is the core-process for creating an ontology. This alone, however, does not suffice. The formation of categories needs to be pursued with method and consideration since an "[o]ntology (...) is the systematic study of existence: categories of things that can exist and the relationships they can bear to one another." [89, p. 1]

Generally, people are well acquainted with categories of things. Consider, for example, things that are alive and things that are not, liquid things and solid things, past things and future things, and so on. Creating categories is a fundamental feature of the functioning our mind and devised to make sense of the world, to establish structure, and to uncover relationships where we only perceive abruptness. In legal contexts, for instance, a skilled creation and use of categories is a key feature of a good investigator, as noted by Kind: "I have formed the impression, over many years of working in the field of crime investigation, that good investigators have this capacity to view events and to classify them, intuitively, in the most informative way." [82, p. 167] However, Kind's observation implies also that humans are at constant risk to employ uninformative, misleading, or even wrong categories. One method to increase our awareness of what categories we are using in a given situation, and to reduce the risk of employing inadequate categories, is to identify and name the criterion that produced a category in the first place. We ask: what is the condition that suggests a differentiation of things into categories in that particular case? Such is the method that we will use in this paper for establishing the different manifestations of evidence and its combinations. It is a way to keep track of the definition of the categories we have set up for a reasoning and not the other way around.

We have argued earlier that uncertainty is an intrinsic property of a reasoner's relationship with evidence and that the use of probability theory is vital to the understanding of the manner in which different uncertainties combine and interact. Thus, it seems natural at this point to investigate if and how a criterion that produced a category of evidence or its combinations affects a related probabilistic argument and whether we can establish on that basis a new structure of kinship that relates to our arguments. To make these analyses explicit, arguments will be represented in the form of BNs. They provide the analytical and structural framework within which the ontology of evidence and its combinations is investigated and presented in this paper. Thus, this paper does not address an ontology of evidence as stipulated by the Anglo-American law of evidence, but aims to provide a probabilistic analysis of fundamental inference-structures associated with different manifestations of evidence. Further, aspects of practical proceedings, such as admissibility, are not addressed. References to ancient Greek and Roman literature on evidence are made only as far as necessary for setting the context of a given subject. Readers interested in the history of evidence are referred to [55, 69, 52].

Table. 2.1 – Definition of events for a reasoning about Scenario 1

Events	
H	h: S murdered V.
	$\bar{h}$: S did not murder V (S has nothing to do with the case).
E	e: There are bloodstains on S's garment.
	$\bar{e}$: There is no bloodstain on S's garment.
R	r: There is a sensory report of reddish stains on S's garment.
	$\bar{r}$: There is no sensory report of reddish stains on S's garment.

2.1.4 Outline of the paper

In Section 2.2 we will investigate the fundamental argument for evidence in its simplest form as proposed by Schum [125]. At its core lies the elementary but often ignored fact that the report of an event is not the same as the event itself. Virtually all other forms of evidence − be it in combination or not − can be derived from this argument by extending and/or retracting one or more elements of that argument. Later in Section 2.3 we will investigate single items of evidence, followed by evidence in combination Section 2.4. Section 2.5 investigates the kinship among the different criteria identified. The paper concludes in Section 2.6. where we highlight our main findings.

2.2 Argument of evidence

It is often held that, for a sensible application of probability theory, events of interest should be clearly defined [83]. That is, regarding the scenario under investigation, the target events should embrace features of the scenario of which we have a clear understanding and that we can articulate appropriately. For the purpose of examining the basic constituents of an argument of evidence, as pursued throughout this paper, consider the following scenario:

> **Scenario 1:** We see S near the place where V lives. S seems upset and on his garment we observe reddish stains. We remember that S had a terrible dispute with V the day before. The possibility that S could have murdered V immediately crosses our mind and becomes object of our inquiry.

Table 2.1 proposes definitions for events of interest for this scenario. H represents the hypotheses of interest. E denotes the event regarding blood staining, relevant for H, and R the report by a source, which − in our case − is our senses. K, not listed in the table, denotes our current state of knowledge so that $K = \{$S is near V's place, S seems upset, S had a dispute with V the day before$\}$. All the beliefs we hold for each event are conditioned by K. Each event has two states, one for an event's occurrence and one for its nonoccurrence, each being denoted by a lower case letter. The question now is, how to order the three events so that a meaningful argument is obtained. General literature on BNs recommends an ordering that reflects causal considerations in order to achieve a consistent conditional dependency relationship between variables and to provide an intuitively interpretable reasoning structure [120, 83].[8] In the case considered here, it appears natural to assume that the knowledge of H influences our beliefs about E. The event, in turn, conditions our view on R. This argument structure is shown on the far left in Figure 2.1 as a BN where the variables are depicted as nodes and the direct dependencies are expressed by edges connecting two nodes. A reasoning structure where we employ a serial ordering of variables only is also called a *singly connected chain of reasoning* [125]. The specified beliefs for the nodes are represented by the node probability tables (NPTs) shown next to each node of the BN depicted in Figure 2.1.

[8]Note, however, that a conditioning order can be established between variables for which we do not conceive a relationship of cause and effect.

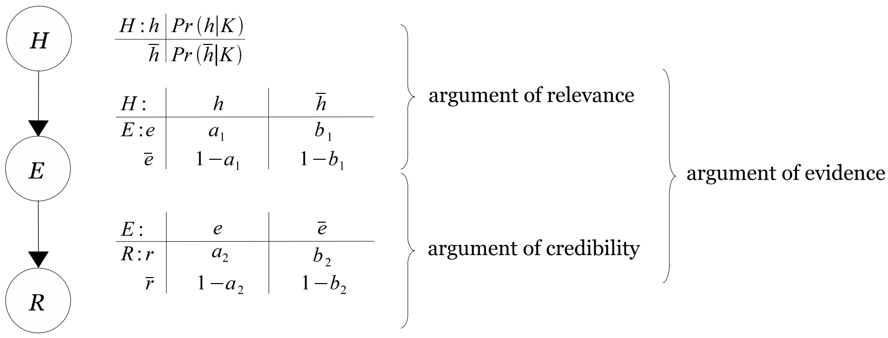


Figure. 2.1 – The BN dealing with the credibility of a report and the relevance of the questioned event E on our hypotheses E. Both the argument of credibility and the argument of relevance involve a single reasoning stage. The corresponding NPTs are listed next to the BN.

2.2.1 Basic reasoning stages of an argument of evidence

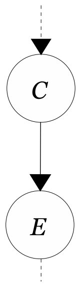
By tracing our inference from a given report R the hypotheses H, we identify two distinct reasoning stages: one from R to E and one from E to H. The first reasoning stage refers to the *argument of credibility* while the second refers to the *argument of relevance* (see Figure 2.1). The argument of credibility requires us to enquire about the following question: what is the nature of the source and what is the relationship between the source of the information and the object of its claim, given K? The assessment of the four probabilities in the NPT of R, each representing an aspect of the source's credibility, expresses the beliefs derived from this enquiry as probabilities, namely a_1 (short for $Pr(e \mid h,K)$) and b_1 ($Pr(e \mid \bar{h},K)$) and their corresponding complementary probabilities (i.e. $Pr(\bar{e} \mid h,K)$ and $Pr(\bar{e} \mid \bar{h},K)$ respectively).

The the questions raised by the argument of relevance are similar to the one of credibility: what is the nature of the event E and what is the relationship between the event and the ultimate hypotheses H given K? Again these questions are addressed in terms of probabilities. The four conditional probabilities in the NPT of E are of interest at this juncture. That is, a_2 (short for $Pr(r \mid e,K)$) and b_2 (short for $Pr(r \mid \bar{e},K)$) and the complementary probabilities $Pr(\bar{r} \mid e,K)$ and $Pr(\bar{r} \mid \bar{e},K)$. Finally, probabilities for the node H are assessed on K, that is, the probabilities of the hypotheses H prior to our reception of any report.

The above describes basic model to evaluate evidence. As will be pointed out in later sections, depending on the case under examination, we might be held to increase the number of reasoning stages to establish a relevance or credibility argument.

2.2.2 The inferential force of the evidence in terms of credibility and relevance

By comparing the beliefs we hold for our hypotheses before and after seeing reddish stains on S's garment we notice two things. First, we may note that these beliefs are different, and second, the probability of h has increased after seeing the reddish stains. Such a change in beliefs, in particular the extent of change in belief, depends on the *inferential force* (or, strength) of the evidence, and is appropriately measured by the likelihood ratio (V, short for 'value'). The likelihood ratio of a report r regarding our argument described above is given by the following ratio of two probabilities $V_{r|H} = Pr(r \mid h,K)/Pr(r \mid \bar{h},K)$. The notation $V_{r|H}$ refers to the likelihood ratio of r. If not stated otherwise, h is the conditioning event of the probability in the numerator, and $\bar{h}$ the conditioning event of the



$C:$	c	$\bar{c}$
$E: e$	a	b
$\bar{e}$	$1-a$	$1-b$

Figure. 2.2 – A BN representing a reasoning stage involving events C and E. The NPT of E is shown on the right-hand side of the BN.

probability in the denominator. According to the model shown in Figure 2.1, R relates to H through E so that the conversation can be extended to include E. By doing, so we can rewrite the likelihood ratio as

$$V_{r|H} = \frac{a_1 + [\frac{a_2}{b_2} - 1]^{-1}}{b_1 + [\frac{a_2}{b_2} - 1]^{-1}} \tag{2.1}$$

where the ratio a_2/b_2 is the local likelihood ratio assigned to the argument of credibility ($V_{r|E}$). The two other probabilities, namely a_1 and b_1, belong to the argument of relevance. Considered separately, their ratio represents the likelihood ratio of the relevance argument $V_{e|H} = a_1/b_1$. As can be seen in Equation 2.1 the reasoning stages are interwoven as is mirrored in $V_{r|H}$

$$V_{r|H} = \frac{a_1 + [V_{r|E} - 1]^{-1}}{b_1 + [V_{r|E} - 1]^{-1}}.$$

The likelihood ratio is a number that shows us if and by how much the evidence allows us to discriminate between hypotheses of interest. The likelihood ratio can take values from 0 to infinity, where values in the range $0 < V < 1$ are said to support the hypothesis in the denominator, values in the range $1 < V < \infty$ are said to support the hypothesis in the numerator (over the specified alternative). A value of 1 does not help discriminate between the two propositions. The evidence is, therefore, *logically irrelevant* [93]. The values 0 and ∞, on the other hand, represent evidence that establish, respectively, the hypothesis in the denominator and the numerator.

The expression $[a_2/b_2 - 1]^{-1}$ in Equation (2.1) is also known as the *drag coefficient* and is denoted as D by Schum [125]. It is a quantitative description of the inferential drag that the evidence r in the reasoning stage $R \to E$ carries on $V_{e|H}$. In general, a positive value of D implies that r favors h. A negative value means that r supports the alternative hypothesis. In any case, however, the overall likelihood ratios of $V_{r|H}$ or $V_{\bar{r}|H}$ are bounded by the local likelihood ratios of $V_{e|H}$ and $V_{\bar{e}|H}$ so that $V_{\bar{e}|H} \leq V_{r|H}, V_{\bar{r}|H} \leq V_{e|H}$.

2.2.3 Examining single reasoning stages in terms of certain, uncertain, and impossible events

To initiate the study of more complex evidential reasoning patterns, consider first the basic relationship between two events. In particular, consider the two events $C = \{c, \bar{c}\}$ and $E = \{e, \bar{e}\}$ organised in a reasoning stage as shown in Figure 2.2 with the relevant NPT for E. Let us further define $a', b' \in]0,1[$ and $a, b \in [0,1]$. This allows us to make the explicit and threefold distinction between certain ($a = 1$, $b = 1$), uncertain (a', b'), and impossible events ($a = 0$, Table. 2.2 – Typology of single reasoning stages with distinct specification of the NPT for the dependant variable E. Note that $a', b' \in (0,1)$.

(a)

C:		c	$\overline{c}$
E:	e	a'	1
	$\overline{e}$	$1-a'$	0

(b)

C:		c	$\overline{c}$
E:	e	0	1
	$\overline{e}$	1	0

(c)

C:		c	$\overline{c}$
E:	e	0	b'
	$\overline{e}$	1	$1-b'$

(d)

C:		c	$\overline{c}$
E:	e	a'	b'
	$\overline{e}$	$1-a'$	$1-b'$

(e)

C:		c	$\overline{c}$
E:	e	1	b'
	$\overline{e}$	0	$1-b'$

(f)

C:		c	$\overline{c}$
E:	e	1	0
	$\overline{e}$	0	1

(g)

C:		c	$\overline{c}$
E:	e	a'	0
	$\overline{e}$	$1-a'$	1

$b = 0$). By applying such a distinction to a single reasoning stage we can identify seven distinct NPT configurations for a single reasoning stage summarised in Tables 2.2 (a) to (g). NPT configurations where $a = b$ are not considered because such a configuration of probability values refers to situations where two events C and E are independent. In other words, knowledge about one event has no bearing whatsoever upon our beliefs about the other, and vice-versa.

Given the knowledge that either e or $\overline{e}$ is true, the NPT configurations (a) to (g) imply different reasoning stages that can be classified in one of two distinct categories of inference, described by Bernoulli in his treatise *Ars Conjectandi* as, respectively, *necessity* and *contingency* [10].[9] Configuration (d) represents a special case that needs further specifications. In particular, we must state which of the two possible situations applies $a' > b'$ or $a' < b'$.

A reasoning stage involving *necessity* refers to situations where the truth or falsehood of E implies to the truth or falsehood of C. For example, we may say, if e then necessarily c. We will use the symbol for a logical implication $\Rightarrow$ to denote such situations (e.g. $e \Rightarrow c$). Probabilistically speaking, we have $V_{e|C} = \infty$ and, conversely, $Pr(e \mid \overline{c})/Pr(e \mid c) = 0$.[10]

As opposed to necessity, *contingency* refers to situations, where the truth or falsehood of E does not establish the truth or falsehood of C. For example, by saying that 'if e then contingently c' we mean that e supports c over $\overline{c}$. In situations of contingency, the symbol $\rightarrow$ is used to indicate that the event on the lefthand side supports the conditioning event on the righthand side (over the alternative), e.g. $e \rightarrow c$. In Bayesian terms, this can be expressed as $Pr(e \mid c) > Pr(e \mid \overline{c})$ which implies that $V_{e|C} > 1$ and, conversely, $0 < Pr(e \mid \overline{c})/Pr(e \mid c) < 1$. Table 2.3 gives an overview of the reasoning stages in terms of necessity and contingency. Type (b) and (f) reasoning stages involve so-called 'deterministic' relationships, that is, dependent variables will take exactly one of their possible states, given the conditioning variable (i.e. the NPT of the dependent variable contains only the logical values zero and one). It is a characteristic feature of type (b) and (f) reasoning stages that they involve necessity irrespective of the direction of reasoning (i.e., diagnostic reasoning (from E to C) or predictive reasoning (from C to E)) and of the particular event we hold as true (i.e., e, $\overline{e}$, c, or $\overline{c}$). Type (d) reasoning stages refer to the opposite extreme as they remain contingent irrespective of the direction of reasoning and of the particular event we hold as true. All other

[9] Of course, Bernoulli did not invent the concepts of necessity and contingency (for the history of these concepts see [52]). However, as will be seen later in Section 2.3.2, he created evidential cases based on these concepts that cover the most fundamental inference types based on evidence.

[10] By convention we adopt $V_{e|C} = \lim_{Pr(e|\overline{c}) \to 0} \frac{Pr(e|c)}{Pr(e|\overline{c})} = \infty$ for a division by zero.

Table. 2.3 – Types of reasoning stages (as defined in Table 2.2), potential evidence (column two), associated likelihood ratio (column three) and nature of the influence (defined in terms of necessity, $\Rightarrow$, and contingency, $\rightarrow$). Note that $a', b' \in (0,1)$.

Reasoning stage	E	$V_{E\|C}$	Nature of influence
(a)	e	$V_{e\|C} = a'$	$e \rightarrow \overline{c}$
	$\overline{e}$	$V_{\overline{e}\|C} = \infty$	$\overline{e} \Rightarrow c$
(b)	e	$V_{e\|C} = 0$	$e \Rightarrow \overline{c}$
	$\overline{e}$	$V_{\overline{e}\|C} = \infty$	$\overline{e} \Rightarrow c$
(c)	e	$V_{e\|C} = 0$	$e \Rightarrow \overline{c}$
	$\overline{e}$	$V_{\overline{e}\|C} = 1/(1-b')$	$\overline{e} \rightarrow c$
(d)	e $(a' > b')$	$V_{e\|C} = a'/b' > 1$	$e \rightarrow c$
	e $(a' < b')$	$V_{e\|C} = a'/b' < 1$	$e \rightarrow \overline{c}$
	$\overline{e}$ $(a' > b')$	$V_{\overline{e}\|C} = a'/b' > 1$	$\overline{e} \rightarrow c$
	$\overline{e}$ $(a' < b')$	$V_{\overline{e}\|C} = a'/b' < 1$	$\overline{e} \rightarrow \overline{c}$
(e)	e	$V_{e\|C} = 1/b'$	$e \rightarrow c$
	$\overline{e}$	$V_{\overline{e}\|C} = 0$	$\overline{e} \Rightarrow c$
(f)	e	$V_{e\|C} = \infty$	$e \Rightarrow c$
	$\overline{e}$	$V_{\overline{e}\|C} = 0$	$\overline{e} \Rightarrow \overline{c}$
(g)	e	$V_{e\|C} = \infty$	$e \Rightarrow c$
	$\overline{e}$	$V_{\overline{e}\|C} = 1 - a'$	$\overline{e} \rightarrow \overline{c}$

types of reasoning stages are intermediate forms between these two extreme cases. These reasoning stages qualify as cases of necessity and contingency depending on the direction of reasoning and on the event we believe to be true.

2.3 Ontology of single items of evidence

In the following, we will investigate how different criteria modify the basic argument of evidence in terms of *credibility*, *relevance*, and *inferential force*.

2.3.1 Criterion based on the occurrence and nonoccurrence of events: positive evidence and negative evidence

Consider the case in which an item of evidence consists of a report regarding a particular conditioning event. Such a report can claim the occurrence or the nonoccurrence of the conditioning event. From an inferential perspective, it does not matter whether the claim relates to the occurrence or nonoccurrence as long as the report allows us to discriminate between competing hypotheses so that it presents some informative value to us. Take, for example, Scenario 1 that focuses on the sensory report of there being reddish stains on S's garment (r), or not ($\overline{r}$). If we were to observe such reddish stains, then this is an instance of what is called *positive evidence*, that is an item of evidence that reports an occurrence of an event. In turn, if we were to observe no such stains, then this is an instance of what is called *negative evidence*: an item of evidence reporting the nonoccurrence of an event. Clearly, both reports are such as to affect, potentially, our belief in the hypotheses of S murdering V. Sections 2.2 and 2.2.2 discuss how to interpret positive evidence. To interpret the nonoccurrence of reddish stains on S's garment, we can employ the same model as depicted in Figure 2.1 with the difference that, at this juncture, we enquire about the effect of $\bar{r}$ upon our belief about the event H. The equation for the likelihood ratio remains largely the same except for the drag coefficient. For cases of negative evidence the drag coefficient is composed of the complementary probabilities of a_2 and b_2, namely $1-a_2$ and $1-b_2$. All other components remain unchanged and the likelihood ratio can be written as follows:

$$V_{\bar{r}|H} = \frac{a_1 + D}{b_1 + D}, \tag{2.2}$$

while for positive evidence we have D_r and for negative evidence we have $D_{\bar{r}}$:

$$D_r = [\frac{a_2}{b_2} - 1]^{-1} = [V_{r|E} - 1]^{-1}, \tag{2.3}$$

$$D_{\bar{r}} = [\frac{1-a_2}{1-b_2} - 1]^{-1} = [V_{\bar{r}|E} - 1]^{-1}. \tag{2.4}$$

2.3.2 Criterion based on necessity and contingency: Bernoulli's four cases of evidence and their sub-cases

Consider again the notions of necessity and contingency in Section 2.2.3. In his treatise *Ars Conjectandi*, Jacques Bernoulli created categories of evidence based on these notions [68, 125].

Bernoulli distinguishes four basic cases of evidence based on the notion of necessity and contingency, namely:

- **Case 1**: The argument of relevance and the argument of credibility are both of necessity.
- **Case 2**: The argument of relevance is one of necessity, while the argument of credibility is one of contingency.
- **Case 3**: The argument of relevance is one of contingency, while the argument of credibility is one of necessity.
- **Case 4**: Both sub-arguments are of contingency.

Bernoulli's four cases can be further subdivided. First, we can distinguish between positive and negative evidence. Thus, in the current notation, we distinguish the reports r from $\bar{r}$. The second and the third distinctions are established by including considerations about the inferential force of the arguments. In particular, as a second consideration, we can make a distinction between the indications a given report (i.e., r or $\bar{r}$) provides on the conditioning event E: does the report indicate e or rather $\bar{e}$? Third, upon receipt of a report, we can focus on the indication an event E makes on our ultimate hypothesis H: does it indicate h or rather $\bar{h}$? Table 2.4 provides an overview of sub-cases resulting from such distinctions by using the notation introduced in Section 2.2.3. Note, however, that only situations corresponding to the natural sense of probabilistic evidence propagation are considered. Take, for instance, the case in which a report r favours e over $\bar{e}$. Subsequently, we will be concerned with how e influences the proposition H (e.g., $r \rightarrow e, e \rightarrow h$). However we will not consider the case, where $\bar{e}$ indicates h (e.g., $r \rightarrow e, \bar{e} \rightarrow h$).

For each possible configuration of reasoning stages discussed in Section 2.2.3, Tables 2.5 and 2.6 show the case and sub-case to which the basic argument of evidence belongs. Consider, for example, an argument of evidence with an argument of relevance and an argument of credibility involving both a type (a) NPT configuration. Since the report r is an instance of positive evidence (Section 2.3.1) Table 2.5 informs us that the argument of evidence in question corresponds to situation 2. Table 2.4, in turn, indicates that in 2 $r \rightarrow \bar{e} \Rightarrow h$ applies.

By comparing positive evidence with negative evidence it becomes clear that arguments of evidence categorise in different cases. Exceptions in that regard are all configurations between reasoning stages of type (b) and (f). These involve exclusively certain and impossible events. They represent configurations of conclusive reasoning stages that may be called *pure* case 1 arguments.

Table. 2.4 – Subcases for each of Bernoulli's four cases for positive and negative evidence in an argument that consists of the three variables H (the ultimate proposition with states h and $\bar{h}$), E (evidence potentially relevant for H with states e and $\bar{e}$) and R (the report of a source with states r and $\bar{r}$).

Bernoulli's cases	Subcases	Positive evidence	Negative evidence
1	1	$r \Rightarrow e \Rightarrow h$	$\bar{r} \Rightarrow e \Rightarrow h$
	2	$r \Rightarrow \bar{e} \Rightarrow h$	$\bar{r} \Rightarrow \bar{e} \Rightarrow h$
	3	$r \Rightarrow e \Rightarrow \bar{h}$	$\bar{r} \Rightarrow e \Rightarrow \bar{h}$
	4	$r \Rightarrow \bar{e} \Rightarrow \bar{h}$	$\bar{r} \Rightarrow \bar{e} \Rightarrow \bar{h}$
2	1	$r \rightarrow e \Rightarrow h$	$\bar{r} \rightarrow e \Rightarrow h$
	2	$r \rightarrow \bar{e} \Rightarrow h$	$\bar{r} \rightarrow \bar{e} \Rightarrow h$
	3	$r \rightarrow e \Rightarrow \bar{h}$	$\bar{r} \rightarrow e \Rightarrow \bar{h}$
	4	$r \rightarrow \bar{e} \Rightarrow \bar{h}$	$\bar{r} \rightarrow \bar{e} \Rightarrow \bar{h}$
3	1	$r \Rightarrow e \rightarrow h$	$\bar{r} \Rightarrow e \rightarrow h$
	2	$r \Rightarrow \bar{e} \rightarrow h$	$\bar{r} \Rightarrow \bar{e} \rightarrow h$
	3	$r \Rightarrow e \rightarrow \bar{h}$	$\bar{r} \Rightarrow e \rightarrow \bar{h}$
	4	$r \Rightarrow \bar{e} \rightarrow \bar{h}$	$\bar{r} \Rightarrow \bar{e} \rightarrow \bar{h}$
4	1	$r \rightarrow e \rightarrow h$	$\bar{r} \rightarrow e \rightarrow h$
	2	$r \rightarrow \bar{e} \rightarrow h$	$\bar{r} \rightarrow \bar{e} \rightarrow h$
	3	$r \rightarrow e \rightarrow \bar{h}$	$\bar{r} \rightarrow e \rightarrow \bar{h}$
	4	$r \rightarrow \bar{e} \rightarrow \bar{h}$	$\bar{r} \rightarrow \bar{e} \rightarrow \bar{h}$

We also observe that Bernoulli's four cases cover all possible configurations of reasoning stages for the basic argument of evidence. As a consequence, every argument of evidence that can be reduced to its basic form can be classified according to Bernoulli's four cases and their sub-cases.

Arguments of evidence belonging to cases 2 or 3 represent *noncascaded inferences*. That is, they involve no intermediary reasoning stage dealing with uncertainties between report and hypotheses (unlike case 4 arguments) [123]. In other words, if the argument of evidence belongs to case 2, then the inferential force (i.e. likelihood ratio) of the argument of evidence reduces to the inferential force provided by the credibility argument (e.g., if $a_1 = 1$ and $b_1 = 0$, then $V_{r|H} = V_{r|E} = a_2/b_2$). If, however, the argument of evidence belongs to case 3, then the inferential force of the argument of evidence reduces to the inferential force provided by the relevance argument (e.g., if $a_2 = 1$ and $b_2 = 0$, then $V_{r|H} = V_{e|H} = a_1/b_1$).

It is also worth emphasizing that certain configurations do not belong to a single sub-case or case (e.g., (a,d) or (d,f)). A common feature of these configurations is that they incorporate a type (d) NPT. As noted earlier in Section 2.2.3, type (d) NPT enable us to specify either $a' > b'$ or $a' < b'$. Depending on which of the two inequalities applies, the argument of evidence may correspond to a different case or sub-case. Hence, if the argument of evidence involves an NPT of type (d) (i.e. (·,d) or (d,·)), and we have not specified which inequality applies, then such an argument of evidence must correspond to two cases or two sub-cases. If the argument of evidence is type (d,d), then such an argument of evidence must correspond to all of the four sub-cases of case 4, since there are 2^2 possible configurations of inequality (i.e. two for the argument of relevance and two for the argument of credibility). For all arguments involving one type (d) reasoning stage, we use the following ordering: $(a' > b')$, $(a' < b')$. For type (d,d) arguments of evidence we use the following ordering: $(a_1 > b_1, a_2 > b_2)$, $(a_1 > b_1, a_2 < b_2)$, $(a_1 < b_1, a_2 > b_2)$, and $(a_1 < b_1, a_2 < b_2)$. The cell-coloring in Tables 2.5 and 2.6 is read from left to right.

Table. 2.5 – Bernoulli's Cases of evidence and their sub-cases for positive evidence

Argument of relevance		(a)	(b)	(c)	(d)	(e)	(f)	(g)
Argument of credibility	(a)	2	2	2	4 2	4	4	4
	(b)	2	2	2	4 2	4	4	4
	(e)	2	2	2	4 2	4	4	4
	(d)	3 2	3 2	3 2	1 4 3 2	1 4	1 4	1 4
	(c)	3	3	3	1 3	1	1	1
	(f)	3	3	3	1 3	1	1	1
	(g)	3	3	3	1 3	1	1	1

Table. 2.6 – Bernoulli's Cases of evidence and their sub-cases for negative evidence

Argument of relevance		(a)	(b)	(c)	(d)	(e)	(f)	(g)
Argument of credibility	(a)	3	3	3	1 3	1	1	1
	(b)	3	3	3	1 3	1	1	1
	(e)	3	3	3	1 3	1	1	1
	(d)	2 3	2 3	2 3	4 1 2 3	4 1	4 1	4 1
	(c)	2	2	2	4 2	4	4	4
	(f)	2	2	2	4 2	4	4	4
	(g)	2	2	2	4 2	4	4	4

Example: Problem of laboratory error in the context of DNA profiling

As an example, consider the analysis by Thompson and his colleagues for the problem of laboratory error in the context of DNA profiling [140]. Their analysis suggests a distinction between the event of corresponding profiles $E = \{e, \bar{e}\}$ (e.g., between the profile of a trace and the profile of a person of interest) and the report $R = \{r, \bar{r}\}$ of a scientist regarding the corresponding DNA profiles. In this analysis, the argument of credibility refers to the reasoning stage that goes from the reported correspondence, r, to the event E. Their analysis supposes no false negatives (i.e., $Pr(r \mid e) = 1$), but allows for false positive reports (i.e., $Pr(r \mid \bar{e}) > 0$). This corresponds to a type (e) NPT configuration in the analysis pursued in this paper (Table 2.2). The reasoning these authors proposed also involves a stage that goes from event E (i.e., corresponding profiles) to a source-level proposition H, that is, the person of interest is (h) or is not ($\bar{h}$) the source of the crime stain. Their analysis supposes that in the event of h a correspondence is certain, hence $Pr(e \mid h) = 1$, while in the event of $\bar{h}$, a correspondence e might apply with probability γ (i.e., the so-called conditional genotype probability). This corresponds to a further NPT of type (e). In combination, the arguments of credibility and relevance as identified in Thompson and his colleagues' [140] development lead to an argument of evidence of type 1. It can readily be seen that changes in the particular assumptions $Pr(r \mid e) = 1$ and $Pr(e \mid h) = 1$, typically the adoption of a type (d) NPT (with $a' > b'$) would lead to another argument classification (Table 2.5).

2.3.3 Criterion based on the reversibility of the argument of evidence: Quintilian's reversibility and nonreversibility, extending reversibility to inconclusive evidence, congruent and inverse reversibility

In his *Institutio Oratoria*, Quintilian distinguishes conclusive arguments of evidence having the same force when reversed (e.g. 'a man who breathes is alive, and a man who is alive breathes' [115, 5.9.6]), from conclusive arguments of evidence that cannot be reversed (e.g. 'because he who walks moves it does not follow that he who moves walks.' [115, 5.9.6]). By relating Quintilian's notion of reversibility to the cases and sub-cases identified in the previous section, the following statements can be made. First, since Quintilian refers to arguments on conclusive evidence (tekmerion), consideration can be limited to case 1 arguments. Second, such a case 1 argument of evidence is reversible, if the implication arrow can be reversed while maintaining the validity of the expression. For example, an argument of evidence corresponding to sub-case 1.1 $r \Rightarrow e \Rightarrow h$, is reversible if and only if $r \Leftarrow e \Leftarrow h$ also applies. Hence, an argument of evidence is reversible according to Quintilian if the relation $r \Leftrightarrow e \Leftrightarrow h$ holds. In Bayesian terms, this means that the value of the inferential force of the argument of evidence ($V_{r|H}$) and the value of the ratio of the posterior probabilities of H are the same, that is

$$V_{r|H} = \frac{Pr(r \mid h)}{Pr(r \mid \bar{h})} = \frac{Pr(h \mid r)}{Pr(h \mid \bar{r})}. \tag{2.5}$$

As noted in Section 2.2.3, arguments of type (b) and type (f) fulfill that requirement as they entail necessity in both diagnostic and predictive reasoning. Thus, arguments of evidence that only involve deterministic relationships of the kind (b,b), (b,f), (f,b), and (b,b) satisfy this requirement and should be called *reversible arguments of evidence*[11].

There is, however, an important difference between the group of sub-cases (b,b) and (f,f), and the group of sub-cases (b,f) and (f,b). The argument of relevance and the argument of credibility are *inferentially congruent* in the first group, while they are *inferentially inverted* in the second (see Table 2.3). In other words, in all the cases where the two sub-arguments are inferentially congruent, r implies the truth of the hypothesis of interest h so that $V_{r|H} = \infty$. In all the cases where the two sub-arguments are inferentially inverted, r implies the truth of the alternative hypothesis $\bar{h}$ so that $V_{r|H} = 0$.

We can formulate general conditions that allow an argument of evidence to become reversible so that Equation (2.5) applies. Such a perspective reveals that arguments on conclusive evidence are particular cases of a larger group of reversible arguments of evidence. In fact, arguments of evidence containing type (d) reasoning stages can also be reversible. Thus, the reversibility of an argument can also be formulated for inconclusive evidence. In general, an argument of evidence on inconclusive evidence becomes reversible if the following two conditions cumulatively apply (for a proof see Appendix B.1):

(i) the prior probabilities have to be equal, that is, $Pr(h) = Pr(\bar{h}) = 0.5$;

(ii) every reasoning stage, that is, every sub-argument in the case considered here, must involve normalized likelihoods (i.e., $a + b = 1$). Normalized likelihoods guarantee *inferential symmetry* in a reasoning stage [125] so that, for example, $Pr(r \mid e)/Pr(r \mid \bar{e}) = Pr(\bar{r} \mid \bar{e})/Pr(\bar{r} \mid e)$. In other words, r supports e to the same extent as $\bar{r}$ supports $\bar{e}$.

The feature of inferential congruence and inferential inversion also applies to arguments on inconclusive evidence. In particular, if the two sub-arguments are inferentially inverted, then $0 < V_{r|H} < 1$. And if the two sub-arguments are inferentially congruent, then $1 < V_{r|H} < \infty$.

We have stated earlier that arguments on conclusive evidence are special cases of a larger group of reversible arguments. They are special for two reasons. First, condition (i) is not necessary for arguments of evidence that

[11] Robertson and Vignaux use the expression 'ideal piece of evidence' that is, an event that always occurs when our hypothesis is true and never otherwise [118]. Robertson and Vignaux's ideal piece of evidence corresponds to the argument of type (f,f). However, it seems appropriate to qualify every piece of evidence as being ideal as long as it is reversible in Quintilian's sense (i.e. (b,b), (b,f), (f,b), and (b,b)).

only contain deterministic relationships (i.e. type (b) and (f) reasoning stages) to be reversible. Irrespective of the prior probabilities, the value of the ratio $Pr(r \mid h)/Pr(\bar{r} \mid h)$ is always either infinity or zero for such arguments. Second, type (b) and (f) reasoning stages already contain normalized likelihoods. The specification of normalized likelihoods is only necessary for type (d) reasoning stages.

Note further that all the other types of reasoning stages do not fulfill condition (ii). As a consequence, a basic argument of evidence containing a reasoning stage other than of type (b), (f), or (d) cannot be reversible.

2.3.4 Criterion based on relevance: ancillary, direct, and circumstantial evidence

Some forms of evidence are directly relevant for the inference task at hand (e.g. inference to H) and others are indirectly relevant. Generally, directly relevant items of evidence are explicitly represented in BNs. The bloodstain on S's garment in Scenario 1 is an example of directly relevant evidence. For directly relevant evidence a distinction is generally made between *direct evidence* and *circumstantial evidence* [125]. Evidence is said to be direct when it is conclusive on some issue, if the source's claim were perfectly credible. In contrast, circumstantial evidence remains inconclusive on some issue, irrespective of the source's claim being perfectly credible or not. The distinction pertains to the argument of relevance: if the latter is of necessity, then the evidence is an instance of direct evidence and if it is of contingency, then the evidence is an instance of circumstantial evidence. Thus, all arguments of evidence belonging to cases 1 or 2 represent direct evidence, while all arguments of evidence belonging to case 3 or 4 represent circumstantial evidence.

An item of indirectly relevant evidence is a particular piece of information that helps us to assess the strength or weakness of reasoning stages. Such items of evidence are called *ancillary evidence* [125]. In the context of BNs this would correspond to all elements of information that we use to assess the probabilities in NPTs such as, scientific research, rules of logic, knowledge about the case at hand and our general professional experience. In Scenario 1, for example, data on the occurrence of blood stains on clothing in the general population or our knowledge that S is a passionate huntsman could serve as ancillary evidence to assess the strength of the argument of relevance. Facts such as S wearing clothing with colorful patterns or S being far away at the time of our observation could serve as ancillary evidence for the argument of credibility. Usually, ancillary evidence is not explicitly represented in a BN. Although it can be represented, it is often not desirable to include such evidence as the model rapidly becomes cumbersome and, by doing so, no deeper understanding is gained for the targeted inference.

2.3.5 Criterion based on the nature of the source of evidence: tangible and testimonial evidence

The primary distinction regarding the nature of the source is whether an item of evidence was produced by a human or a non-human source. One might argue that, ultimately, all evidence is produced by a human source since no thing can be called a sign without there being a mind who calls it as such. This 'truism', as Wigmore calls it [146], is not ignored or disputed by endorsing a distinction between evidence from human and non-human sources. The aim of such a distinction is to identify the inferentially prevalent and, thus, the more influential element in an item of evidence.

Items of evidence produced by non-human sources are sometimes called *tangible evidence* [125] or *physical evidence* [82]. More precisely, these names comprise all items of evidence that relate to physical traces resulting from an event and includes measurements, test results, images, and other results of scientific analysis.[12] Items of evidence that are produced by human sources are commonly called *testimonial evidence* [125] or *personal evidence* [82]. They cover opinions, eyewitness reports, secondhand evidence (e.g., hearsay evidence, rumor or

[12] For items of tangible evidence that are directly presented in court, Wigmore used the term *autopic proference*. Such items include, for example, "... the production of a blood-stained knife; the exhibition of an injured limb; the viewing of premises by the jury; the production of a document." [146, p. 5-6] More commonly, this is also termed *real evidence* although for Wigmore himself real evidence included testimonial evidence given in court [6].

gossip), and expert testimony[13]. The distinction based on the nature of the source of evidence affects the argument of credibility, but not the argument of relevance [6]. Notably, credibility-issues that arise when confronted with tangible evidence include the *authenticity* of the item (is the item what it claims to be?) and the *accuracy* (does the device measure what it claims to measure?) and *reliability* (are the measurements reproducible?) of the sensory device that perceives the item. Credibility attributes associated with testimonial evidence, on the other hand, include *observational sensitivity*, *objectivity*, and *veracity* [125]. These attributes will be discussed in detail in Section 2.3.5.

Tangible evidence

In cases of tangible evidence, the argument of credibility usually has the same form as depicted in Figure (2.1), but the various probabilities involved are named in a particular way. For example, when using diagnostic devices to analyse the nature of biological fluids, the probability a_2 is called the *sensitivity* (of the method) and the probability of $1 - b_2$ the *specificity* (of the method). In situations where we consider a given report to be a sensor report or closely related to a sensor report (such as the cue given by a detection dog, or human colour vision), a_2 is also sometimes known as the probability of a hit and denoted as h [14]. In turn, b_1 is called the false positive probability and denoted as f (or FPP). The complementary probabilities $1 - a_2$ and $1 - b_2$ are called, respectively, the probability of a miss (m) and the probability of correct rejection (c).

An application of a reasoning that considers false positive rates in the evaluation of DNA typing results is discussed in [134, 140]. As noted earlier in Section 2.3.2, the argument of evidence put forth in these publications corresponds to case (c,c).

Testimonial evidence

In cases of testimonial evidence the argument of credibility is build on multiple attributes. The attributes in question are observational sensitivity, objectivity, and veracity[15]. Schum presents a reasoning model that incorporates these attributes [125]. Figure 2.3 shows the reasoning pattern as BNs. In essence, the model is a chain of reasoning with an argument of credibility composed of three sub-sub arguments. Each additional reasoning stage captures a particular attribute of human sources. The conditioning order from top to bottom is based on the temporal succession of the events. Thus, it encodes the following reasoning: an event e occurs, e is perceived, a belief about e is formed based on what was observed, and then the belief about e is communicated. The argument of relevance remains unchanged with respect to Figure 2.1.

The inductive reasoning stage from S to E represents the argument of observational sensitivity involving the probabilities a_2 and b_2. The first refers to the probability of the event s, the source's senses receiving evidence of the event e given that e occurred. The second, refers to the probability of the event s given that $\bar{e}$ holds. There is an important point to be made regarding the conditional probability involving the alternative proposition $\bar{s}$. Here we do not enquire about whether the source's senses gave evidence of $\bar{e}$. We focus on whether the source's senses did not receive evidence of e. These two things are not the same. The negation relates to the sensory reception of evidence of e and not to the event itself.

The reasoning stage from O to S relates to the argument of objectivity and involves the probabilities a_3 and b_3. At this juncture, we enquire about how probable it is that the source believes that e occurred given that the source received sensory evidence of the event ($Pr(s \mid o)$) and given that the source did not receive such sensory evidence $Pr(o \mid \bar{s})$, respectively. The latter probability includes the possibility that the source received no evidence at all

[13] Note that authoritative records are usually considered as a distinct category apart from testimonial evidence. In the context of forensic science, one could reasonably consider peer reviewed publications to be authoritative records.

[14] The probability of a hit h is not to be confounded with the hypothesis h.

[15] Wigmore used the terms *perception*, *recollection*, and *communication*. An insightful, but an essentially non-probabilistic, discourse on testimonial evidence can be found in [146].

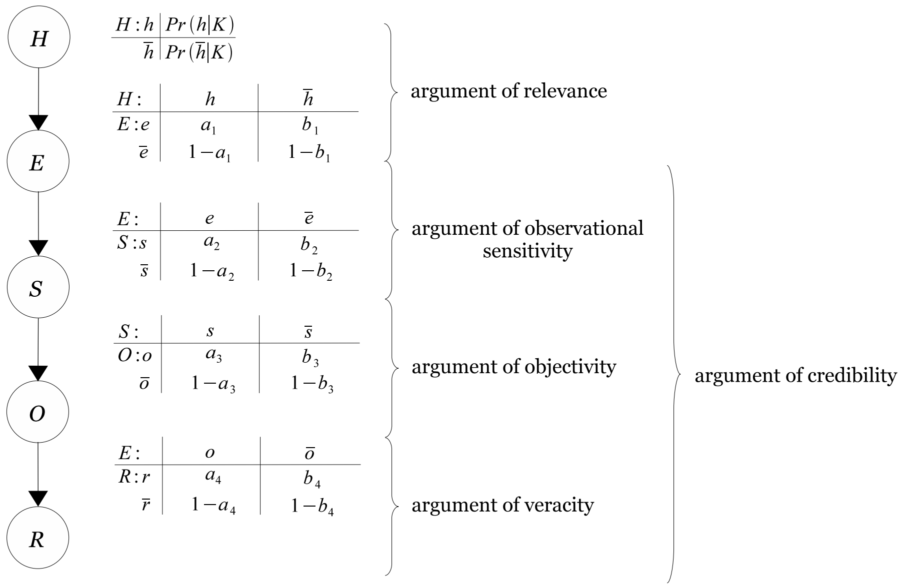


Figure. 2.3 – The argument of credibility for testimonial evidence is composed of three sub-sub-arguments: the argument of observational sensitivity, the argument of objectivity, and the argument of veracity.

regarding e. As for the alternative proposition $\bar{o}$, note that we do not focus on whether the source believes that $\bar{e}$ holds, but whether the source does not believe that e holds.

The final reasoning stage from R to O refers to the argument of veracity. The probabilities involved are a_4 and b_4. The former is the probability of the source reporting e, given that she believes that e occurred ($Pr(r \mid o)$). The latter represents the probability of the source reporting e, given that the source does not believe that e occurred ($Pr(r \mid \bar{o})$). The latter probability includes the possibility that the source holds no belief whatsoever regarding the event. Similar to before, the alternative proposition $\bar{r}$ refers to the source's report of e, and not to the event e itself.

Consider some important aspects of this model. First, the sensory perception (S) of an event is directly influenced by the event itself and the conditions in which the sensory evidence is obtained. These conditions cover two dimensions: (1) the mental and emotional state of the source and (2) the physical environment in which the event occurred. These conditions are acknowledged by K when expressing our current state of belief, including an explicit reference to ancillary evidence. Second, a source's belief regarding the occurrence of some event (O) is directly influenced by what was observed by the source. A source's objectivity or non-objectivity is assessed by examining the extent to which a source's belief on an event deviates from (or concurs with) the content that the source's senses have received. Third, and most importantly, a source's veracity regarding some claim cannot be meaningfully evaluated in a framework that dissociates it from the source's belief on that matter. Indeed, the truth or otherwise of a source's claim is irrelevant for the source's veracity as long as the source communicates according to his or her belief. A source is untruthful if and only if he or she exhibits an intent to deceive by making the effort of communicating a claim that is contrary to his or her belief. These are all vital aspects of Schum's and Wigmore's studies on testimonial evidence.

All other forms of testimonial evidence incorporate the credibility argument described in this section. Take, for example, an item of secondhand evidence, where a witness reports the occurrence of an event that he or she

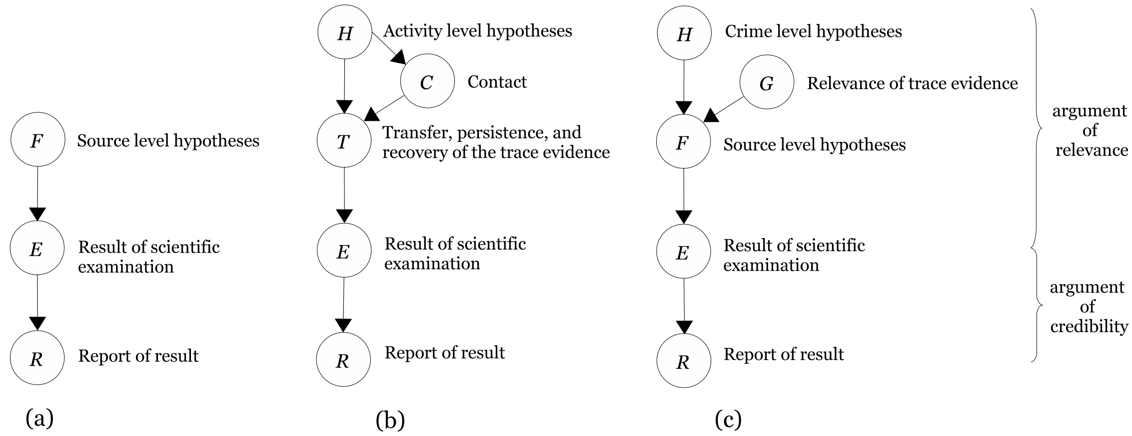


Figure. 2.4 – (a) generic BN for interpreting an item of evidence given source-level or sub-source-level propositions. (b) generic BN for interpreting an item of evidence given activity-level propositions. (c) generic BN for the interpretation of an item of evidence given crime-level propositions.

had been told about by another person. Note that only the latter person has made an observation regarding the event in question and not the witness him- or herself. Hence, the person who made the actual observation is the *primary source* and the person who gives the testimony is an *intermediate source*. In such a case, two arguments of credibility are concatenated, that is, one for the primary source, which is directly connected to the argument of relevance, and one for the intermediate source, which is connected to the credibility argument of the primary source. Following this logic one can evaluate a situation for any number of intermediary sources by adding an argument of credibility for each additional source (for a more detailed explanation on testimonial evidence see [123, 125, 128]).

One has to keep in mind that, in general, the longer the chain of reasoning the smaller the inferential force resulting from the argument of evidence. It follows, therefore, that the witness and the primary source should ideally coincide or that the witness should be at least as close as possible to the primary source. The definition of the inferential force of an item of testimonial evidence is given in Appendix B.2. Note that technically it is possible to reduce the chain length for such reasoning chains (see Appendix B.3 for a general account on bypassing intermediary variables and Appendix B.4 for a particular account on bypassing intermediary variables in testimonial evidence). Such a reduction in chain length, however, does not make up for the loss of inferential force because the reduced model still accounts for all the uncertainties involved in the original reasoning chain.

2.3.6 Criterion based on the level of propositions: sub-source-level, source-level, activity-level, and crime-level propositions

A primary aim of forensic science is to pursue a contextualized evaluation. It is an expression of the view that an item of tangible evidence should be embedded in the most meaningful way into the case at hand, and that this is crucial. A key element to achieve this goal is the appropriate choice of target questions and related propositions. Cook and his coworkers analysed and described different propositions that forensic scientists commonly encounter when evaluating tangible evidence. They were able to observe common characteristics among different groups of propositions. This led to a categorization of propositions that became known as 'the hierarchy of propositions' [23]. Originally, there were three categories, also called 'levels': source-level (level I), activity-level (level II), and crime-level (level III)[16]. Evett et al. [43] introduced a new level in the hierarchy called sub-level I or sub-source-level. Sub-source-level propositions concentrate on the source of DNA, without reference to the nature of the biological stain, such as blood, sperm, or saliva. This additional distinction may be necessary in cases involving trace quantities of DNA that cannot be associated with visible staining.

Figure 2.4 shows examples of BNs for different levels of propositions. Figure 2.4 (a) corresponds to the argument of evidence discussed in Section 2.2. It is a generic model to evaluate an item of evidence given source-level propositions. The root node F represents the proposition regarding the source of some trace (e.g. h: the fingermark comes from S; $\overline{h}$: the fingermark comes from an unknown person).

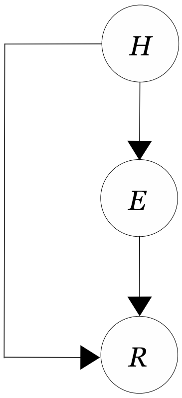
The event of interest E relates to the result of a forensic examination such as the result of a comparison between the DNA profile of a crime stain and the profile of a reference (or control) material. When the DNA profile from the specimen and the DNA profile of a reference material correspond in every marker, then this is written e, otherwise $\overline{e}$. Similarly, the fibers found in a suspect's car may either correspond (e) or not ($\bar{e}$) in substance, shape and color to fibers composing the victim's clothing. In the general context of tangible evidence, the probability b_1 expresses the rarity of an analytical feature in a relevant population. The scientist's report, represented by node R, has the same definition as that given in Section 2.2.

Figure 2.4 (b) shows a generic BN for the evaluation of transfer material given activity-level propositions. Part of it was first described in [56] and later presented in [133] in different variations. The BN focuses on material recovered on a suspect, potentially transferred from the scene or from the victim. H relates to the main proposition of interest. It always relates to a clearly defined event (e.g. 'S transported V in the trunk of his car' or 'S assaulted V with a bat'). The alternative proposition is typically derived from the statements made by the suspect (e.g. 'S tried to help V' or 'S never met V'). The proposition represented by the node C enquires about whether the suspect has been in contact with the victim, and the proposition T about whether some residue was transferred during the action (including considerations on the persistence and recovery of the residue).

Figure 2.4 (c) depicts a generic BN for the evaluation of an item of tangible evidence given crime-level propositions. A detailed discussion of this BN is available in [56, 133]. The node H enquires about the identity of the criminal, such as 'S is the offender' (h) and 'an unknown person is the offender' ($\overline{h}$). The node G is the so-called 'relevance node'. It not to be confused with the argument of relevance as discussed throughout this paper (Section 2.2). In the context of evaluations given crime-level propositions, the term 'relevance' accounts for the assumption that there is a true connection between the offender and the recovered trace material [130]. 'The residue was left by the offender during the commission of the crime' for g, and 'the residue was not left by the offender' for $\overline{g}$ are common formulations for G. Hence, the fact that we consider the possibility of presence of a material unrelated to a crime implies that a correspondence between a recovered material and a control material might be coincidental. The probability that describes this possibility is incorporated into the probability table of the node E. The assessment of relevance relies heavily on circumstantial information and on extrinsic features of the recovered material. The argument of relevance can be reduced to a single stage of reasoning and then classified according to the typology of single reasoning stages (see Appendix B.5).

The choice of the level of propositions has a substantial effect on the argument of relevance, while the argument of credibility remains unchanged. There is widespread agreement in forensic science that the higher the level of propositions, the more one needs to rely on extrinsic features, circumstantial information, and empirical studies. This is due to the fact that higher levels of propositions incorporate more uncertainties, such as T, C, and G for each of which probabilities must be assessed. Numerous publications have shown that these probabilities crucially affect the inferential force of the argument of evidence ([e.g., 4, 133]). Imagine, for instance, a case in which scientific findings are evaluated with respect to crime-level propositions, that is inference about the identity of the offender. Suppose further that the scientific findings are perfectly credible and the analytical feature examined is extremely

[16] It is worth noting that the difference between these categories or levels is not always clearcut. There is a common understanding in the field that the hierarchy of propositions should be considered as an interpretative guide and not as a rigid interpretative edit format [23].



$H:$	h		$\bar{h}$	
$E:$	e	$\bar{e}$	e	$\bar{e}$
$R: r$	a_2^h	b_2^h	$a_2^{\bar{h}}$	$b_2^{\bar{h}}$
$\bar{r}$	$1-a_2^h$	$1-b_2^h$	$1-a_2^{\bar{h}}$	$1-b_2^{\bar{h}}$

Figure. 2.5 – BN for evaluating a report R about an event E that is deemed to be relevant for an inference about a proposition H (as defined earlier in Figure 2.1. An additional arc from H to R, called 'warp', is added in order to account for a dependence of the source's credibility on the state of H. All nodes are binary. The conditional probabilities are defined in Section 2.3.7.

rare. On a probabilistic account, this would imply an extremely strong support for the proposition that recovered and control material come from the same source, rather than from different sources. However, if the recovered material was *not* left by the offender, then the result will have no inferential force at all with respect to propositions at crime-level.

2.3.7 Criterion based on the nature of the source of evidence and on the hypotheses of interest: state-dependent credibility

In some cases the credibility of a source can directly depend on the hypotheses of interest. This means that the credibility is not equal under h and $\bar{h}$. This is also known as 'state-dependent credibility of a source' [125]. Such a situation requires an additional arc from variable H to R as is depicted in Figure 2.5. Schum named this kind of additional connection 'warp'. The addition of warps can be necessary in any argument of evidence especially in testimonial evidence, where the sensitivity relative to the hypotheses may affect any attribute of the argument of credibility[17].

An important consequence of a warp is that the global inferential force of a report is no longer bounded by the inferential force of the argument of relevance. This becomes evident by considering the equation for the inferential force of the argument depicted in Figure 2.5

$$V_{r|H} = \frac{Pr(e \mid h)[Pr(r \mid e,h) - Pr(r \mid \bar{e},h)] + Pr(r \mid \bar{e},h)}{Pr(e \mid \bar{h})[Pr(r \mid e,\bar{h}) - Pr(r \mid \bar{e},\bar{h})] + Pr(r \mid \bar{e},\bar{h})} = \frac{a_1[a_2^h - b_2^h] + b_2^h}{b_1[a_2^{\bar{h}} - b_2^{\bar{h}}] + b_2^{\bar{h}}}. \tag{2.6}$$

As can be seen, the probabilities a_2 and b_2 are different under h and under $\bar{h}$ as indicated by the superscripts. Thus,

[17] An example is the a situation in which a witness gives testimony on a crime that might involve a relative or a friend of the witness.

the ratio produced by different probabilities a_2 and b_2 for a hypothesis can vary greatly from the one produced by an identical drag coefficient in the numerator and the denominator.

A further observation is worthy of emphasis: In the case where $a_2^h = a_2^{\bar{h}}$ and $b_2^h = b_2^{\bar{h}}$ we have the same model as discussed in Section 2.2.

2.4 Ontology of combinations of items of evidence

In the following sections we will extend our considerations made on single items of evidence to the combination of multiple items of evidence.

2.4.1 Criterion based on natural redundance of events

Consider the BN in Figure 2.6 (a). The two reports $R_1 = \{r_1, \bar{r}_1\}$ and $R_2 = \{r_2, \bar{r}_2\}$ given by two separate sources make claims on the same event $E = \{e, \bar{e}\}$. In Schum's notation, this is called *natural redundance* [125]. The situation depicted in Figure 2.6 (b) is different. Here, R_1 and R_2, also given by two separate sources, concern *different* events, notably $E_1 = \{e_1, \bar{e}_1\}$ and $E_2 = \{e_2, \bar{e}_2\}$. Hence, R_1 and R_2 relate to *naturally nonredundant events*.

The criteria of natural redundance of events has an impact on the argument of evidence that can be readily anticipated from the structures of the BNs of Figure 2.6. In cases where two reports concern naturally redundant events there can and *must* only be a single argument of relevance since there is only one event that is relevant to the hypotheses of interest. The argument of credibility is then a composite argument of the two reports. Cases where two sets of reports concern two different events, require – besides an argument of credibility for each report – an argument of relevance for each event also.

Assume that we receive two reports r_1 and r_2. According to Schum, the argument of evidence involving reports that concern naturally redundant events translates into a pattern of reasoning that is depicted in Figure 2.6 (a) as a BN. The inferential force of such an argument is given by

$$V_{r_1,r_2|E} = \frac{a_1 + [V_{r_1,r_2|E} - 1]^{-1}}{b_1 + [V_{r_1,r_2|E} - 1]^{-1}}, \tag{2.7}$$

where $V_{r_1,r_2|E} = V_{r_1|e} V_{r_2|E}$. The term $V_{r_1|E}$ and $V_{r_2|E}$ denote the local inferential force that each report exerts on E. It can also be seen that the inferential force of the argument of relevance is not influenced and remains defined by a_1 and b_1 as noted in Section 2.2. Irrespective of how credible the reports are, the global inferential force $V_{r_1,r_2|H}$ remains bounded by the inferential force of the argument of relevance ($V_{e|H} \geq V_{r_1,r_2|H} \geq V_{\bar{e},|H}$). The inferential force of an argument of evidence depicted in Figure 4.1 (b) presents itself differently

$$V_{r_1,r_2|H} = V_{r_1|H} \times V_{r_2|H} \tag{2.8}$$

$$= \frac{a_1^{E_1} + [V_{r_1|E_1} - 1]^{-1}}{b_1^{E_1} + [V_{r_1|E_1} - 1]^{-1}} \times \frac{a_1^{E_2} + [V_{r_2|E_2} - 1]^{-1}}{b_1^{E_2} + [V_{r_2|E_2} - 1]^{-1}}. \tag{2.9}$$

Evidence of this type has two separate arguments of evidence for each event and the corresponding report.

Several BNs involving naturally nonredundant events have been proposed in forensic literature [133]. Such networks can focus either situations involving combinations of different items of tangible evidence or situations involving combinations of different aspects of a single item of tangible evidence. The former include, for instance, the joint evaluation of results of comparative handwriting and fingermark examination, the joint evaluation of a data base search result and correspondence between the profile of crime stain, potentially left by the offender, and the profile of a person of interest. The latter situation involves the notion of *distinct components*, which was introduced by Evett et al. in 1998 [46] in the context of shoe mark evidence, where it is common to distinguish between features

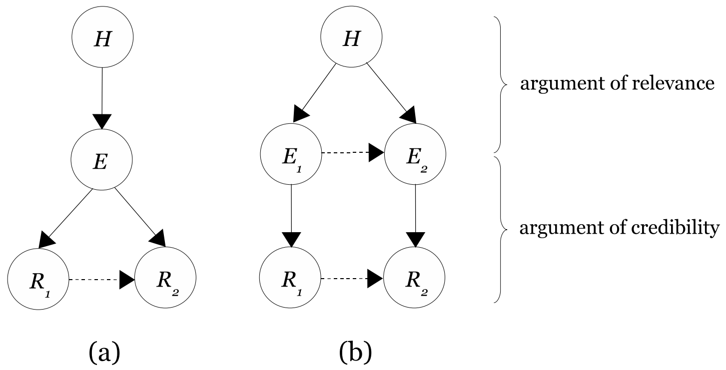


Figure. 2.6 – (a) BN for two reports from different sources (R_1, R_2) that concern naturally redundant events (i.e. E). (b) BN for two reports from different sources (R_1, R_2) that concern naturally nonredundant events (E_1 and E_2). The dashed arcs connect two separate lines of sub-arguments and are called wefts.

due to the manufacturing process (e.g. shoe size and sole pattern) and features present due to the wear (e.g. cuts and abrasions)[18].

As noted previously, two reports, which concern naturally redundant events only have a single argument of evidence. The reason for that is that if we were to assign a separate argument of relevance to each report we would compute $V_{r_1,r_2|h}$ by accounting for the relevance argument twice, although there is only one event. As a result, $V_{r_1,r_2|h}$ would not be bounded by the inferential force of the argument of relevance as is stipulated by Equation 2.10. This type of flawed reasoning is also called *double counting* [125], and can lead to a serious over- or understatement of the global inferential force.

Consider Figure 2.6 again. The dashed arcs indicate dependency relationships that may be required in some cases. Unlike warps introduced in Section 2.3.7, the dashed arcs indicated here link two separate lines of reasoning. Schum calls these kinds of arcs *wefts* [125]. Moreover, we recognize that there are two kinds of wefts. One kind connects two arguments of credibility and the other connects two arguments of relevance. The former kind of weft describes situations, where the report given by a first source, say r_1, directly affects the credibility of the second source (and vice versa). In turn, the latter kind of weft expresses the view that the knowledge on the first event, say e_1, directly affects the argument of relevance of the second event E_2 (and vice versa). The impact of the first kind of weft is not discussed here as this would exceed the scope of the paper. Readers interested in the impact of such wefts are referred to consult [125]. The impact of wefts connecting two separate arguments of relevance is considered in Section 2.4.3.

2.4.2 Criterion based on the direction of inferential force: contradicting, corroborating, conflicting, and convergent evidence

When evaluating multiple items of evidence in combination, one can distinguish between two categories of combined evidence, depending on the directions of inferential forces produced by the items. Either they all point in the same

[18] Such distinctions between different aspects of a particular type of marks are not limited to shoes. Similar observations can be made for firearms, for example. Yet another area of application are fingermarks, where forensic scientists usually make a threefold distinction with categories labelled 'level I', 'level II', and 'level III'. These refer to, respectively, the general ridge pattern, the forms of ridge configurations called 'minutiae', and the forms of pore arrangements along ridges [22].

direction (e.g. proposition h), or they point in different directions (e.g. propositions h and $\bar{h}$). In the first case, the evidence is said to be 'harmonious'. In the second case, the evidence is said to be 'dissonant'. Extending these considerations to include the concept of natural redundance, we can identify four distinct categories of evidence. These have been described by Schum [125] and are summarised in Table 2.7. Reports concerning naturally redundant events can be either contradictory or corroborating (Section 2.4.2). Reports concerning naturally nonredundant events are either conflicting or convergent (Section 2.4.2).

Contradiction and corroboration

For reports concerning naturally nonredundant events the reasoning scheme depicted in Figure 4.1 (a) applies. The overall inferential force of two reports implied by this reasoning structure, with respect to the proposition h is given by

$$V_{r_1,\bar{r}_2|H} = \frac{a_1 + [V_{r_1,\bar{r}_2|E} - 1]^{-1}}{b_1 + [V_{r_1,\bar{r}_2|E} - 1]^{-1}},$$

where $a_1 = Pr(e \mid h)$ and $b_1 = Pr(e \mid \bar{h})$. Assume that both sources $S1$ and $S2$ are credible to at least some extent so that $a_2 > b_2$ (see Figure 2.1). Now the first source $S1$ reports r_1, that e occurred (positive evidence), and the second source $S2$ reports $\bar{r}_2$ that e did not happen (negative evidence). The likelihood ratio for these contradicting reports, with respect to e, is:

$$V_{r_1,\bar{r}_2|E} = V_{r_1|E} V_{\bar{r}_2|E} = \frac{a_2^{S1}}{b_2^{S2}} \times \frac{1 - a_2^{S2}}{1 - b_2^{S2}}.$$

If the drag coefficient $D = [V_{r_1|E} V_{\bar{r}_2|E} - 1]^{-1}$ is negative, then the joint effect of the contradictory reports will favor $\bar{e}_1$. If D is positive, the contradictory reports will favor e. If, individually, both reports would support opposite events with the same inferential force, then the contradictory reports considered in combination have *no* inferential force, i.e. $V_{r_1,\bar{r}_2|E} = 1$.

Next, assume that the second source reported r_2, thus supporting e. In such a case the reports are corroborative and the joint inferential force is given by

$$V_{r_1,r_2|H} = \frac{a_1 + [V_{r_1,r_2|E} - 1]^{-1}}{b_1 + [V_{r_1,r_2|E} - 1]^{-1}}, \tag{2.10}$$

where

$$V_{r_1,r_2|E} = V_{r_1|e} V_{r_2|E} = \frac{a_2^{S1}}{b_2^{S2}} \times \frac{a_2^{S2}}{b_2^{S2}}.$$

Table. 2.7 – Four basic categories of evidence based on natural redundance criteria and on the direction of inferential force criteria

	Dissonance	Harmony
Naturally redundant	Contradiction	Corroboration
Naturally nonredundant	Conflict	Convergence

Whether a contradiction or a corroboration exists in a given situation depends on the arguments of credibility and not on whether given evidence is positive or negative *per se*. In order to find out whether evidence is corroborative or contradictory, we can employ the typology of single reasoning stages presented in Section 2.2.3 and outlined in Table 2.3. For example, if we were given two items of positive evidence r_1 and r_2, and the arguments of credibility were of type (a), $r_1 \rightarrow \bar{e}$, and of type (c), $r_2 \rightarrow e$, then the evidence is contradictory although being positive in both cases.

In this Baysesian network a special kind of redundance can arise. If one argument of credibility is of necessity, then all the other reports, be they contradictory or corroborative, become redundant. In such a case we have $V_{r_1,r_2|H} = V_{e|H}$. In a case where all arguments of credibility are contingent, redundance appears in a more moderate form as long as the evidence is corroborative. Schum named this kind of redundant evidence *corroboratively redundant* [125]. Suppose two corroborative reports r_1 and r_2 for which $a_2 > b_2$ holds (see Figure 2.1). The inferential force of the first report is ([125]):

$$V_{r_1|H} = \frac{a_1 + [V_{r_1|E} - 1]^{-1}}{b_1 + [V_{r_1|E} - 1]^{-1}}.$$

The inferential force of the second report given the first report is given by

$$V_{r_2|,r_1,H} = \frac{Pr(r_2 \mid r_1, h)}{Pr(r_2 \mid r_1, \bar{h})} \tag{2.11}$$

$$= \frac{Pr(e \mid r_1, h) + [\frac{a_2^{S2}}{b_2^{S2}} - 1]^{-1}}{Pr(e \mid r_1, \bar{h}) + [\frac{a_2^{S2}}{b_2^{S2}} - 1]^{-1}}, \tag{2.12}$$

where

$$Pr(e \mid r_1, h) = \frac{Pr(e \mid h)a_2^{S1}}{Pr(e \mid h)a_2^{S1} + Pr(\bar{e}_1 \mid h)b_2^{S1}}, \tag{2.13}$$

$$Pr(e \mid r_1, \bar{h}) = \frac{Pr(e \mid \bar{h})a_2^{S1}}{Pr(e \mid \bar{h})a_2^{S1} + Pr(\bar{e}_1 \mid \bar{h})b_2^{S1}}. \tag{2.14}$$

The ratio $Pr(e \mid r_1, h)/Pr(e \mid r_1, \bar{h})$ measures the inferential force of e with respect to h, given that we know r_1 (i.e., $V_{e|r_1,H}$). It represents the maximal inferential force that r_2 can provide at any given time. In other words, r_2 can contribute the proportion of the inferential force $V_{e|H}$ that r_1 left over. For example, suppose that the first argument of credibility is one of necessity (Section 2.2.3). That is, for r_1 we have $b_2^{S1} = 0$, so that $Pr(e \mid r_1, h) = Pr(e \mid r_1, \bar{h}) = 1$ and $V_{r_2|,r_1,H} = 1$. Thus, r_1 has absorbed all the inferential force, $V_{r_1|H} = V_{e|H}$, so that there is no inferential force left that r_2 could provide for h, irrespective of how credible r_2 is. If, however, the first report has zero inferential force, that is $a_2^{S1} = b_2^{S1}$, then $Pr(e \mid r_1, h) = Pr(e \mid h)$ and $Pr(e \mid r_1, \bar{h}) = Pr(e \mid \bar{h})$. The second report has the entire inferential force of e (i.e., $V_{e|H}$) at disposal. In summary, thus, if the first report is given by a credible source, then the second report has less inferential force than it could have when no first report would be available, or if the first report would have no credibility. Arguably, the second report is corroboratively redundant to some degree. In general, the stronger the first report, the smaller $V_{e|r_1,H}$ and the less the second report can contribute and the more it becomes corroboratively redundant.

Figure 2.7 shows a numerical example on how the values of $V_{e|r_1,H}$, $V_{r_1|H}$, and $V_{r_2|r_1,H}$ change as a function of $V_{r_1|E}$. The likelihood ratios are shown as logarithms at the base of ten. To the event E, the conditional probabilities $Pr(e \mid h) = 0.99$ and $Pr(e \mid \bar{h}) = 0.001$ have been assigned. The local likelihood ratio for the argument of relevance

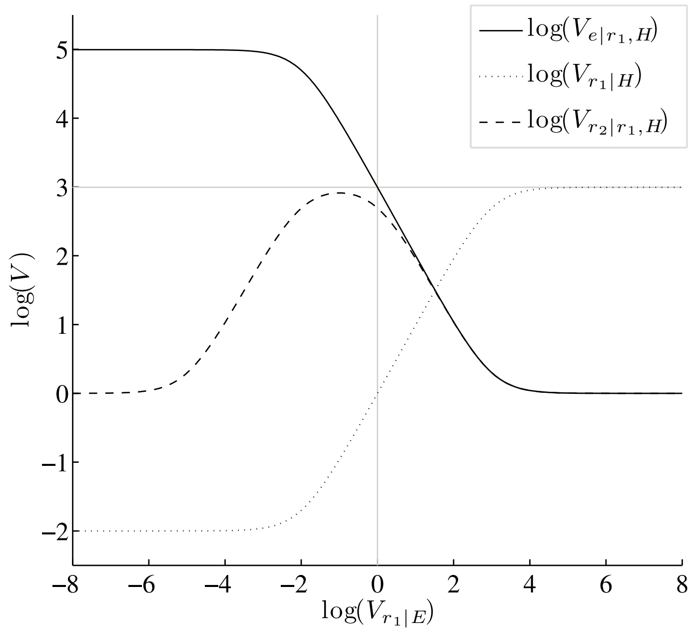


Figure. 2.7 – Graphical representation of the values of $V_{e|r_1,H}$, $V_{r_1|H}$, and $V_{r_2|r_1,H}$ as a function of $V_{r_1|E}$. The vertical grey line indicates where $Pr(r_1 \mid e) = Pr(r_1 \mid \bar{e})$ (i.e., the argument of credibility of the first report has no inferential force: $\log(V_{r_1|E}) = 0$). The horizontal grey line indicates the inferential force of the argument of relevance given that we knew for sure that e applies ($\log(V_{e|H}) = 2.9956$). The intersection of the grey lines coincides with the curve of $\log(V_{e|r_1,H})$, implying that the maximal inferential force that the report r_2 has at its disposal corresponds to the inferential force of the argument of relevance $\log(V_{e\ |H})$. Note that at any time $\log(r_1 \mid H) + \log(V_{e|r_1,H}) = \log(e \mid H)$. This follows from the fact that $V_{r_1,r_2|H}$ is bounded by the inferential force of the argument of relevance ($\log(V_{e|H}) \geq \log(V_{r_1,r_2|H}) \geq \log(V_{\bar{e}|H})$).

is, therefore, $V_{e|H} = 990$ (or $\log(V_{e|H}) = 2.9956$). The credibility of the report given by the second source is characterised by the probability assignments $Pr(r_2 \mid e) = 0.99$ and $Pr(r_2 \mid \bar{e}) = 0.001$. The local likelihood ratio relating to the argument of credibility is, thus, $V_{r_2|E} = 990$ (or $\log(V_{r_2|E}) = 2.9956$). The values of the likelihood ratio associated with the credibility of the report of the first source range from $V_{r_1|E} = 10^-8$ to $V_{r_1|E} = 10^-8$ (or $\log(V_{r_1|E}) = [-8, 8]$).

Conflict and convergence

Evidence is said to be conflicting if arguments of evidence support different hypotheses. Evidence is said to be convergent if all arguments of evidence support the same hypothesis over some alternative. Whether evidence is conflicting or converging can be examined by considering Tables 2.5 and 2.6, as long as the events E are conditionally independent given H. In such cases, Equation (2.8) applies. If the events E are conditionally dependent given H, then the following Equation applies

$$V_{r_1,r_2|H} = V_{r_1|H} V_{r_2|r_1,H}, \tag{2.15}$$

where $V_{r_1|H}$ is given in Equation (2.8) and

$$V_{r_2|r_1,H} = \frac{Pr(e_1 \mid r_1,h)[Pr(e_2 \mid e_1,h) - Pr(e_2 \mid \bar{e}_1,h)] + Pr(e_2 \mid \bar{e}_1,h) + [V_{r_2|E_2} - 1]^{-1}}{Pr(e_1 \mid r_1,\bar{h})[Pr(e_2 \mid e_1,\bar{h}) - Pr(e_2 \mid \bar{e}_1,\bar{h})] + Pr(e_2 \mid \bar{e}_1,\bar{h}) + [V_{r_2|E_2} - 1]^{-1}}. \tag{2.16}$$

Situations where the events are conditionally dependent given H allow us to further distinguish between categories of evidence based on the inferential interaction between these events. These categories of evidence are studied in the next section.

2.4.3 Criterion based on inferential interaction between events: synergistic evidence, redundant evidence, and evidence inducing directional change

A direct dependency between two separate arguments of relevance as indicated by the dashed arrow in Figure 4.1 (b) allows these two arguments to inferentially interact. Inferential interaction can be used as a criterion to distinguish between different forms of combined evidence involving naturally nonredundant events. The discussion will focus on the structural relationships between the events E_1, E_2, and H and how they affect the inferential force of the argument of relevance.

The likelihood ratio $V_{e_2|H}$ represents the inferential force that e_2 exerts on H. The likelihood ratio $V_{e_2|e_1,H}$ represents the inferential force that e_1 exerts on H, given knowledge of e_2. One can ask now in what way the knowledge of e_1 affects the inferential force of e_2: does it change the inferential force of e_2? If so, does the inferential force increase or decrease? Thus, the focus is on comparing $V_{e_2|H}$ and $V_{e_2|e_1,H}$. To do so, one can use the logarithm of the inferential forces that ought to be compared, and then normalize by $\log V_{e_2|H}$. This leads to a new term that Schum called *event redundance* $R_{e_2|e_1}$, which is the redundance of e_2 given e_1

$$R_{e_2|e_1} = \frac{\log V_{e_2|H} - \log V_{e_2|e_1,H}}{\log V_{e_2|H}} = 1 - \frac{\log V_{e_2|e_1,H}}{\log V_{e_2|H}}. \tag{2.17}$$

To analyse this expression in some further detail, it is useful to distinguish between two situations. One is that e_2 supports h over $\bar{h}$, and the other is that e_2 supports $\bar{h}$ over h. In the former case $V_{e_2|H} > 1$, and in the latter case $V_{e_2|H} < 1$. The case where $V_{e_2|H} > 1$ is examined below. Cases where $V_{e_2|H} < 1$ can be examined analogously.

If $V_{e_2|e_1,H} > V_{e_2|H}$, that is the knowledge of e_1 increases the inferential force of e_2, then $R_{e_2|e_1} < 0$ and the relationship between e_2 and e_1 is called *synergystic*. If $V_{e_2|e_1,H} = V_{e_2|H}$, that is the knowledge that e_1 does not change the inferential force of e_2, then $R_{e_2|e_1} = 0$. In other words, e_2 and e_1 are conditionally independent given H. If $V_{e_2|H} > V_{e_2|e_1,H} > 1$, then the inferential force of e_2 decreases given the knowledge of e_1, but not to the extent that e_2 starts to support $\bar{h}$ over h. In that case, e_2 is less informative and said to be *redundant* given e_1, measured by the degree of redundance $R_{e_2|e_1}$. If $V_{e_2|e_1,H} = 1$, that is given e_1, the event e_2 exerts no inferential force on H. In such a case, $R_{e_2|e_1} = 1$ and e_2 is entirely redundant. If $V_{e_2|H} > 1$ and $V_{e_2|e_1,H} < 1$, that is e_2 alone supports h over $\bar{h}$, but given e_1 the opposite hypothesis is supported, then $R_{e_2|e_1} > 1$ and indicates a *directional change*. Another way to look at directional change is to say that both arguments of relevance are conflicting within a joint evaluation, while each argument supports the same hypothesis when considered separately. An overview on how to interpret $R_{e_2|e_1}$ is given in Table D.1.

Through consideration of $R_{e_2|e_1}$, one can distinguish between arguments of relevance that are synergistic, conditionally independent, redundant, and imply a directional change. It follows that there are combinations of evidence of naturally nonredundant events that fall into these categories.

Table. 2.8 – Relationship between the measure $R_{e_2|e_1}$ and the inferential force of e_2.

$R_{e_2\|e_1}$	e_2 supports h over $\bar{h}$ ($V_{e_2\|H} > 1$)	e_2 supports $\bar{h}$ over h ($V_{e_2\|H} < 1$)	Type of interaction
$R_{e_2\|e_1} < 0$	$V_{e_2\|e_1,H} > V_{e_2\|H}$	$V_{e_2\|e_1,H} < V_{e_2\|H}$	Synergy
$R_{e_2\|e_1} = 0$	$V_{e_2\|e_1,H} = V_{e_2\|H}$	$V_{e_2\|e_1,H} = V_{e_2\|H}$	Cond. independence
$1 > R_{e_2\|e_1} > 0$	$V_{e_2\|H} > V_{e_2\|e_1,H} > 1$	$V_{e_2\|H} < V_{e_2\|e_1,H} < 1$	Redundance
$R_{e_2\|e_1} = 1$	$V_{e_2\|e_1,H} = 1$	$V_{e_2\|e_1,H} = 1$	Redundance
$R_{e_2\|e_1} > 1$	$V_{e_2\|e_1,H} < 1$	$V_{e_2\|e_1,H} > 1$	Directional change

2.5 Kinship among different manifestations of evidence

Some criteria seem to entail a structural variation in the basic argument of evidence (nature of the source, hypotheses of interest, natural redundance, and state-dependency of credibility), while this is not the case for other criteria (occurrence/nonoccurrence, necessity/contingency, reversibility, relevance, direction of inferential force, inferential interaction between naturally nonredundant events). The first type of criteria is *inferentially prescriptive*. This means that invoking such a criterion provides a binding definition of how to structurally arrange an argument of evidence in terms of credibility and relevance. The second type of criteria is *inferentially descriptive*. Such criteria clarify distinctions that arise within a structure of an argument of evidence. It appears that distinctions produced by such criteria are a result of the types of reasoning stages (Section 2.2.3) that are employed within a given argument structure, and of the inferentially prescriptive criterion that suggested the argument structure in the first place. This implies a hierarchical relationship between inferentially prescriptive criteria and inferentially descriptive criteria in the sense that an assessment of the latter requires a prior specification of the former. Figure 2.8 depicts a hierarchical classification of the different criteria of the manifestations of evidence given the distinction between inferentially prescriptive and inferentially descriptive criteria. It elicits, at the same time, the argument that is affected by a criterion.

Inferentially descriptive criteria themselves can further be subdivided into two classes. The first class of inferentially descriptive criteria covers all the criteria that relate to specific configurations of probability values assigned in a given argument. It is denoted as 'Probability assignment' in Figure 2.8 and comprises the criteria of necessity and contingency (Section 2.3.2), reversibility (Section 2.3.3), direction of the inferential force (Section 2.4.2) and inferential interaction between naturally nonredundant events (Section 2.4.3). The criterion based on relevance regarding direct and circumstantial evidence can be considered as a result of the criterion based on necessity and contingency applied to the argument of relevance as pointed out in Section 2.3.4. The criterion based on relevance is therefore depicted as a subordinate of the more fundamental criterion based on necessity and contingency. The criterion based on inferential interaction between naturally nonredundant events (Section 2.4.3) is an inferentially descriptive criterion based on the probability assessment and at the same time a subordinate of the inferentially prescriptive criterion based on natural redundance (Section 2.4.1). This kinship structure accounts for the fact that the notion of natural redundance is a conceptual prerequisite for the criterion based on inferential interaction between naturally nonredundant events.

The second class of inferentially descriptive criteria is based on the actual observation or – to use the more technical term used in the context of BNs – variable instantiation. As can be seen from Figure 2.8, the criterion based on the occurrence or nonoccurrence of events (Section 2.3.1) is the only inferentially descriptive criterion based on variable instantiation in this paper.

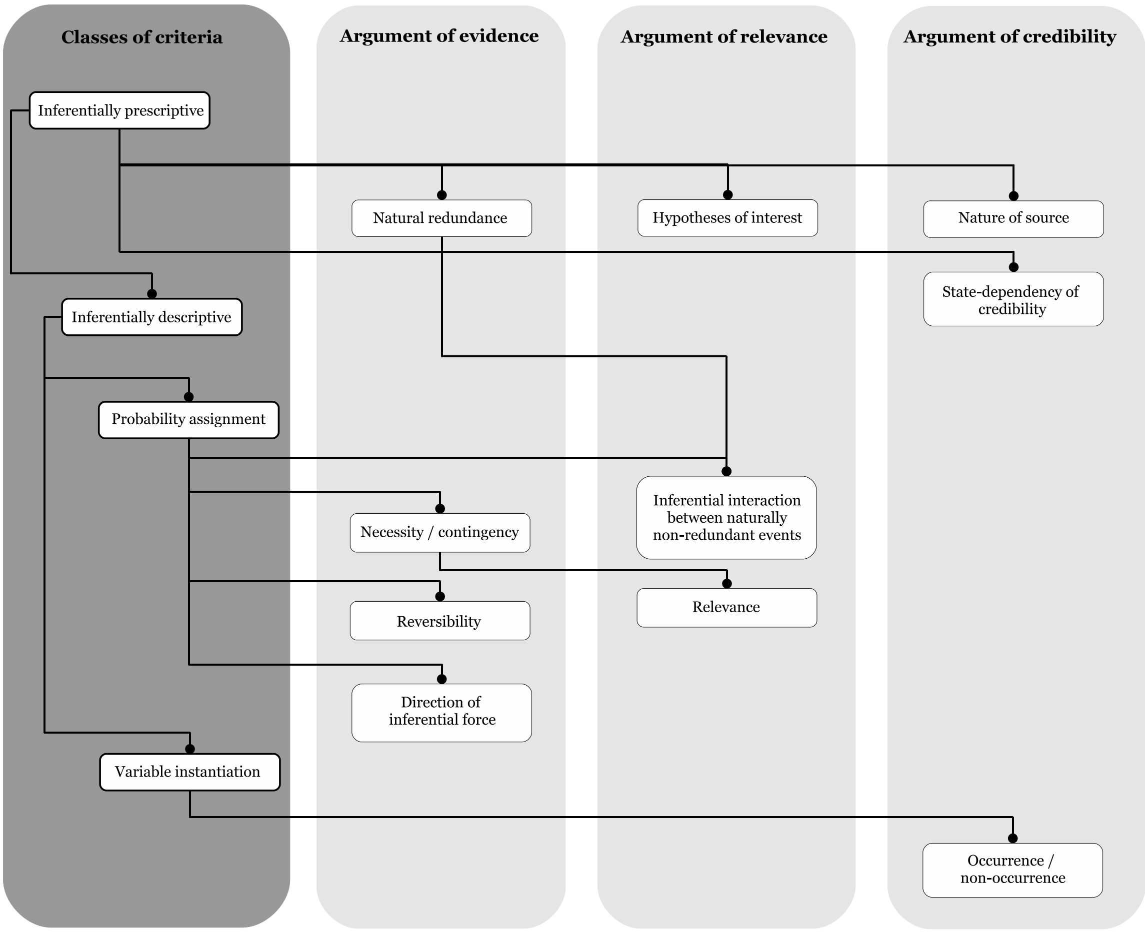


Figure. 2.8 – Hierarchical relationship among different criteria that give rise to different manifestations of evidence. A classification of different criteria is shown on the far left. The criteria themselves are shown in the three following columns in relation to the argument they affect.

2.6 Conclusion

"Some persons have grown very old in their attempts to find comprehensive categorizations of various subjects. (...) In my present attempt to categorize evidence, I may fare no better (I am certainly growing older)." [125, p.115] As recognised by Schum, finding a comprehensive categorisation of evidence is a far reaching attempt, and was not the intention of this paper. In view of Schum's foundational works on categorizing evidence, the ontology proposed in this paper is incomplete. Notwithstanding, it represents a complementary perspective with extensions to the existing bulk studies realised by Schum and Kind so far.

Single reasoning stages (Section 2.2.3) and by extension, Bernoulli's four cases of evidence and their sub-cases (Section 2.3.2) provide an insightful framework to accommodate a broad range of arguments of evidence. Reversible evidence, direct evidence and circumstantial evidence, testimonial evidence, source-level evaluation of DNA profiling results, evaluation of tangible evidence given crime-level propositions, as well arguments of evidence involving convergence and conflict (as long as they contain no wefts) can be described and classified in terms of Bernoulli's four cases and their sub-cases. Arguments of evidence involving naturally redundant events can be classified into corroboration or contradiction based on the typology of single reasoning stages as long as they contain no wefts. This shows that the understanding of the simplest arguments of evidence is crucial for the understanding of more complex arguments.

As argued in Section 2.5, inferentially prescriptive criteria precede inferentially descriptive criteria. It may, therefore, be helpful for future research on categorisation of evidence to examine whether a newly identified inferentially prescriptive criterion can give rise to further inferentially descriptive distinctions.

All the inferentially prescriptive criteria regarding single items of evidence encountered in this paper affect either the argument of credibility or the argument of relevance, but never both. A question that naturally emerges at this point is whether there exists an inferentially prescriptive criterion that affects both sub-arguments and if so, to identify the prerequisite(s) that allow that particular criterion to do so.

Inferentially descriptive criteria could be further refined into those that are based on particular configurations of probability assignments and those that are based on the variable instantiation. In this paper, we have only discussed the criterion based on the occurrence and nonoccurrence of an event as a member of the latter descriptive criterion.

Acknowledgements

The authors are grateful to Professor Paolo Garbolino from the University Iuav of Venice for proofreading the present research.

This work has been supported by the Swiss National Science Foundation [grant number 100014_150276].

3 Graphical probabilistic analysis of the combination of items of evidence

ABSTRACT. Unlike the evaluation of single items of scientific evidence, the formal study and analysis of the joint evaluation of several distinct items of forensic evidence has, to date, received some punctual, rather than systematic, attention. Questions about the (i) relationships among a set of (usually unobservable) propositions and a set of (observable) items of scientific evidence, (ii) the joint probative value of a collection of distinct items of evidence as well as (iii) the contribution of each individual item within a given group of pieces of evidence still represent fundamental areas of research. To some degree, this is remarkable since both forensic science theory and practice, yet many daily inference tasks, require the consideration of multiple items, if not masses of evidence. A recurrent and particular complication that arises in such settings is that the application of probability theory, that is the reference method for reasoning under uncertainty, becomes increasingly demanding. The present paper takes this as a starting point and discusses graphical probability models, that is Bayesian networks, as a framework within which the joint evaluation of scientific evidence can be approached in some viable way. Based on a review of existing main contributions in this area, the article here aims at presenting instances of real case studies from the author's institution in order to point out the usefulness and capacities of Bayesian networks for the probabilistic assessment of the probative value of multiple and interrelated items of evidence. A main emphasis is placed on underlying general patterns of inference, their representation as well as their graphical probabilistic analysis. Attention is also drawn to inferential interactions, such as redundancy, synergy and directional change. These distinguish the joint evaluation of evidence from assessments of isolated items of evidence. Together, these topics present aspects of interest to both domain experts and recipients of expert information, because they have a bearing on how multiple items of evidence are meaningfully and appropriately set into context.

Keywords. Bayesian networks · combining items of evidence · likelihood ratio

3.1 Introduction

3.1.1 Graphical approaches to judicial inference modeling

Legal disciplines largely, including more specialist areas such as forensic science, are characterised by inference problems that involve a variety of events among which distinct relationships, affected by uncertainty, are assumed to hold. The use of visual representation schemes for analysing, depicting and communicating such aspects, typically occurring in relation with legal cases, has a remarkably long history. An approach frequently quoted in this context is that of Wigmore, which relies on an extensive, however non-probabilistic and non-quantitative, hierarchical representation system for capturing the potentially large range of issues involved in legal cases [145, 146]. Recent decades have seen the introduction of modern graphical probabilistic networks, in particular *Bayesian networks*. This concept has been studied for the analysis of such complex and historically famous cases like the Collins case [38], the Sacco and Vanzetti case [79, 71], the Omar Raddad case [94] and the O.J. Simpson trial [137]. Bayesian networks will be introduced more formally later in Section 3.3. At this point we solely note that 'Bayesian', useable both as an adjective and as a noun, is a notion that stems from a theorem – Bayes' theorem[1] – that is a logical consequence of the basic rules of probability and related concepts. It is a result that helps to understand how to treat new evidence. In turn, the term 'network' is taken to refer here to a set of nodes and arcs that stand for, respectively, propositions[2] of interest and assumed relationships between propositions. Such a graph, or network, provides an abstract, but rigorous, representation of one's view of and attitude towards an inferential problem.

A main guiding idea of studies in graphical inference modeling, using Bayesian networks, is that they allow one to capture rigorously core aspects of situations of reasoning under uncertainty. That is, as a kind of graphical model based upon concepts from graph and probability theory, a Bayesian network's graphical structure accounts for possible relevance relationships between the various aspects or events of a problem domain under investigation. The underlying probabilistic architecture of these models expresses beliefs about the strengths of the assumed relationships. Newly acquired information about a given inference problem – in legal contexts typically referred to as evidence – can then be used to update the probability of the various specified uncertain propositions. Bayesian networks operate this updating according to Bayes' theorem, the fundamental rule for assessing the discriminative value of evidence in forensic science [117, 116]. During the past 20 years, there has been a regular stream of publications on the use of Bayesian networks in forensic and legal theory and practice. These contributions converge in their opinions that Bayesian networks provide valuable assistance to their user in coping with inferential issues that are marked by uncertainty.

In the particular context of forensic science too, Bayesian networks have, since their development in the area of artificial intelligence in the early 1980s, soon found their applications. Aitken and Gammerman [2], followed by Dawid and Evett [30], were among the first to show how probabilistic network-guided thinking and analysis can support the organisation and implementation of an evaluative framework that is not restricted to single items of forensic evidence, but naturally extends to one that allows for a combination of evidence from a variety of sources. The examples provided by these authors include scientific evidence in the form of fibers and bloodstains, described in the context of well defined cases.

These formative studies have subsequently led to further developments that pertain to more generic forensic case settings. Garbolino and Taroni [56], for example, proposed Bayesian network models for fibre scenarios with a particular emphasis on general patterns of inference, such as evidential relevance and potential innocent transfer (i.e., legitimate contact). In turn, Evett et al. [43] described Bayesian network approaches in order to deal with particular complications encountered in connection with DNA profiling analyses applied to small quantities of DNA. Yet further developments in rather specific contexts, such as the (reciprocal) cross-transfer of materials between two persons (or objects), have been proposed by Aitken at al. [5].

Generally, the use of Bayesian networks for evaluating DNA profiling results represents a particularly lively area of research, to which Dawid et al. [31] have contributed a seminal paper. It proposes a methodology for deriving appropriate Bayesian network structures from initial pedigree representations of forensic identification problems. This approach has subsequently been used in a series of other works that concentrated on selected aspects of the assessment of forensic DNA evidence. Mortera et al. [104, 103], for example, have studied this Bayesian network modeling approach for mixed DNA traces, including a discussion of issues such as missing individuals and silent alleles. Following the same ideas, Bayesian network models have also been proposed for situations in which (i) the alternative proposition is that a close relative of the suspect left the crime stain (in agreement with a probabilistic

[1] Although Bayes' theorem has about a 250-year history, the attribute 'Bayesian' as a descriptor of a particular class of inference methods appears to have gained more widespread use only since the middle of the twentieth century [51].

[2] In a rather broad sense, the term proposition is interpreted here as a statement or assertion that such-and-such is the case (e.g., an outcome or a state of nature). It is assumed that personal degrees of belief can be assigned to it. At times, we may also use the term 'hypothesis', which we use interchangeably.

approach previously described by Evett [40]), or (ii) multiple propositions need to be considered (e.g., that the crime stain comes from a brother of the suspect or an unrelated member of the suspect population) [131]. A further topic approached in the book by Taroni et al. [131] is that of partial matches, that is a situation in which a suspect matches a crime stain only partially and when a proposition of interest is that a close relative of the suspect, such as a brother, is the source of the crime stain.

More recently, the modeling approach of Dawid et al. [31] has been extended to the object-oriented Bayesian network environment [32]. It is based on the idea of defining generic 'classes' of networks, parts of which can be used, as required, within other networks. This allows one to describe inference problems in terms of interrelated objects, and to structure them hierarchically at different levels of abstraction. An advantage of this property is that it is well adapted for supporting human reasoning, which tends naturally to proceed in terms of hierarchies of abstractions, in particular where it is difficult to mentally capture all aspects of a problem simultaneously.

3.1.2 Contents and aims of this paper

Much of the formalised graphical analyses mentioned in the previous section involve substantial probabilistic analyses. This may distract the view from the fact that there are some very general patterns of evidential reasoning that can be set forth and discussed without reference to a particular graphical probability modeling framework. This is illustrated, for example, by Schum's [125] graphs for evidence analysis (also used in Kadane and Schum [79]). Unlike Bayesian networks, that are directed in a top-down mode from uncertain target propositions to observational instances (i.e., evidence nodes), so-called 'Schum graphs', as they will be termed here, work in the other direction. That is, they are modeled in a bottom-up way in the the same way as Wigmore's [145] reference charts. It is the belief of the authors here that these graphs can be instructive to become acquainted with the general idea of graphical representations for relationships among variables that are retained in a formal probabilistic analysis. Stated otherwise, the probabilistic inference steps undertaken by a reasoner when proceeding from one item of evidence to another can be reflected through a graphical display. The ways in which such inferential steps are taken in different situations allows one to shed light on some generic patterns of combination. This insight also helps to recognise viable graphical structures for more advanced graphical modeling formalisms, such as Bayesian networks. It is for this reason that the paper here will start, in Section 3.2, with a presentation of elements of Schum's [125] foundational theory on probabilistic evidence combination. The aim at this point is to discuss, using several examples, the relevance of these influential works for joint inference analyses encountered in forensic contexts. The chosen examples will also show, however, that even basic model structures, involving only a few nodes and a light connective structure, can readily lead to extended formulae for probabilistic calculations of evidential value. Section 3.3 will take this as an instance to illustrate a particular feature of Bayesian networks, that consists in their underlying probabilistic architecture. It constitutes an integral part of this class of graphical models and distinguishes it from Schum graphs. This means, stated otherwise, that in Bayesian networks, probabilistic calculations can be confined to the model while the user can concentrate efforts on model elicitation and structuring. Section 3.4 will set these arguments into context by presenting and analysing aspects drawn from two real cases involving scientific evidence. These case studies will also serve the purpose of discussing the nature of inferential interactions that may arise when the analysis of evidential value is extended beyond single and isolated items of evidence. A general discussion and conclusions are presented in a final Section 3.5.

The intention to approach uncertainty in evidence evaluation through probability requires the consideration of some notational convention. Besides, it is also necessary to accept main elements and results of probability theory. These aspects will not be reproduced here in much detail essentially because extensive literature on this topic now is widely available [e.g., 116, 118, 4]. Throughout this paper, in particular in Section 3.3, notation commonly used in forensic literature will be followed. Although some of the formulaic expressions may be perceived as difficult to apprehend, it is important to emphasise that this should not be taken as a deficiency of the proposed formal framework – that is graph and probability theory – but rather as a consequence of the level of difficulty associated with the real-world problems that are being addressed. This viewpoint has already been put forward, elsewhere in literature, by Friedman [53] in a discussion on the relevance of Bayesian reasoning applied to realistic settings: “If applied to take into account all the information we have about a situation, Bayesian analysis requires unrealistically complex calculations, but this does not suggest a problem with the theory. On the contrary, the complexity is in the world surrounding us, and the theory would have limited value if it could not in principle represent that complexity. Probability is a flexible template. It can take into account as much complexity as its user is able to handle.” [53, at p. 1818]

The case studies proposed in this paper involve footwear mark evidence. This category of scientific evidence was also involved in the recent judgment of the Court of Appeal in R v T [106]. In particular, debates involved the handling of aspects such as general pattern and size, as well as the rarity of such descriptors and how they may be informed by data. In a wider sense, this touches on questions of more fundamental importance, which go beyond the instances debated within this particular judgment. Forensic scientists need a clear view of the inferential issues that are associated with particular items of evidence, and questions of the combination of evidence are an essential aspect of this. Such combinations may be required within a given item of evidence (i.e., combination of distinct aspects of a given item of evidence), but questions may also extend to problems of relating several separate items of evidence. Graphical probability modeling, and analyses based on such models, may help scientists to refine their understanding of the various evaluative issues that are involved in a given case. The suggestion at this juncture, and throughout this paper in general, is not, however, that graphical models ought to be part of or substitute a scientist’s (written) report. Their primary role could be that of assisting the logical reasoning, discussing and clear drafting of reports. This represents an important preliminary for coherent, concise and informed reporting on forensic examinations. It is thus thought to offer support when addressing the challenging task of communicating scientific evidence clearly and convincingly, so as to favour the correct understanding of the meaning of evidence among recipients of expert information.

3.2 ‘Schum graphs’ for evidence analysis

3.2.1 Preliminaries

When reasoning about an item of evidence, one may find that it favours a certain hypothesis rather than others. This can be conceptualized as the ascription of an inferential vector to a given item of evidence. Such an inferential vector can be characterized by two main aspects, that is an inferential direction and an inferential force. When extending this idea to practical reasoning, however, one comes to realise that one will be required to consider several items of evidence and this will generate a whole batch of such vectors.

With respect to inferential directions, two situations can be distinguished. Either the inferential vectors will point towards more than one hypothesis or they point towards one unique hypothesis. Following Schum [125], the first situation is said to involve *dissonant evidence* whereas the second situation involves *harmonious evidence*. The probabilistic underpinnings of these distinctions are considered hereafter in some further detail.

3.2.2 Dissonant evidence: Contradiction and conflict

All dissonant evidence incorporates an inferential divergence, although only some situations of dissonance can properly be called contradictory. Schum [125] considers dissonant evidence that is not contradictory as ‘being in conflict’. Properly speaking, a *contradiction* is given only if the occurrence of mutually exclusive events are reported. In order to clarify this, let us say that source S_1 states E^*, that is ‘Event E occurred’. A second source S_2 states E^{c*}, that is ‘Event E did not occur’.

Example 1. – In a case involving questioned documents it may be of interest to learn something about the proposition E, that a given suspect wrote a signature on a handwritten document. Denote by E^c the proposition

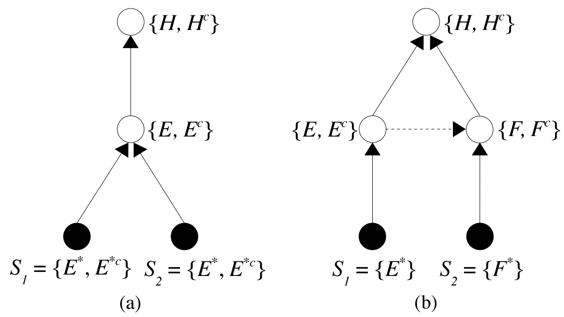


Figure. 3.1 – Generic models for (a) contradictory and corroborative inference and (b) conflicting and converging inference. The dotted arrow applies whenever one assumes a dependency between the two events $\{E,E^c\}$ and $\{F,F^c\}$ conditional upon $\{H,H^c\}$. Notice that these graphical models do not represent Bayesian networks.

that the suspect did not write the questioned signature. One cannot directly know whether or not the suspect is the author of the questioned signature. One may therefore rely on an opinion presented by, for example, an eyewitness. Let this source of information be denoted by S_1 and the report given by this source in terms of E^*, that is a statement that E occurred. Next, one may also have a further source of information, denoted by S_2. This source, too, reports about the proposition E, but affirms that its complement, E^c, holds. An example for such a second source of information could be another eyewitness or a forensic document examiner.[3]

Given this outset, a question of interest may be how to draw an inference about a pair of ultimate propositions H and H^c, while allowing uncertainty about the true state of the intermediate variable E. For the example introduced above, the variable H could be, for example, the commission of a fraud, or another criminal activity, which requires the establishment of authorship of the questioned signature at hand.

A common way to approach such a question relies on a likelihood ratio (LR), that is a fraction of two likelihoods, each of which expresses the probability of obtaining a certain outcome, here the evidence $\{E^*,E^{c*}\}$, given a proposition of interest. Applied to the situation here, one would thus focus on the probability of the two reports, $\{E^*,E^{c*}\}$, given that H holds, and compare this assignment to that made under the assumption that H^c holds. More formally, this is written as follows:

$$LR_{E^*,E^{c*}} = \frac{Pr(E^*,E^{c*} \mid H)}{Pr(E^*,E^{c*} \mid H^c)} \; . \tag{3.1}$$

There is now a broad agreement among legal and forensic researchers and practitioners that this fraction - independently of the level at which propositions are formulated - represents the key element for reasoning processes that seek to evaluate propositions in judicial contexts [e.g., 118, 3]. In particular, it is recognised that legal reasoning can be reconstructed as inferences in accordance with Bayes' theorem. That is, for updating odds in response to evidence, one needs to assess the probability of that evidence relative to each of the two competing propositions and then compare the resulting likelihoods. If the ratio of the likelihoods is one, then the evidence would be said to be neutral, that is, it would leave the prior odds unchanged. Likelihood ratios greater (or smaller) than one would be said to favour H over H^c (H^c over H).

Assuming a relationship of dependence between the variables as shown in Figure 3.1 (a), the likelihood ratio in

[3]Forensic document examiners commonly present (categoric) opinions about selected propositions of interest [84]. There is ongoing debate about whether or not this is helpful for guiding courts in assessing the probative value of scientific evidence [132]. This relates to the wider topic of how to actually present results of forensic examinations. This topic is beyond the scope of this paper.

Equation (3.1) can be presented in some further detail, as proposed in [125]:[4]

$$LR_{E^*,E^{c*}} = \frac{Pr(E^*,E^{c*} \mid H)}{Pr(E^*,E^{c*} \mid H^c)} = \frac{Pr(E \mid H) + [\frac{h_1 m_2}{f_1 c_2} - 1]^{-1}}{Pr(E \mid H^c) + [\frac{h_1 m_2}{f_1 c_2} - 1]^{-1}} \,. \tag{3.2}$$

Here, $h_1 = Pr(E^* \mid E)$, $m_2 = Pr(E^{c*} \mid E)$, $f_1 = Pr(E^* \mid E^c)$ and $c_2 = Pr(E^{c*} \mid E^c)$. The extended form of the likelihood ratio shown in Equation (3.2) is reproduced here because it contains the expression $[(h_1 m_2/f_1 c_2) - 1]^{-1}$. This part of the formula is also referred to as "drag coefficient", as it acts like a drag upon LR_E, that is the quantity of inferential force that E exerts towards $\{H, H^c\}$. As will later be pointed out in a separate Section 3.2.4, the drag coefficient accounts for the credibility of the statements made by the sources of interest. In particular, it will determine the degree to which $LR_{E^*,E^{c*}}$ will approach the value of LR_E.[5]

The result shown in Equation (3.2) can be further understood by considering local likelihood ratios for drawing an inference about E, on the basis of the distinct items of evidence E^* and E^{c*}. More specifically, there is, respectively, a likelihood ratio for item of evidence E^*, written LR'_{E^*}, and one for the item of evidence E^{c*}, written $LR'_{E^{c*}}$:

$$LR'_{E^*} = \frac{Pr(E^* \mid E)}{Pr(E^* \mid E^c)} = \frac{h_1}{f_1}, \;\; LR'_{E^{c*}} = \frac{Pr(E^{c*} \mid E)}{Pr(E^{c*} \mid E^c)} = \frac{m_2}{c_2}.$$

A prime ('$'$') is used here to indicate that the likelihood ratio concentrates on an inference to E only, rather than to the ultimate proposition H.

When taking the inverse of the latter likelihood ratio, then one has an expression of the degree to which E^{c*} favours E^c: $LR'^{-1}_{E^{c*}} = Pr(E^{c*} \mid E^c)/Pr(E^{c*} \mid E) = c_2/m_2$. It can now be seen that the overall support of the two statements $\{E^*, E^{c*}\}$ for E depends on the relative magnitude of LR'_{E^*} and $LR'_{E^{c*}}$. In particular, in all the cases where $(h_1/f_1) > (c_2/m_2)$, the evaluation of the statements will strengthen the proposition E. Conversely, if $(h_1/f_1) < (c_2/m_2)$, then the other proposition, E^c, will be favored.

More generally, notice further that the global inferential force $LR_{E^*,E^{c*}}$ is bound by LR_{E^c} and LR_E so that $LR_{E^c} \leq LR_{E^*,E^{c*}} \leq LR_E$. That is, LR_E represents the capacity of E to discriminate between H and H^c, given by $Pr(E \mid H)/Pr(E \mid H^c)$, whereas LR_{E^c} that of E^c, given by $Pr(E^c \mid H)/Pr(E^c \mid H^c)$. But usually, one will not have confirmed knowledge of the occurrence of either E or E^c, only evidence in the form of the statements $\{E^*, E^{c*}\}$.

Situations of evidence in *conflict* are different as they imply events that are not mutually exclusive. This is pointed out in Figure 3.1 (b). For this model, suppose that source S_1 states E^*, that is the occurrence of event E, which is one that favours the proposition H. A second source, S_2, states F^*, that another proposition F, favoring proposition H^c, occurred. The example given hereafter illustrates this outset.

Example 2. – Consider again, as in Example 1 given above, a report E^* that event E occurred, that is, a given suspect wrote a signature on a questioned document. Imagine further that the questioned document bears ridge skin marks (i.e., 'fingermarks'). Let F denote the proposition according to which the fingermarks come from some person other than the suspect and let F^* denote a scientist's report of such a conclusion.[6] Conversely, let F^c denote the proposition according to which the fingermarks come from the suspect. Assuming that the fingermarks are found in a position (on the document) where marks from the author of the crime of interest would be expected to be found, the proposition F can be considered relevant in an inference about the proposition H, that is 'the suspect is the author of the fraud'. Clearly, proposition F would favour H^c here because the probability of F can reasonably be taken to be greater given H^c than given H. That is, stated otherwise, the likelihood ratio for F,

[4] Additional information about the derivation of this result is provided in C.1.

[5] The likelihood ratio LR_E describes the inference about H on the basis of the intermediate variable E and is given by the fraction of the two likelihoods $Pr(E \mid H)$ and $Pr(E \mid H^c)$.

[6] Again, there is no suggestion at this point that this is an appropriate way of reporting conclusions from fingermark analyses, although we concede that probably it still represents, currently, the most widespread practice.

written $LR_F = Pr(F \mid H)/Pr(F \mid H^c)$, is smaller than 1. This represents support for H^c. In turn, the proposition E, which relates to the authorship of the questioned signature, provides support for H. In fact, following Example 1, the likelihood ratio for E is $LR_E = Pr(E \mid H)/Pr(E \mid H^c) > 1$.[7]

In this example, the evidential values of the reports E^* and F^* by respectively, source S_1 and source S_2 are given by:

$$LR_{E^*} = \frac{Pr(E^* \mid H)}{Pr(E^* \mid H^c)} = \frac{Pr(E^* \mid E)Pr(E \mid H) + Pr(E^* \mid E^c)Pr(E^c \mid H)}{Pr(E^* \mid E)Pr(E \mid H^c) + Pr(E^* \mid E^c)Pr(E^c \mid H^c)}, \quad (3.3)$$

$$LR_{F^*} = \frac{Pr(F^* \mid H)}{Pr(F^* \mid H^c)} = \frac{Pr(F^* \mid F)Pr(F \mid H) + Pr(F^* \mid F^c)Pr(F^c \mid H)}{Pr(F^* \mid F)Pr(F \mid H^c) + Pr(F^* \mid F^c)Pr(F^c \mid H^c)}. \quad (3.4)$$

These individual likelihood ratios suppose a conditional independence[8] upon knowledge of the ultimate proposition H. In addition, they incorporate uncertainty about the actual − but unobserved − state of the events E and F. This is achieved by writing a given report, for example E^*, conditioned on both E and E^c, weighted by the probability of, respectively, E and E^c.

The two likelihood ratios, Equations (3.3) and (3.4), can also be written in a more compacted form [125]:

$$LR_{E^*} = \frac{Pr(E \mid H) + [\frac{h_1}{f_1} - 1]^{-1}}{Pr(E \mid H^c) + [\frac{h_1}{f_1} - 1]^{-1}}, \quad (3.5)$$

$$LR_{F^*} = \frac{Pr(F \mid H) + [\frac{h_2}{f_2} - 1]^{-1}}{Pr(F \mid H^c) + [\frac{h_2}{f_2} - 1]^{-1}} \quad (3.6)$$

where $h_1 = Pr(E^* \mid E)$, $f_1 = Pr(E^* \mid E^c)$, $h_2 = Pr(F^* \mid F)$ et $f_2 = Pr(F^* \mid F^c)$. As may be seen, the fractions h_1/f_1 and h_2/f_2 represent the evidential values − that is the likelihood ratios − of the reports E^* and F^* for discriminating between the states of the individual events E and F.

Given the stated assumption of conditional independence, the overall evidential value of the two reports E^* and F^*, that is LR_{E^*,F^*}, is given by the product of the individual likelihood ratios: $LR_{E^*,F^*} = LR_{E^*} \times LR_{F^*}$. For the currently discussed Example 2, such an assumption seems reasonable. In fact, ridge skin surface characteristics can be considered to be independent of handwriting characteristics.

If, however, in a more general case the events $\{E, E^c\}$ and $\{F, F^c\}$ need to be considered as not conditionally independent upon $\{H, H^c\}$, then the overall likelihood ratio will be of the form $LR_{E^*} \times LR_{F^*|E^*}$. That is, the likelihood ratio for the second report, F^*, is conditioned upon knowledge of the first report, E^*. More formally, this is written as $LR_{F^*|E^*}$. While LR_{E^*} is as defined above in Equation 3.3, the term $LR_{F^*|E^*}$ involves a more extended development that can be shown to reduce to [125]:

$$LR_{F^*|E^*} = \frac{Pr(E \mid E^*, H)[Pr(F \mid E, H) - Pr(F \mid E^c, H)] + Pr(F \mid E^c, H) + [\frac{h_2}{f_2} - 1]^{-1}}{Pr(E \mid E^*, H^c)[Pr(F \mid E, H^c) - Pr(F \mid E^c, H^c)] + P(F \mid E^c, H^c) + [\frac{h_2}{f_2} - 1]^{-1}}. \quad (3.7)$$

Here $h_2 = Pr(F^* \mid F)$ and $f_2 = Pr(F^* \mid F^c)$. These latter two terms represent, respectively, the numerator and denominator of a local likelihood ratio LR'_{F^*} that expresses the degree to which the report F^* discriminates between the intermediate propositions F and F^c.

[7] Further discussion of such an example, using Bayesian networks and a consideration of multiple propositions, can also be found in Taroni and Biedermann [132].

[8] Conditional independence describes, broadly speaking, a setting in which the truth or otherwise of a proposition would not affect one's belief in another proposition, given that a third proposition is already known. It is a property that characterises one's system of beliefs, if one maintains, for example, the following: once proposition C is known, one's belief in proposition A would not be altered if, in addition, proposition B would be known. More formally stated, a variable A is said to be conditionally independent of B, given C, if and only if $Pr(A \mid B, C) = Pr(A \mid C)$ for all the states of A, B and C.

There is a close relationship that one can observe with respect to the previous Equation (3.6). In fact, when E is irrelevant for the assessment of F conditional on H, then the latter Equation (3.7) reduces to the former Equation (3.6). That is, more formally expressed, when knowledge of E is irrelevant, then

$$Pr(F \mid E,H) = Pr(F \mid E^c,H) = Pr(F \mid H) \text{ and } Pr(F \mid E,H^c) = Pr(F \mid E^c,H^c) = Pr(F \mid H^c)$$

hold, and this eliminates the product in the numerator and the denominator of the likelihood ratio $LR_{F^*|E^*}$.

3.2.3 Harmonious evidence: Corroboration and convergence

Schum [125] distinguishes two main cases of harmonious evidence, notably corroborating evidence and convergent evidence. The former, *corroboration*, applies to evidence from sources that state the occurrence of the same event. As illustrated by Example 3 below, consider two sources S_1 and S_2 that each state E^*, that event E occurred. Suppose further that $Pr(E \mid H) > Pr(E \mid H^c)$, that is event E is one that is more probable to occur if the ultimate probandum H is true, rather than when the specified alternative, H^c, is true. Using notation introduced so far, this expression of evidential value can also be written as LR_E.

Example 3. – An illustration of a setting in which evidence is corroborating can be obtained by modifying the previous Example 1. When assuming two independent handwriting experts, that each report E^*, that is the proposition E defined as 'the suspect is the source of the signature on the questioned document', evidence from two distinct sources is available. In such a setting, each expert reports the occurrence of the *same* event. In turn, the proposition E is relevant in an inference about H, that is the proposition according to which the suspect is the author of a given criminal event of interest.

By supposing a relation of dependence between the variables as shown in Figure 3.1 (a), the likelihood ratio for the reports E_1^* and E_2^* by, respectively, source S_1 and S_2, follows the general structure defined earlier in Equation (3.1). For the case considered here, the expression can again be developed further and shown to be as follows [125]:

$$LR_{E_1^*,E_2^*} = \frac{Pr(E_1^*,E_2^* \mid H)}{Pr(E_1^*,E_2^* \mid H^c)} = \frac{Pr(E \mid H) + [\frac{h_1 h_2}{f_1 f_2} - 1]^{-1}}{Pr(E \mid H^c) + [\frac{h_1 h_2}{f_1 f_2} - 1]^{-1}} . \tag{3.8}$$

As may be seen, the overall inferential force of the two reports E_1^* and E_2^* not only depends on the the capacity of event E to discriminate between H and H^c, expressed by the likelihoods $Pr(E \mid H)$ and $Pr(E \mid H^c)$, but also on the conditional probabilities of the reports given E, that is the local likelihood ratios $LR'_{E_1^*} = h_1/f_1$ associated with report 1, and $LR'_{E_2^*} = h_2/f_2$ associated with report 2.

Notice further that Equation (3.8) can also be extended to multiple, say n, independent sources. For such a situation, the likelihood ratio can be shown to lead to the following:

$$LR_{E_1^*,\dots,E_n^*} = \frac{Pr(E_1^*,\dots,E_n^* \mid H)}{Pr(E_1^*,\dots,E_n^* \mid H^c)} = \frac{Pr(E \mid H) + [\prod_{j=1}^n \frac{h_i}{f_i} - 1]^{-1}}{Pr(E \mid H^c) + [\prod_{j=1}^n \frac{h_i}{f_i} - 1]^{-1}} . \tag{3.9}$$

Such a setting is typically encountered in so-called 'testing cases', where distinct examiners work on a well defined question. As pointed out by Example 4 here below, that may be an actual case or an experiment under predefined testing conditions (such as a proficiency test).

Example 4. – Imagine a situation in which it is of interest to infer something about a proposition of the kind 'the suspect's photocopier (some other printing device) was involved in the production (i.e., printing) of the questioned document'. Next, suppose a series of experts who have all examined both questioned and known samples. In addition, each expert provides a report on whether or not the suspect's photocopier was involved. For the purpose of illustration, consider some recent proficiency testing data reported by Collaborative Testing Services Inc. in 2010. As part of their questioned documents test No. 10-521, 149 participants correctly reported a given known source as 'was involved' in the production of a given questioned document. However, there were also 12 participants that incorrectly reported that the known source at hand 'was not involved'.[9] If one now takes the proposition E, that is 'the suspect's printing device was involved', as evidence in support of a higher level proposition H, which states the involvement of the suspect in some criminal activity, because one maintains $Pr(E \mid H) > Pr(E \mid H^c)$, then all statements E^* of 'was involved' can be considered as corroborating. They are, however, in *contradiction* with statements of the kind E^{c*}, that is 'was not involved', following definitions and discussion presented earlier in Section 3.2.2.

A corroboration, with respect to the proposition H, needs to meet $h_i > f_i$ for every source i in order to take place. This implies that the examination of the credibility of the sources must not be neglected even when confronted with a case of corroboration. Notice further that the likelihood ratio in Equations (3.8) and (3.9) cannot exceed LR_E or LR_{E^c}. That is, the joint value in an inference about H, based on a given number of individual sources, that report on E, cannot be higher than that for confirmed knowledge about E (i.e., a situation in which the actual state of E would be known). Stated otherwise, the value of individual reports for discriminating about H depends on the capacity of individual reports in discriminating between the states of the variable E. For example, if a report E^* is capable of 'establishing' E, then the likelihood ratio for E^*, that is LR_{E^*}, would equate that for E, that is LR_E. However, as long as E^* – or, by extension, a collection of reports $E_1^*,...,E_n^*$ – cannot 'establish' E with certainty, which should be the regular case, $LR_{E*} < LR_E$.

A convergence is given when two or more sources state the occurrence of distinct events that do not support the same intermediate hypothesis. As depicted by Figure 3.1 (b), sources S_1 and S_2 may report the occurrence of the events E and F which are conditionally independent given the proposition H. This is equivalent to having two independent strains of inference of the kind $E^* \rightarrow E \rightarrow H$, as illustrated in Figure 3.1 (a). In such a case, the overall likelihood ratio for the two reports E^* et F^* is given by the product of the likelihood ratios associated with the individual reports. That is, $LR_{E^*,F^*} = LR_{E^*} \times LR_{F^*}$, and Equations (3.3) and (3.4) can again be applied. An illustration for convergence can be obtained by reconsideration of Examples 1 and 2.

Example 5. – Suppose a scientist's report E^*, that event E occurred, that is, a given suspect wrote a signature on a questioned document. In addition, assume further that the questioned document bears ridge skin marks. Let F now denote – unlike in Example 2 – the proposition according to which the fingermarks *come* from the suspect. With regard to this, let F^* denote a scientist's report of such a conclusion. Assuming again that the fingermarks are found in a position (on the document) where marks from the author of the crime of interest would be expected to be found, the proposition F can be considered relevant in an inference about the proposition H, that is 'the suspect is the the author of the fraud'. Consequently, proposition F would now favour H because the probability of F may be taken to be greater given H than given H^c. That is, stated otherwise, the likelihood ratio for F, written $LR_F = Pr(F \mid H)/Pr(F \mid H^c)$, is greater than one. Along with a likelihood ratio for the E, written $LR_E = Pr(E \mid H)/Pr(E \mid H^c) > 1$, this presents a further element in support of H, and thus implies *convergence*.

If, however, the events E and F are conditionally dependent upon the ultimate proposition H, then Equations (3.3) and (3.7) need to be employed. In particular, one needs to account for the fact that when evaluating the probative value of F, it is necessary to account for what has been observed in relation with the first source, and this is expressed by the conditional likelihood ratio $LR_{F|E}$. According to the specified probabilistic underpinning, this may lead to the observation that the second observation F has more evidential value when E is already known, compared to a situation in which nothing is known about the first source. In such a case, the evidence is called 'synergic'. However, it may also be the case that knowledge about E diminishes the inferential force of F and this would be a situation of redundancy. This may go as far as to entail a directional change, rather than only reducing the inferential force of F. That is, an individual consideration of a supportive event F, that is $LR_F > 1$ (i.e., supporting H), may turn into a support for the alternative proposition, H^c, that is $LR_{F|E} < 1$.

[9] Summary Report available at `http://www.ctsforensics.com/assets/news/3021_Web.pdf`, last accessed March 8th 2011.

3.2.4 A closer look at the drag coefficient

Consider again a situation as in Example 1, discussed in Section 3.2.2, where the report E^* of a single expert (source S_1) is used to infer something about the occurrence of an event E. As shown in Figure 3.1 in terms of a path starting at E^*, the event E is in turn of interest in an inference about $\{H, H^c\}$. The inferential force of the scientist's report is as defined earlier in Equation (3.5):

$$LR_{E^*} = \frac{Pr(E \mid H) + [\frac{h_1}{f_1} - 1]^{-1}}{Pr(E \mid H^c) + [\frac{h_1}{f_1} - 1]^{-1}}.$$

Here, the term called 'drag coefficient' is given by $[(h_1/f_1) - 1]^{-1}$. It is part of both the numerator and denominator and written as D, for short. As mentioned earlier in Section 3.2.2, D acts like an inferential drag on $Pr(E \mid H)$ and $Pr(E \mid H^c)$. The drag coefficient is also encountered in other likelihood ratio formulae considered so far in this section, differing only with respect to the probabilities that are incorporated in this expression. The underlying mechanism that generates an inferential drag is, however, the same. There is also no difference with respect to how the bound of the likelihood ratios based on reports comes about.

It is useful to take a closer look at some limiting cases in order to illustrate how D generates a so-called inferential drag. Suppose that a given source, for instance S_1, states E^*, but has no credibility. That is, stated otherwise, the evidence given by S_1 does not enable one to discriminate between E or E^c. This is the case whenever S_1 is equally likely to provide report E^* given E and E^c. Alternatively, one may also say that the 'hit probability', that is, the probability of report E^* when E is in fact true, $Pr(E^* \mid E)$, equals the 'false positive probability', that is the probability of report E^* when E^c is actually true. Let us also recall that, previously, the latter two probabilities have been written, for short, h_1 and f_1. So, for a situation in which $h_1 = f_1$ is assumed to hold, the drag coefficient is $[1-1]^{-1} = 1/0$, a term which tends towards infinity. Consequently, the likelihood ratio for report E^* becomes:

$$LR_{E^*} = \frac{Pr(E \mid H) + \infty}{Pr(E \mid H^c) + \infty} \approx 1.$$

Hence, the influence of $Pr(E \mid H)$ and $Pr(E \mid H^c)$, which both assume values from the range between zero and unity, becomes negligible. The drag coefficient dominates the numerator and the denominator so that the likelihood ratio tends towards a value of one. As may thus be seen, the failure of E^* to discriminate between E and E^* deprives LR_{E^*} to draw advantage from the capacity of E to discriminate betwen H and H^c.

In order to pursue this analysis, now suppose a situation where S_1 has maximal credibility. That is, its hit probability is unity and that of a false positive is zero. This is just another way to say that the source S_1 provides perfect evidence for disrciminating between E and E^c: it always reports E^* when in fact E is true (i.e., $Pr(E^* \mid E) = h_1 = 1$) and never reports E^* otherwiese (i.e., $Pr(E^* \mid E^c) = f_1 = 0$). It can now be seen that in such a case the drag coefficient is $[(1/0) - 1]^{-1}$. While $1/0$ tends towards infinity, the drag coefficient becomes $[\infty - 1]^{-1}$, which is virtually zero. The likelihood ratio thus becomes:

$$LR_{E^*} = \frac{Pr(E \mid H) + 0}{Pr(E \mid H^c) + 0} = \frac{Pr(E \mid H)}{Pr(E \mid H^c)} = LR_E.$$

This result shows that the likelihood ratio for the report E^* of a perfectly credible source equates that for knowing the occurrence of E for sure. The likelihood ratio LR_{E^*} thus has an upper bound given by LR_E.

Finally, imagine yet another situation where S_1 has a hit probability of zero, but a false positive probability of unity. This is a situation in which a given source would systematically report the opposite of what it should state in order to be right. More formally, such a source would report E^* whenever E^c is true ($Pr(E^* \mid E^c) = 1$) and report E^{c*} when E is true ($Pr(E^{c*} \mid E)$. In such a case, the drag coefficient is $[(0/1)-1]^{-1} = -1$. Consequently, the likelihood ratio becomes

$$LR_{E^*} \quad = \quad \frac{Pr(E \mid H) - 1}{Pr(E \mid H^c) - 1} = \frac{Pr(E^c \mid H^c)}{Pr(E^c \mid H)} = LR_{E^c}.$$

This represents the likelihood ratio for knowing the occurrence of E^c for sure and shows why the lower bound of LR_{E^*} is given by LR_{E^c}.

3.2.5 From Schum graphs to Bayesian networks

Throughout the previous sections it has become apparent that the extension of probabilistic value of evidence analyses to more than one item of evidence requires consideration to be given to additional and subtle aspects. These are not encountered when items of evidence are looked at in isolation. These aspects relate to notions that characterise the joint occurrence of several items of evidence, such as evidential harmony and dissonance. Moreover, the joint probative force of several items is clearly to be distinguished from the evidential value associated with an item of evidence considered in isolation. For evidence analyses Schum graphs are important in this context because they provide a concise representation of the variables involved as well as the structure of the argument that is invoked for progressing from observations to unobserved propositional variables of interest.

Above all, Schum graphs remain, however, an essentially representational technique wherein probability and the dynamics of its calculations are not explicitly incorporated. As such, these models offer only limited assistance to their users in defining the relevant probabilistic computations. In particular, they offer no means to actually execute these computations. This is a major difference with respect to Bayesian networks. These models have an underlying probabilistic architecture. That is, each node contains a probability table that specifies the nature (which is, as the name says, probabilistic) as well as the strength of the relationship to connected neighboring nodes. In addition, the probabilistic underpinning of these models is defined in such a way that ‘entering’ evidence at some node (or group of nodes) – that is, communicating to the model which variables have been observed – will update the probability distributions associated with all remaining nodes according to Bayes’ theorem. This the reference rule for reasoning in the light of uncertainty, most notably also in forensic science [118], and this is the reason why these models are of particular interest for studying questions about forensic inference. Hereafter, this topic is pursued in further detail in Sections 3.3 and 3.4.

3.3 Bayesian networks

A Bayesian network is a graphical model that is constructed on the basis of two main ingredients: nodes (or vertices) and arcs (directed edges). A node can represent a propositional variable of interest that possesses mutually exclusive states to which a probability distribution[10] is associated. This probability distribution has the form of a so-called node probability table. A requirement that stems from probability theory is that the sum of the probabilities associated with each state of a variable must sum up to unity. That is, a variable is in exactly one of its possible states, although it may not be known which.

In turn, directed links are represented by arrows and these connect pairs of nodes. Such connections between nodes reflect probabilistic relevance relationships. Variables that are not directly dependent are connected through a chain of nodes. In summary, the topology of a Bayesian network thus represents the dependance relationships between the variables that are retained in a probabilistic analysis [108]. The notion of ‘directed path’ refers to a

[10]The discussion in this paper will concentrate on discrete models, since genuine continuous nodes in Bayesian networks can only be used – at the current state of their development – with several constraints [77]. Among the principal constraints is that one can only handle conditional Gaussian (Normal) distributions. Another constraint, a structural one, forbids the specification of a discrete node as a child of a continuous parent.

sequence of connections between nodes. And for a Bayesian network structure to be valid, nodes and arcs must be combined in a way that does not lead to cycles. For this reason, Bayesian networks are also called 'directed acyclic graphs'.

On the basis of these definitional elements, one can extend the consideration from a static description of relevance relationships between nodes, as it is given by a graph's structure, to the analysis of the flow of information within a given network structure. That is, depending whether or not to a set of nodes evidence[11] is given, a path may be 'activated' or 'blocked'. This determines whether or not evidence is propagated through a path.

In order to exemplify this on a more formal account, consider a path from a node X to a node Y that leads through a node Z to which evidence is given. In such a case, X is said to be directionally separated (or 'd-separated') form Y given Z when X and Y are serially connected via Z (i.e., by $X \rightarrow Z \rightarrow Y$) or when they are divergently connected (i.e., by $X \leftarrow Z \rightarrow Y$). This means that X and Y are conditionally independent, given Z. The path will thus be said to be 'blocked'. If, however, X and Y are linked via Z through a converging connection (i.e., $X \rightarrow Z \leftarrow Y$), then X and Y are not d-separated, given Z, but 'd-connected'. Unlike in the two situations mentioned above, here X and Y are marginally dependent given information about the intermediate node Z. Accordingly, the path is said to be 'activated' [108, 77].

The technical description of a relevance relationship between nodes is often based on kinship terminology. For example, when a node A has an arrow pointing towards another node B, then A may be called a 'graphical parent' and B is called a 'child node'. A node without parents is a called 'root node' and its associated node probability table contains probabilities that are not conditioned (except on circumstantial information that is otherwise not explicitly represented by a node). All other nodes, that is nodes with entering arcs, have probability tables that contain conditional probabilities.

A main asset of Bayesian networks consist in their ability to compute a joint probability distribution by taking account of assumed dependencies between variables. Generally, a distribution Pr of n discrete variables $X_1, X_2, ..., X_n$ can be decomposed by the chain rule of probability calculus:

$$Pr(A_1, ..., A_n) = \left[\prod_{i=2}^{n} Pr(A_i \mid A_1, ..., A_{i-1}) \right] Pr(A_1) \; . \qquad (3.10)$$

But if one can assume that X_j is influenced exclusively by certain predecessors and is insensitive to other variables, one can reformulate Equation (3.10) to:

$$Pr(A_1, .., A_n) = \prod_{i=1}^{n} Pr(A_i \mid par(A_i)) \; . \qquad (3.11)$$

Here, $par(A_j)$ stands for the group of parents of x_j. This allows for a considerable reduction of the complexity of computations as well as the quantity of probabilities that have to be stored. The result is a joint probability distribution that is broken up into several local distributions. Such a local distribution contains a variable with its parents and all the distributions conditioned by every combination of the values of the parents.

3.4 Case studies in evidence combination using Bayesian networks

3.4.1 Case example 1: Footwear mark evidence

The casework example analysed in this section focuses on the joint evaluation of size and general pattern observable on footwear marks. Consider the following outset:

[11] With respect to Bayesian networks, the term 'evidence' refers to a statement about the certainties of a node's states. A variable whose actual state is known is also called 'instantiated'.

Case example 1:[12] A woman was found dead in her bed. She died because of severe injuries and wounds inflicted by a third person. At the crime scene, footwear marks from a left and right sole of a Nike Multi Court III of size 13US have been detected. These marks were found on the upper surface of a toilet that was located under a bathroom window, which was suggested as the point of entry of a burglar. Marks were also found on the window and the tiled floor of the bathroom. During subsequent investigations, the husband became the focus of attention as it was thought that he mimicked a burglary by gaining entry through the bathroom window. The husband himself possessed a pair of Nike Multi Court III of size 13US. They were seized at his office. Examination of the crime scene marks indicated that all marks were made by a Nike Multi Court III, because the observed mould-design was specific to Nike, that model and that size. This could be taken as a reliable information. Subsequent comparative examinations with the pair of shoes seized from the suspect did not allow to 'exclude' this pair as being the source of the marks found on the crime scene.

This case covers several interesting aspects, relating in part to issues in the combination of evidence, that an analysis through Bayesian networks can help to set further into context. As a first aspect, it is often useful to start by focusing on the definition of target propositions, which is an important requirement for a probabilistic approach to value of evidence analyses. For the purpose of the case considered here, suppose that it is of interest to draw an inference about propositions at the source-level, that is whether or not the suspect's pair of shoes (some other pair of shoes) is at the source of the marks found at the crime scene. At this juncture it seems important to emphasise that the contextual information has an important bearing on the formulation of the alternative proposition. In fact, if it is not the suspect's pair of shoes that left the crime marks (i.e., proposition F_p), then, following the proposition put forward by the defence, it is not just some other pair of shoes that left the crime marks, but a pair of shoes worn by a burglar (referred to hereafter as F_d). This stems from the husband's suggestion that this home was burglarized and that his wife was killed by a burglar. This definitional detail is important because it determines the relevant population on which one should focus. In turn, this will have a bearing on the kind of data that will be used to inform the numerical specification of the inference model proposed here below. More generally, this allows one to insist on the importance of defining propositions *not* on the basis of observations made on the crime marks, but on the basis of actual case circumstances.

In a further step, it is necessary to capture observations upon which an inference about the proposition F is to be based. As mentioned in the case description, there are multiple marks found on the crime scene. However, in order to keep the analysis and discussion at a tractable level, it is decided here to regroup distinct source-level propositions for individual marks into one. This is considered as an acceptable simplification here because, on the basis of information available from the scene investigation and subsequent mark examination, relevance of the marks and a single source could be reasonably allowed as assumptions. In the currently discussed case, observations can be broadly divided into two parts, that is (i) a Nike Multi Court III general sole pattern, and (ii) a size of 13US. Let information about general pattern be denoted by E_1 with e_1 referring to 'Nike Multi Court III'. Accordingly, $\bar{e}_1$ will refer to any sole pattern other than 'Nike Multi Court III'. In turn, let the variable for size be denoted by E_2 with e_2 referring to 13US, and $\bar{e}_2$ all sizes other than 13US.

Following discussion about relevance relationships among variables of interest, presented earlier in Section 3.2, a dependency structure for the variables considered here could be as shown in Figure 3.2. The likelihood ratio for both items of evidence, E_1 and E_2, can thus be formulated as follows (omitting circumstantial information I from notation):

$$LR_{E_1,E_2} = \frac{Pr(E_1 = e_1, E_2 = e_2 \mid F_p)}{Pr(E_1 = e_1, E_2 = e_2 \mid F_d)} = \frac{Pr(E_1 = e_1, E_2 = e_2 \mid F_p)}{Pr(E_1 = e_1 \mid F_d)Pr(E_2 = e_2 \mid E_1 = e_1, F_d)}. \quad (3.12)$$

In this way of providing the likelihood ratio, the numerator is not written in extended form. In fact, if the suspect's pair of shoes with known characteristics is the source of the crime marks, then certainly we would expect to find marks with the same characteristics. Therefore, the value of 1 is assigned to the numerator here. This is an

[12]The scenario is adapted from a real case analysed at the author's institution by Prof. Champod.

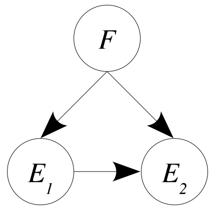


Figure. 3.2 – Bayesian network for inference about a binary source-level proposition F (defined as 'the suspect's pair of shoes is the source of the crime marks') on the basis of observations about general sole pattern (E_1) and size (E_2).

expression of assumptions that cover stability over time and in substance of shoe characteristics, as well as their reliably discernible reproduction in terms of marks.

The denominator, however, is written in more detail by invoking the third law of probability. In particular, general pattern is chosen as a conditional for size, as implied by the graph structure adopted in Figure 3.2. This conditioning could also have been chosen differently, but generally better data are available for sizes among general patterns rather than for general patterns among sizes. Proceeding in this way, attention is thus first drawn to $Pr(E_1 = e_1 \mid F_d)$, that is the probability of encountering the general sole pattern of a Nike Multi Court III, if another pair of shoes, worn by a burglar, is at the source of the marks found at the crime scene. According to available data in this case (i.e., a regional database on footwear observed on individuals that came to police attention), four individuals among 21621 were seen to wear Nike Multi Court III shoes (i.e., one pair of each of the sizes 10US, 11US, 12US and 13US). For the purpose of the current discussion, we thus accept the coarse probability assignment $Pr(E_1 = e_1 \mid F_d) = 4/21621$. A summary of the probability assignments for the table of the node E_1 is given in Table 3.1.

Table. 3.1 – Conditional probabilities assigned to the table of the node E_1, where e_1 denotes the observation of a Nike Multicourt III general pattern and F denotes propositions at the source-level.

	$F:$	F_p	F_d
E_1:	e_1	1	0.000185
	$\bar{e}_1$	0	0.999815

Next, attention is directed to a second term, $Pr(E_2 = e_2|E_1 = e_1, F_d)$, that describes the conditional probability of finding a pair of shoes of size 13US among Nike Multi Court III shoes, if another pair of shoes, worn by a burglar, is at the source of the marks found at the crime scene. Referring again to the data mentioned above, one can see that there is one such pair among the four instances of this model of shoes. But adopting a value of 1/4 seems somewhat delicate here because the sample size is very limited (4 individuals). Notice that if, for example, no instance would have been observed, a reasonable probability assignment would not have been possible without considering a procedure that is capable of dealing with zero observations [e.g., 27, 135]. For this reason, data on sales of Nike Multi Court III shoes in neighboring countries have been collected. These suggested a value of 3.3% which was retained here for $Pr(E_2 = e_2|E_1 = e_1, F_d)$.[13] This value enters the probability table of the node E_2 (Table 3.2) and is

[13] This assignment is based on the assumption that the preferences of criminals (i.e., burglars) in choosing shoes does not deviate from that of the 'average' shoe customer, as reflected by the general data on sales.

part of the quantitative specification of the Bayesian network shown in Figure 3.2. Table 3.2 shows further that the observation of size 13US is taken to be certain under the assumption that the suspect's pair of shoes is the source of the crime marks. The value 0.005 for a size 13US observation given the shoe of a burglar with another sole pattern, $Pr(E_2 = e_2 \mid E_1 = e_1, F_d)$, is also derived from an appropriate database. At this point, the value is primarily added in order to comply with the definitional requirement of Bayesian networks of having fully specified node tables. As seen from Equation (3.12), the value is however irrelevant for the kind of likelihood ratio calculations pursued here, essentially because the evaluation in the scenario here is based on the observation $E_1 = e_1$.[14] This will be different for further analyses discussed towards the end of this section.

Table. 3.2 – Conditional probabilities assigned to the node E_2, that is shoe size, as a function of general pattern (E_1) and assumptions about the source of the marks (i.e., 'F_p: the suspect's pair of shoes is the source of the crime marks' and 'F_d: some other pair of shoes, from a burglar, is the source of the crime marks').

	$F:$	F_p		F_d	
	$E_1:$	e_1	$\bar{e}_1$	e_1	$\bar{e}_1$
$E_2:$	e_2	1	1	0.033	0.005
	$\bar{e}_2$	0	0	0.967	0.995

Introducing these assignments in the likelihood ratio (3.12) leads to the following result:

$$LR_{E_1,E_2} = \frac{1}{Pr(E_1 = e_1 \mid F_d)Pr(E_2 = e_2 \mid E_1 = e_1, F_d)} = \frac{1}{4/21621 \times 0.033} \approx 160'000. \tag{3.13}$$

This means that the joint consideration of the two descriptors, general pattern and size, of the footwear mark evidence supports the proposition F_p by a factor of 160'000, rather than the stated alternative F_d.

The specified dependency between E_1 and E_2, as well as the conditioning of these variables upon the proposition F allows for effects of *redundance*, *synergy* or *directional change*. In order to examine if, and to what degree, one of these effect applies in the case studied here, one can compute an expression R, defined as follows [125]:

$$R_{E_2=e_2|E_1=e_1} = 1 - \frac{logLR_{E_2=e_2|E_1=e_1}}{logLR_{E_2=e_2}}, \quad (LR_{E_2=e_2} \neq 0) \ . \tag{3.14}$$

This formula compares the likelihood ratio for E_2 given E_1, that is $LR_{E_2=e_2|E_1=e_1}$, against the likelihood ratio for E_2 for a situation in which nothing is known about E_1, that is $LR_{E_2=e_2}$. It can be seen that if the numerator of the fraction is larger than the denominator, this means that the evidential value associated with E_2 is stronger by knowing the state of the variable E_1, than in a situation in which nothing would be known about the latter variable. If this condition holds, then E_2 and E_1 are said to be *synergic* in nature and R becomes smaller than zero.

When the likelihood ratios in the numerator and the denominator of the fraction take the same numerical value, which would mean that knowledge about E_1 would not influence the probative value associated with E_2, then E_1 and E_2 are conditionally independent on F. Consequently, R would become zero.

There is yet another situation for which the expression R allows for further insight, that is when it becomes unity. In order for R to take this value, the fraction must equate zero. This is the case when the numerator becomes zero, and this requires the term $LR_{E_2=e_2|E_1=e_1}$ to be one. If the latter likelihood ratio is one, this means that E_2 is entirely *redundant* in an inference about F given E_1.

[14] As a consequence of this, any probability assignments in the columns of $E_1 = \bar{e}_1$ do not enter the considerations.

If, however, the likelihood ratio in the numerator, $LR_{E_2=e_2|E_1=e_1}$ becomes smaller than one, and the likelihood ratio associated with E_2 when nothing is known about the state of E_1 is greater than one, then one is confronted with an effect of *directional change*, and R would become larger than one.

The expression R thus is a measure of redundancy only if R takes a value between zero and one ($0 < R \leq 1$). Notice, however, that the interpretation of R as put forth here is only valid for cases where both items of evidence, E_1 and E_2, favour F_p over F_d.[15]

Applied to the case considered here, we start by finding the conditional likelihood ratio $LR_{E_2=e_2|E_1=e_1}$, given by:

$$LR_{E_2=e_2|E_1=e_1} = \frac{Pr(E_2 = e_2 \mid E_1 = e_1, F_p)}{Pr(E_2 = e_2 \mid E_1 = e_1, F_d)} = \frac{1}{0.033} \approx 30 \ .$$

Next, we continue by finding the likelihood ratio for the denominator of the fraction in Equation (3.14), that is $LR_{E_2=e_2}$. If nothing is known about the state of the first variable, E_1, then uncertainty about the actual state of this variable needs to be accounted for. The likelihood ratio $LR_{E_2=e_2}$ can thus be developed as follows:

$$\begin{aligned}
LR_{E_2=e_2} &= \frac{Pr(E_2 = e_2 \mid F_p)}{Pr(E_2 = e_2 \mid F_d)} \\
&= \frac{Pr(E_2 = e_2 \mid E_1 = e_1, F_p) Pr(E_1 = e_1 \mid F_p) + Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_p) \overbrace{Pr(E_1 = \bar{e}_1 \mid F_p)}^{1-Pr(E_1=e_1|F_p)}}{Pr(E_2 = e_2 \mid E_1 = e_1, F_d) Pr(E_1 = e_1 \mid F_d) + Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d) \underbrace{Pr(E_1 = \bar{e}_1 \mid F_d)}_{1-Pr(E_1=e_1|F_d)}} \\
&= \frac{Pr(E_1 = e_1 \mid F_p)[Pr(E_2 = e_2 \mid E_1 = e_1, F_p) - Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_p)] + Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_p)}{Pr(E_1 = e_1 \mid F_d)[Pr(E_2 = e_2 \mid E_1 = e_1, F_d) - Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)] + Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)}
\end{aligned}$$

Using the data defined so far in Tables 3.1 and 3.2 thus allows one to find the following:

$$LR_{E_2=e_2} = \frac{1 \times (1-1) + 1}{4/21621 \times (0.033 - 0.005) + 0.005} \approx 200 \ .$$

One can thus see that by knowing that the general pattern is that of a Nike Multi Court III ($E_1 = e_1$), the likelihood ratio for the information about shoe size diminishes from approximately 200 to approximately 30. This means that the evidence of the general pattern incorporates already some information about the size of the shoe and this renders the latter aspect inferentially redundant to some degree. On the basis of these calculations, one can now proceed with calculating $R_{E_2|E_1}$, for which one obtains:

$$R_{E_2=e_2|E_1=e_1} = 1 - \frac{logLR_{E_2=e_2|E_1=e_1}}{logLR_{E_2=e_2}} = 1 - \frac{1.4771}{2.3010} \approx 0.3581 \ .$$

The expression R thus takes a value between 0 and 1 and this indicates an effect of redundancy. Following Schum [125], one could thus say that information on shoe size 13US, that is $E_2 = e_2$, is redundant at 0.3581 with respect to information about Nike Multi Court III general pattern ($E_1 = e_1$).

More generally, it is interesting to note that knowledge about E_1 affects the assessment of the probative value of $E_2 = e_2$ only as long as $Pr(E_2 = e_2 \mid E_1 = e_1, F_d) \neq Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)$. Stated otherwise, information about the general pattern e_1 is only relevant for assessing the probative value of the observed size $E_2 = e_2$ as long as that size occurs at a different rate on pattern $E_1 = e_1$ than on other patterns. It is thus not only relevant to have accurate information about the occurrence of the size 13US among Nike Multicourt III, but also among shoes

[15] An interpretation of R for the opposite case can be found, for example, in Schum [125].

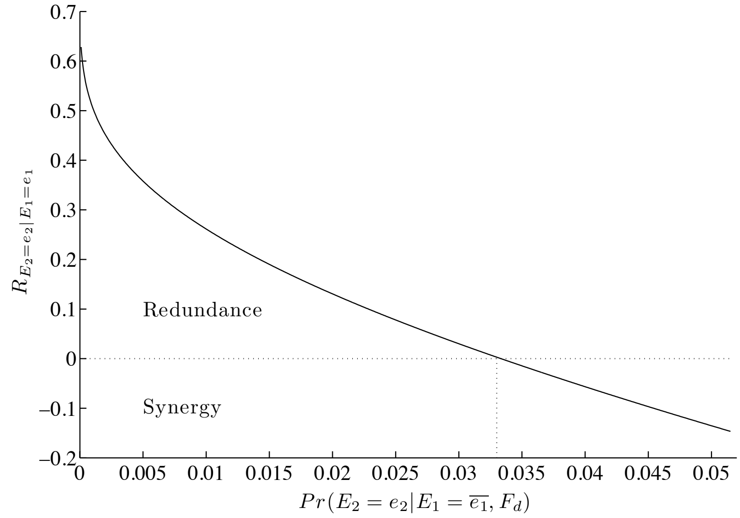


Figure. 3.3 – Representation of $R_{E_2=e_2|E_1=e_1}$ as a function of uncertainty about $Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)$. Information about the first item of evidence E_1 is irrelevant for the assessment of the second item of evidence $E_2 = e_2$ when $Pr(E_2 = e_2 \mid E_1 = e_1, F_d) = Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)$, that is equal to probability 0.033.

with other general patterns (different from Nike Multicourt III). To some degree, this may seem counterintuitive because, generally, under the alternative proposition (i.e., F_d) one is inclined to inquire about the occurrence of a target characteristic only among other *compatible* (in terms of general pattern) potential sources. A graphical illustration of this shown in Figure 3.3, which represents the value of $R_{E_2=e_2|E_1=e_1}$ as a function of uncertainty about $Pr(E_2 = e_2 \mid E_1 = \bar{e}_1, F_d)$. As may be seen, when $Pr(E_2 \mid E_1 = e_1, F_d) = Pr(E_2 \mid E_1 = \bar{e}_1, F_d)$, that is equal to probability 0.033, then information about E_1 is irrelevant for the assessment of the probative value of $E_2 = e_2$. This would also become clear from Table 3.2, which would contain the same values in the last two columns. This would correspond to a situation of independence and the arc from node E_1 to the node E_2 would entail no inferential effect.

This example emphasises the importance of examining potential dependency relationships between distinct items of evidence. Suppose that one would have assumed, for simplicity, that the shoe size and the general pattern are conditionally independent on the specified set of target propositions $\{F_p, F_d\}$. In such a case, one would have obtained a likelihood ratio of 1’081’050[16] instead of 160’000. This represents a difference by a factor of approximately 6.8. Besides, this example also provides an illustration of a case in which the likelihood ratios for each item of evidence favor F_p over F_d and this implies, consequently, a situation of convergence.

3.4.2 Case example 2: Fingermark and footwear mark evidence

Case description

The example pursued in this section is based on a case reported in Champod [21]. It involves two main and distinct items of evidence, that is fingermark and footwear mark evidence.

[16] $LR_{E_1,E_2} = \frac{Pr(E_1=e_1,E_2=e_2|H_p)}{Pr(E_1=e_1,E_2=e_2|H_d)} = \frac{Pr(E_1=e_1|H_p)}{Pr(E_1=e_1|H_d)} \times \frac{Pr(E_2=e_2|H_p)}{Pr(E_2=e_2|H_d)} = \frac{1}{0.000185} \times \frac{1}{0.005} = 1'081'050$

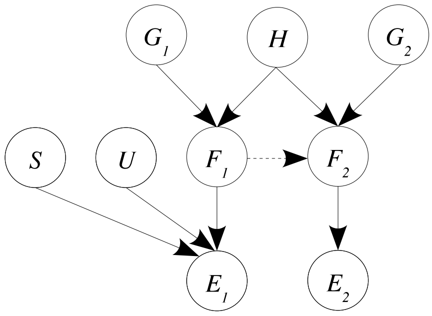


Figure. 3.4 – Bayesian network for the joint evaluation of finger (E_1) and footwear mark (E_2) evidence under crime-level propositions (H). Uncertainty about the marks is accounted for by the nodes G. Intermediate source-level propositions that specify the suspect as the source of the marks are incorporated in terms of the nodes F. The dotted arc between the latter two nodes indicates a *possible* relevance relationship (depending on the probabilities assigned to the node F_2), conditional upon H. The nodes S and U model the ridge skin configurations of, respectively, the suspect and an unknown person.

Case example 2: After a burglary in a shop, crime scene investigators detected two footwear impressions located between flower pots at the back of the shop. In addition three fingermarks were detected on a sliding door. The fingermarks consisted of three arches. These were thought to represent the anatomical sequence index-middle-ring finger of a right hand (either a triple arch, denoted A-A-A, or of an arch, a tented arch and another arch, denoted A-T-A). After cross-checking with staff members from the shop, the police investigators retained that, on the basis of additional circumstantial information collected on the scene (e.g., on modus operandi), the evidential marks were in direct relation with the burglary (i.e., relevant to the incident under investigation). The same day, another burglary was committed in the same region. Following a description provided by the victim, a suspect was arrested. Subsequently it was found that the shoesoles and the fingerprints of the suspect 'corresponded' to the marks recovered on the scene of the first burglary. On their own, the fingermarks found at the scene did not offer sufficient quality to allow for 'individualising' the suspect as the source of the crime marks. Likewise, the footwear marks could not be unequivocally associated with the suspect's shoes. The combination of the available evidence (sequence of three arches; two corresponding footwear impressions) nevertheless provided an interesting link between the suspect and the crime scene.

From an investigator's or evaluator's point of view, a main question of interest in this scenario is that of the suspect's involvement in first burglary. This is a principal difference to the case discussed in the previous section where the proposition of interest was formulated on the so-called 'source-level'. In the case considered in this section, propositions are defined on the 'crime-level', that is 'H_p: the suspect is the criminal' and 'H_d: some person other than the suspect is the criminal'. For the forensic scientist, this implies questions of the following kind: 'What is the value of the fingermark evidence for discriminating between H_p and H_d?', 'What is the probative value of the footwear mark evidence?', and 'What is the joint probative value of these two items of evidence?'.

Structure for a Bayesian network

Reasoning about questions as mentioned at the end of the previous section can be represented and supported by a Bayesian network as shown in Figure 3.4. It provides an outline of two main strains of argument that are explained in some further detail here below. Consider first the Bayesian network component for the fingermark evidence, that is the path leading from the node for the target proposition H to the observational variable E_1. In this local network fragment, the node G_1 takes the task of modeling the relevance of the fingermark with respect to the burglary under investigation. The node G_1 is binary with the states $G_1 = g_1$ and $G_1 = \bar{g}_1$, representing the propositions according to which, respectively, the fingermarks come (g_1) and do not come ($\bar{g}_1$) from the person who committed the burglary (i.e., the offender). The probabilities associated with the two possible node states are $Pr(G_1 = g_1) = 0.99$ and $Pr(G_1 = \bar{g}_1) = 0.01$. They are thought to reflect a firm belief that the fingermarks are relevant to the case. In turn, the node F_1 defines a pair of source-level propositions where $F_1 = f_1$ specifies that the suspect is the source of the crime marks and $F_1 = \bar{f}_1$ specifies that some person other than the suspect is the source of the crime marks. The probabilities associated with this node are essentially assignments of zero and one, as shown in Table 3.3. For example, if the suspect is the author of the burglary (H_p) and the crime marks come from the burglar ($G_1 = g_1$), then the crime marks must come from the suspect: $Pr(F = f_1 \mid G_1 = g_1, H_p) = 1$. One value, however, is different from zero and one, that is the probability that the suspect left the marks for innocent reasons, $Pr(F_1 = f_1|G_1 = \bar{g}_1, H_d)$. Here, a value of 0.01 is assigned.

Table. 3.3 – Table with conditional probabilities for the node F_1 (representing propositions at the source-level), as a function of the nodes G_1 (a proposition modeling uncertainty about the relevance of the crime marks) and H (representing propositions at the crime-level).

	H:	H_p		H_d	
	G_1:	g_1	$\bar{g}_1$	g_1	$\bar{g}_1$
F_1:	f_1	1	0	0	0.01
	$\bar{f}_1$	0	1	1	0.99

The source-level node F_1 acts as a conditioning for the observational variable E_1, which accounts for the observations made on the crime marks. Let $E_1 = e_1$ denote the observation of an anatomical sequence A-A-A or A-T-A, and $E_1 = \bar{e}_1$ the observation of an other sequence (different from A-A-A and A-T-A). In addition, let U be a binary node that accounts for the proposition that an unknown person possesses the sequences in question (A-A-A or A-T-A), with $U = u$ denoting the truth and $U = \bar{u}$ the negation of this proposition. The probabilities assigned to the latter two states are $Pr(U = u) = 0.0071$ and $Pr(U = \bar{u}) = 0.9929$. These values have been derived on the basis of data obtained from the AFIS system of the Swiss Central Police Bureau [21]. A further node S models the observation of the general patterns observed on the suspect, with $S = s$ denoting the proposition that the suspect possesses the sequence A-A-A or A-T-A, and $S = \bar{s}$ denoting the negation of this proposition. The probability values assigned to this node do not require further consideration because subsequent analyses can assume the suspect's ridge skin configuration as known (i.e., the node S will be fixed to the state s so that any initial probability distribution for this node will become irrelevant). A summary of the probabilities assigned to the node E_1 is given in Table 3.4. As may be seen, the assignments of zero and one imply that the node E_1 will 'copy' the actual ridge skin configuration of the suspect whenever f_1 holds, and that of an unknown person in all cases where $\bar{f}_1$ holds. Besides, notice that the explicit representation in terms of distinct nodes of the ridge skin configuration of the suspect and an unknown person allows one to complete the node table of E_1 with zeros and ones. In a model without nodes S and U, the probability of the occurrence of the target ridge skin configuration would need to be specified directly in the table of the node E_1.

A second strain of argument that makes up the Bayesian network shown in Figure 3.4 pertains to the footwear mark evidence. Here, the node G_2 models uncertainty about the relevance of the footwear mark. This binary node has two states, $G_2 = g_2$ for a situation in which the footwear marks come from the offender and $G_2 = \bar{g}_2$ for a setting in which they do not come from the offender. In analogy to the network fragment for fingermark evidence, 'relevance probabilities' of $Pr(G_2 = g_2) = 0.99$ and $Pr(G_2 = \bar{g}_2) = 0.01$ are defined here. The node F_2

Table. 3.4 – Conditional probabilities assigned to the node E_1 (observations made on the crime marks) as a function of the source-level proposition F_1 as well as the ridge skin configuration of the suspect (S) and that of an unknown person (U).

	F_1:	f_1				$\bar{f}_1$			
	S:	s		$\bar{s}$		s		$\bar{s}$	
	U:	u	$\bar{u}$	u	$\bar{u}$	u	$\bar{u}$	u	$\bar{u}$
E_1:	e_1	1	1	0	0	1	0	1	0
	$\bar{e}_1$	0	0	1	1	0	1	0	1

defines a pair of source-level propositions. The state $F_2 = f_2$ defines the suspect as the source of the crime marks whereas $F_2 = \bar{f}_2$ defines some person other than the suspect as the source. Notice that the probability that the suspect was wearing the shoe corresponding to the mark, given that he was the offender and that he left the mark (i.e., $Pr(F_2 = f_2 \mid G_2 = g_2, H_p)$) is defined as a function of the number of pair of shoes that the suspect could have potentially worn. In the case considered here, it is assumed that the suspect is in possession of two pairs of shoes regularly worn, so that the assignment $Pr(F_2 = f_2 \mid G_2 = g_2, H_p) = 0.5$ is retained here (later on, this probability will be abbreviated by w). The probability that the suspect left the marks for innocent reasons, $Pr(F_2 = f_2 \mid G_2 = \bar{g}_2, H_d)$, is also accounted for in this node. This probability depends on whether or not the suspect is the source of the fingermarks (proposition F_1) and is expressed by, respectively, a_1 and a'_2. A summary of all assignments is given in Table 3.5.

Notice that the directed link from F_1 to F_2 entails inferential force only when $a_2 \neq a'_2$. An assignment of the kind $a_2 \neq a'_2$ may be necessary to express the belief that the probability for the suspect being the source of the footwear marks, if he is innocent (H_d) and the footwear marks do not come from the offender ($G_2 = \bar{g}_2$), is different according to the truth or falsity of the proposition according to which the suspect is the source of the fingermarks (F_1). This would be a case of asymmetric independence, which occurs when variables are independent for some but not all of their values [131].

Table. 3.5 – Conditional probabilities assigned to the source-level node F_2 as a function of the propositions at the crime-level (node H), the relevance of the crime marks (node G_2) and the source-level proposition for the fingermark evience (node F_1).

	H :	H_p				H_d			
	G_2 :	g_2		$\bar{g}_2$		g_2		$\bar{g}_2$	
	F_1 :	f_1	$\bar{f}_1$	f_1	$\bar{f}_1$	f_1	$\bar{f}_1$	f_1	$\bar{f}_1$
F_2:	f_2	0.5	0.5	0	0	0	0	a_2	a'_2
	$\bar{f}_2$	0.5	0.5	1	1	0	0	$1-a_2$	$1-a'_2$

The node E_2 represents the observations made on the crime marks. The pattern actually observed in this case (and found to correspond with the suspect's shoe) is represented by $E_2 = e_2$. All other patterns are represented by the state $E_2 = \bar{e}_2$. The principal probabilities associated with this node are $Pr(E_2 = e_2 \mid F_2 = f_2) = 1$ and $Pr(E_2 = e_2 | F_2 = \bar{f}_2) = 0.015$, derived from a relevant regional database.[17]

[17] In order to avoid that the latter assignment is directly specified in the node table of E_2, it would be possible to specify the occurrence of the

The separate lines of argument from E_1 and E_2, via source-level propositions F and relevance considerations G, to ultimate propositions of interest, H, are instances of a general Bayesian network fragments for evaluating scientific evidence under crime-level propositions, initially described in Garbolino and Taroni [56]. Here, a logical combination of these two local lines of reasoning can be operated because of a common target node H, that is the proposition according to which the suspect is the offender [131]. This leads to an overall network structure that can be recognised as the classical conflict/convergence model proposed by Schum [125], discussed earlier in Section 3.2 (Figure 3.1(b)). In fact, there are the two event classes here, F_1 and F_2, that are not mutually exclusive, but both are directly dependent on H. Inference about those two event classes is made on the basis of E_1 and E_2, respectively. Notice further that the Bayesian network structure proposed in Figure 3.4 assumes that F_1 and F_2 are taken to be asymmetrically independent on H. As will be investigated in more detail further on, effects such as redundancy, synergy or directional change can appear in this network whenever $a_2 \neq a_2'$.

Likelihood ratio analyses

Start by considering the overall likelihood ratio $LR_{E_1=e_1,E_2=e_2}$ for the two items of evidence E_1 and E_2. It can be written as the product of the individual likelihood ratios $LR_{E_1=e_1}$ and $LR_{E_2=e_2|E_1=e_1}$:

$$LR_{E_1=e_1,E_2=e_2} = \frac{Pr(E_1=e_1,E_2=e_2|S=s,H_p)}{Pr(E_1=e_1,E_2=e_2 \mid S=s,H_d)} = \underbrace{\frac{Pr(E_1=e_1 \mid S=s,H_p)}{Pr(E_1=e_1 \mid S=s,H_d)}}_{LR_{E_1=e_1}} \times \underbrace{\frac{Pr(E_2=e_2 \mid E_1=e_1,S=s,H_p)}{Pr(E_2=e_2 \mid E_1=e_1,S=s,H_d)}}_{LR_{E_2=e_2|E_1=e_1}}. \tag{3.15}$$

A more detailed examination of this likelihood ratio is pursued hereafter in two steps, by separately considering the likelihood ratios associated with the individual items of evidence. When incorporating the model described so far in this section in a Bayesian network software, the numerator of $LR_{E_1=e_1}$ can be found by fixing the state of the node H to H_p and that of the node S to s. The conditional probability of interest is then obtained at the node E_1 and takes the value 0.990071. The value for the denominator of $LR_{E_1=e_1}$ is obtained at the same node, by leaving the node S in state s,[18] but changing that of H to H_d. The result of this propagation is 0.00719929. The high number of decimals in this result is solely retained here in order to allow for a subsequent comparison with an algebraic approach. In summary, thus, the following numerical result can be derived from a Bayesian network:

$$LR_{E_1=e_1} = \frac{0.990071}{0.00719929} = 137.5234 \approx 138.$$

The coherence of this result can be checked against published algebraic solutions for probabilistic inference about crime-level propositions that allow for uncertainty about the relevance of the crime mark. As established in [56], the case considered here relates to the following likelihood ratio originally developed in [41]:

$$LR = \frac{r\{1+(k-1)\gamma\}+k(1-r)\gamma'}{k[r\gamma+(1-r)\{a_1+(1-a_1)\gamma'\}]}. \tag{3.16}$$

In this expression, r accounts for the probability of relevance of the mark, incorporated in the currently discussed Bayesian network in terms of the value 0.99 for $Pr(G=g_1)$. The variable k represents the number of offenders which, in the case here, is one. The probability that the mark was left by the suspect for innocent reasons, designated

observed pattern on some other shoe in terms of a distinct node, in analogy to what has been done above for the ridge skin configuration of an unknown person.

[18] The reason for this is that, even though in the denominator one assumes an offender different from the suspect ($H = H_d$), it may still be the case that the suspect is the source of the crime stain ($F_1 = f_1$ is the case). This stems from the conditional probabilities assigned to the node F_1 (Table 3.3). It is for this reason that it is important to keep information about the suspect's ridge skin configuration as 'known' ($S = s$).

by a_1, corresponds to $Pr(F_1 = f_1 \mid G_1 = \bar{g}_1, H_d)$ of the Bayesian network (and was assigned the value 0.01). The terms γ and γ' represent the population proportions of the target analytical characteristic (here, the ridge pattern configuration) among criminal and non-criminal individuals. For the kind of ridge skin characteristics considered here, no differences in occurrence are assumed in the latter two categories of persons (criminal and innocent persons). Stated otherwise, it is assumed that there is no association between ridge skin characteristics and criminal behaviour. Analogous assumptions are regularly invoked for DNA evidence. Therefore, a single value $\gamma' = \gamma$ is retained here. It corresponds the the value assigned to the state u of the node U, that is the probability that an unknown person would possess the ridge skin characteristics of interest (taken to be 0.0071 here). Applying these numerical values in Equation (3.16) allows one to obtain 137.5234 which is in entire agreement with the numerical values derived from the proposed Bayesian network.

In a second step, one can proceed analogously for the footwear mark evidence, but there are two options to this. One consists in assuming $a_2 = a_2'$, which is a setting in which the footwear mark evidence and the fingermark evidence would be considered independent conditionally on H. A second option consists of allowing for an asymmetrical independence, which applies whenever $a_2 \neq a_2'$.

Consider the former case first and set the probability for leaving the marks for innocent reasons, a_2, to 0.01. Then, setting the node H to, alternatively, H_p and then to H_d, allows one to find the numerator and the denominator of the likelihood ratio $LR_{E_2=e_2}$:[19]

$$LR_{E_2=e_2} = \frac{0.502575}{0.0150985} = 33.28642 \approx 33.$$

Again, the coherence of this result can, following analyses presented in [131, 14], be examined by comparison with an algebraic approach previously published by Evett et al. [46]:

$$LR = \frac{[p_{mrk}w + \gamma(1-w)]r + \gamma(1-r)}{\gamma + [p_{mrk}a_2 + \gamma(1-a_2)](1-r)}. \tag{3.17}$$

This formula is obtained by updating Equation (3.16) with p_{mrk}, that is the probability of observing the crime mark given that the suspect's shoe is at the source, w, the probability of the shoe available for comparison purposes being the source, assuming the mark being left by the offender and that the suspect is the offender as well as by setting $\gamma' = \gamma$.

In the Bayesian network discussed here, p_{mrk} corresponds to $Pr(E_2 = e_2|F_2 = f_2)$ which was assigned the value of 1. The term w corresponds to $Pr(F_2 = f_2|G_2 = g_2, H_p)$ and is taken to be 0.5. The rarity of the characteristics, γ, is implemented in terms of $Pr(E_2 = e_2 \mid F_2 = \bar{f}_2)$, using a value of 0.015. The relevance term r corresponds to $Pr(G_2 = g_2)$ and is set to 0.99. Applying these values to Equation (3.17) leads to 33.28642. This result, too, is in entire agreement with the result derived from the proposed Bayesian network.

Next, consider the second option for a likelihood ratio development for E_2. As soon as one allows the probability of innocently leaving the footwear mark to be affected by knowledge about the source-level proposition of the fingermarks, the likelihood ratio for the second item of evidence E_2 needs to include the conditioning on E_1, as noted earlier in Equation (3.15). For such a case, assuming a structural relationship of the variables as implied by the Bayesian network shown in Figure 3.4, the following likelihood ratio applies:[20]

$$LR_{E_2=e_2|E_1=e_1} = \frac{wr + (h_2/f_2 - 1)^{-1}}{\left\{\frac{a_1(1-r_1)}{a_1(1-r_1)+\gamma[r_1+(1-a_1)(1-r_1)]}(a_2 - a_2') + a_2'\right\}(1-r_2) + (h_2/f_2 - 1)^{-1}}. \tag{3.18}$$

[19] Notice that the conditioning on $E_1 = e_1$ is omitted from notation because $a_2 = a_2'$ implies conditional independence of the two items of evidence.

[20] Further details on the derivation of this result is given in C.2.

This result clearly illustrates that, on a purely formal account, likelihood ratio formulae may become increasingly complex whenever dependencies (with respect to other items of evidence) need to be accounted for. Confining such calculations to Bayesian networks thus provides substantial support in evidential assessment.

For the currently discussed case, suppose that one can agree on an increased probability a_2, that is the suspect being the source of the footwear mark, if he is innocent (H_d) and the crime mark does not come from the offender (G_2), whenever it is already known that he is the source of the fingermarks (F_1). For the purpose of illustration, let $a_2 = 0.4$ while keeping $a_2' = 0.01$. In such a case, the likelihood ratio $LR_{E_2=e_2|E_1=e_1}$ is 33.1692 (result rounded), which is almost equivalent to $LR_{E_2=e_2}$ under conditions of conditional independence. This result should however not be taken as a suggestion to avoid the analysis of possible dependencies. It is often useful to investigate the effect of potential dependency relationships using varying assumptions, prior to deciding whether simplified calculations, based on independence assumptions, can reliably be retained.

In summary, thus, the likelihood ratio for the fingermark evidence, $LR_{E_1=e_1}$, and for the footwear mark evidence, $LR_{E_2=e_2|E_1=e_1}$, both support H_p rather than H_d. One is thus confronted with a situation of *convergence*. By multiplying the two component likelihood ratios as defined by Equation (3.15), a value of 4562 (value rounded) is obtained. The result for an assumption of conditional independence of the two items of evidence would be 4577, which would be slightly less conservative. In either case, the result shows that the probative value, which is limited to moderate for the individual items of evidence, becomes strong when considering the two items of evidence in combination.

Analysis of redundance

In cases of asymmetric independence between F_1 and F_2, one can investigate possible effects of redundance. In view of the expression defined earlier in Equation (3.14), the following relation can thus be invoked here:

$$R_{F_2=f_2|F_1=f_1} = 1 - \frac{\log LR_{F_2=f_2|F_1=f_1}}{\log LR_{F_2=f_2}}. \tag{3.19}$$

Start by considering the component likelihood ratio $LR_{F_2=f_2|F_1=f_1}$, which requires an extension to uncertainty about the relevance of the footwear mark evidence (G_2):

$$\begin{aligned} LR_{F_2=f_2|F_1=f_1} &= \frac{Pr(F_2=f_2|F_1=f_1,H_p)}{Pr(F_2=f_2|F_1=f_1,H_d)} \\ &= \frac{Pr(F_2=f_2|F_1=f_1,G_2=g_2,H_p)\overbrace{Pr(G_2=g_2)}^{r_2}+\overbrace{Pr(F_2=f_2|F_1=f_1,G_2=\bar{g}_2,H_p)}^{0}Pr(G_2=\bar{g}_2)}{\underbrace{Pr(F_2=f_2|F_1=f_1,G_2=g_2,H_d)}_{0}Pr(G_2=g_2)+\underbrace{Pr(F_2=f_2|F_1=f_1,G_2=\bar{g}_2,H_d)}_{a_2}\underbrace{Pr(G_2=\bar{g}_2)}_{1-r_2}}. \end{aligned}$$

Invoking notation introduced earlier in Section 3.4.2, that is $Pr(F_2=f_2|F_1=f_1,G_2=g_2,H_p)=w$, one thus obtains:

$$LR_{F_2=f_2|F_1=f_1} = \frac{wr_2}{a_2(1-r_2)}.$$

This result shows that, given H_p, the probability of the suspect's shoe being the source of the footwear mark is a consideration of the uncertainty about relevance of the crime mark, expressed in terms of r_2, and the probability of the shoe of interest being worn, expressed by w. Under the assumption that the suspect is not the author of the crime, H_d, the probability that the suspect's shoe is the source of the crime marks depends on the probability that the crime marks do *not* come from the offender (i.e., $1-r_2$) as well as the probability of leaving the mark for innocent reasons, a_2.

Invoking similar developments, one can find the likelihood ratio for F_2 without a conditioning on knowledge about F_1:

$$LR_{F_2=f_2} = \frac{wr_2}{a_1(1-r_1)(1-r_2)(a_2-a_2')+a_2'(1-r_2)} \; .$$

It appears worth noting that, in the case where $a_2 = a_2'$, the likelihood ratio $LR_{F_2=f_2}$ becomes

$$LR_{F_2=f_2} = \frac{wr_2}{a_1(1-r_1)(1-r_2)\underbrace{(a_2-a_2')}_{0}+a_2'(1-r_2)} = \frac{wr_2}{a_2'(1-r_2)} = \frac{wr_2}{a_2(1-r_2)} = LR_{F_2=f_2|F_1=f_1} \; .$$

It is readily seen that this will set the expression $R_{F_2=f_2|F_1=f_1}$, that is Equation (3.19), to zero. A graphical representation of this behaviour is given in Figure 3.5.

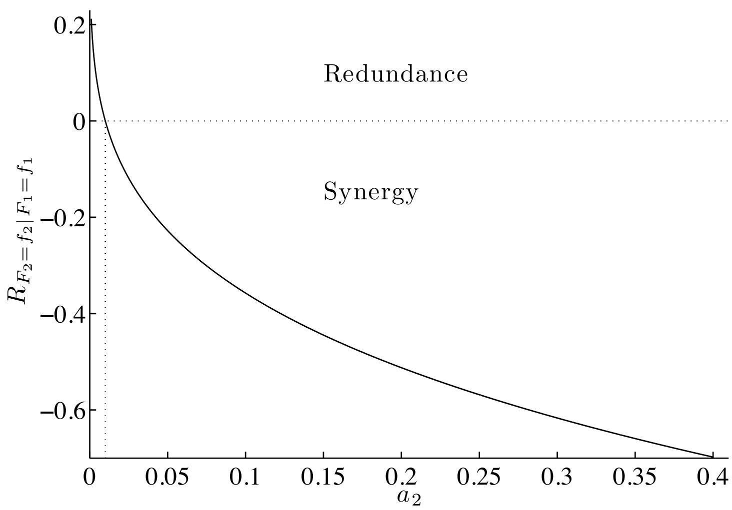


Figure. 3.5 – Representation of $R_{F_2=f_2|F_1=f_1}$ as a function of a_2

3.5 Discussion and conclusions

The joint evaluation of several items of evidence, as well as the examination of their individual contribution, is a requirement that follows naturally from the fact that items of evidence usually do not appear in isolated, but in concurrent ways. This gives rise to inferential interactions that go beyond those that may be encountered when looking only at isolated items of evidence. Inferential interactions such as redundance, synergy or directional change, are subtle topics that require an assessment on a case-based level in order to avoid possible instances of over- or underestimations. The isolated evaluation of evidence – when there is more than one item involved – may thus be unsafe when one is unaware of how they interact. Famous cases like People vs Collins provide illustrative examples for this [48]. In order to capture potential effects due to evidential interactions and to set collections of evidence appropriately in context, both forensic scientists as well as legal practitioners should thus take interest in approaching questions in the combination of evidence with particular awareness.

An additional layer of complication that adds to this outset is that the general context in which a joint assessment of scientific evidence ought to be operated, is probabilistic. This normative, prescriptive and inevitable requirement [97] helps to guard against a potentially fallacious, intuitive handling of uncertainties that is typical for unaided human reasoning. When considering one item of evidence, scientists should thus assure that their evaluative framework is amenable, on the one hand, for building upon existing knowledge, based on particular evidence, and, on the other hand, for a logical combination with forthcoming evidence, that is evidence that has not yet been considered. In addition, their approach must be able to cope with the fact that sources of uncertainty associated with the evaluation of one item of evidence may be relevant when considering another piece of evidence. This task is analytically demanding, but past and current directions of research in graphical (probabilistic) modeling offer some viable directions for approaching this challenge.

As argued throughout this paper, graphical models allow one to clarify and to obtain a concise representation of the structural dependencies assumed to hold among different aspects of evidence. The seminal works of Schum [125] constitute instances of analytical approaches to direct fundamental thinking about defining appropriate expressions of evidential value (i.e., in terms of likelihood ratio formulae). The generic patterns of reasoning defined and instantiated in terms of so-called ‘Schum graphs’ emerge also in current Bayesian network-based approaches. As a main additional feature with respect to Schum graphs, Bayesian networks offer a full coverage of an underlying probabilistic architecture that allows one to confine calculations entirely to the model (i.e., when implemented within a software environment). This allows reasoners to concentrate efforts and attention to building appropriate network structures. A probabilistic evaluation of the kind presented in this paper supports a refined understanding of the evidence one is examining and clarifies the inferential effects and entities one is confronted with. It also points out, contrary to the opinion of the Court of Appeal in R v T [106], that an evaluation does not necessarily depend upon data itself or precise numbers. Other levels of consideration, such as general principles of coherent reasoning about a given problem and logical formalization are also integral parts of evidential reasoning. It is indeed the Bayesian approach that even allows one to exploit tacit knowledge correctly.

As pointed out throughout this paper, even apparently ‘simple’ examples involving only two items of information (or, evidence) may entail a great variety of inferential issues to be explored. The Schum graphs allow one to recognise general patterns of inference towards a given target proposition whereas Bayesian networks support the extension of these ideas to more complex strains of inference. The latter include means for the probabilistic assessment of additional aspects such as the relevance of evidential material. A further important aspect of the use of Bayesian networks is that they can be shown to provide results that are in agreement with existing probabilistic procedures for the evaluation of single items of evidence. Bayesian networks thus provide a framework for implementing these approaches in practice, but also offer a possibility to combine distinct evaluative procedures coherently. This latter task would become increasingly difficult if it were approached on a purely algebraic level. Besides, the agreement between results from Bayesian networks and existing inferential approaches is also important for justifying particular Bayesian network models. This way of deriving inference models serves the purpose of illustrating that one can propose testable Bayesian networks whose properties are not arbitrary.

It thus appears that a primary advantage of a Bayesian network based approach to analyses of the combination of evidence consists in facilitating the locating, formal articulating and handling of relevant parameters. This can aid scientists to bring in a more secure position whenever they are required, for example, to explain the foundations of their reasoning and to evaluate the effect of specific parameter uncertainties. The recent judgment of the Court of Appeal in R v T [106] clearly illustrates that there is an ongoing need for this.

The intention of the authors here is not, at the moment, to propose such models for use in written reports or for presentation before trial. Working with Bayesian networks in forensic science is largely concerned with thinking about the way in which scientists assess evidence in the light of propositions relevant for a given actor within legal proceedings, and at a given juncture within a legal process. Something worthwhile has been gained if Bayesian networks can increase the level of insight in the inductive nature of this thought process.

4 Investigating evidential phenomena in combined evidence

ABSTRACT. This paper fuses elements of Irving J. Good's research on weight of evidence and David A. Schum's studies on evidence-based reasoning. The three main findings provided by such a fusion are methods to measure inferential interactions, dissonances among items of evidence, and relative contributions of items of evidence. All three measures are pragmatically defined in order to ensure a versatile applicability in inference tasks involving combined evidence or masses of evidence. These measures enable a more in-depth examination of recurrent phenomena in evidence-based reasoning, such as convergence, contradiction, redundancy, and synergy. Most of these phenomena have – to the best of the author's knowledge – either not been formally described to this date, or the formal description proposed remained of limited use for evidence-based reasoning tasks. The present research addresses this deficit in the current understanding and treatment of these evidential phenomena.

Keywords. Weight of evidence · evidence-based reasoning · inferential interaction · dissonance · relative contribution

4.1 Introduction

Schum's teaching on evidence-based inferences in general, and his book 'The Evidential Foundations of Probabilistic Reasoning' by David A. Schum [125] in particular, are references on reasoning about combined evidence and masses of evidence with an importance difficult to overstate. The present article intends to extend Schum's teachings on combined evidence and masses of evidence with a focus on what might best be summarized by 'evidence-based reasoning about masses of evidence given by its probabilistic foundation'. This phrase contains three underlying characteristics of this paper.

The first is the probabilistic foundation of evidence-based reasoning[1], as emphasized in existing discourses on the importance of probability theory for the reasoning about evidence [e.g., 125, 57, 97]. With respect to this, the object of investigation in this paper is to extend the consideration to the different evidential phenomena emerging in settings of combined evidence or masses of evidence.

Second, evidence-based reasoning is understood here, as an instance of probabilistic reasoning. That is, a probabilistic reasoning is said to be an evidence-based reasoning if the bearing of some observation(s) (i.e., evidence) on an issue of interest (e.g., hypotheses) is examined, but not if, say, the bearing of an issue of interest on possible observations is studied. In other words, the term 'evidence-based reasoning' is taken to refer to inductive (or abductive), diagnostic, or bottom-up inferences. The notion of inferential force of evidence, measured by the likelihood ratio (see Section 4.1.4), is directly relevant in the context of this kind of reasoning [e.g., 125, 93, 134].

[1]The present research adopts a subjectivist or Bayesian perspective [e.g., 9, 37].

Another form of the inferential force known in the discipline of law is the weight of evidence [81]. The notion of weight of evidence was coined by Irving J. Good [e.g., 58, 60]. The logarithm of the likelihood ratio corresponds to the weight of evidence. It has the advantage to be additive[2]. Good himself suggested the possibility to view the expected weight of evidence as a quasiutiliy like money that can be maximized if other utilities are unavailable or difficult to assess [66, 64].[3] One could consider the weight of evidence as a currency in the world of evidence-based reasoning. Thus, the weight of evidence has not only an intuitive appeal, but is also very instructive for investigating evidential phenomena occurring in masses of evidence.

Third, the concepts of 'combined evidence' and 'masses of evidence' are essentially the same in nature. The latter merely emphasizes the large number of items of evidence involved in a reasoning task so that a given mass of evidence is a particular instance of combined evidence standing out by its size. Indeed, all the evidential phenomena that this paper addresses can also be observed in the smallest possible combination of evidence, that is, within two items of evidence. Schum was the first to describe most of the evidential phenomena investigated here [125]. The contribution of this paper is to capture these phenomena more succinctly and render them measurable by using the notion of weight of evidence. The novel analysis methods of these phenomena proposed here, renders evidential subtleties more accessible, reveals evidential structures within reasoning patterns, and suggests general implications for evidence-based reasoning.

4.1.1 Relevance of the present research

The use of probabilistic expert systems for evaluating has become more and more common in forensic and legal practice [e.g., 134, 30, 94]. Most of the time the evaluation of evidence is confined to the derivation and examination of likelihood ratios or posterior probabilities of hypotheses. Combined evidence and masses of evidence, however, often exhibit evidential phenomena particular to a setting of combined evidence. These phenomena are likely to produce results that are generally difficult to grasp and trace [e.g., 38, 139]. In such cases, it might not be satisfying to confine the examination to the inferential force (or posterior probabilities) of the combined evidence. The authors believe that clear descriptions and analysis methods of such evidential phenomena can contribute greatly to a better understanding of evidential phenomena recurrently occurring in combined evidence and masses of evidence.

The next three sections outline basic notions of combined evidence, inferential force, and weight of evidence that form the starting point of the present investigation.

4.1.2 Natural redundance: A criterion for the generic combinations of arguments of evidence

Schum identified two generic structures of combinations of evidence based on the concept of natural redudnance [125]. They are depicted in Figure 4.1 as Bayesian networks [e.g., 108, 28, 83]. Each Bayesian network connects two reports, representing the actual evidence, to a hypothesis, representing the issues of interest. Schum calls such structures 'arguments of evidence' [125]. According to P. Tillers, the use of this word "... serves as a reminder that inference is a human construct and that complex formal analyses of evidence are also constructed, assembled, and constituted by human actors". [141, p. 1253] In this sense, arguments of evidence are often built upon multiple sub-arguments, where each sub-argument tackles a different aspect of evidence. Such arguments serve as frameworks for cascaded or hierarchical inferences [e.g., 127]. Prominent examples are the relevance of an event E for hypotheses of interest H, or the credibility of a given report R regarding the occurrence of an event E. The first type of sub-arguments are called *arguments of relevance* whereas the second are called *arguments of credibility*. The distinction stems from the epistemologically crucial observation that the perception of a relevant fact is different from the relevance of a fact for a hypothesis. The argument of credibility is concerned with observer

[2] The term 'weight' is often used shorthand for weight of evidence in this paper.

[3] However, the paper restricts the attention to the weight of evidence. That is, it does not extend the discussion to utility or loss functions.

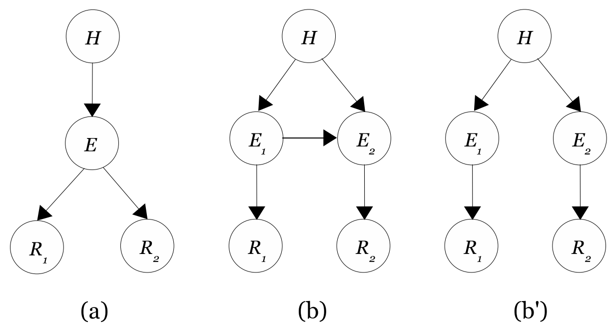


Figure. 4.1 – Generic ways of evidence combination in the form of Bayesian networks: (a) representation of a situation, in which reports R_1 and R_2 refer to the same event E; (b) situation where reports R_1 and R_2 refer to different events E_1 and E_2, which are further directly connected to each other; (b') situation where the events E_1 and E_2 are conditionally independent given H.

reliability and observer errors, but not the question about the relevance of a fact[4].

In situation (a), each of two reports R_1 and R_2, such as observations or sensor responses, refer to the same event E, which in turn, represents the relevant fact for the hypotheses of interest H. Hence, there are two arguments of credibility but only one argument of relevance. In situation (b), however, the two reports refer to different events each, that is E_1 and E_2 respectively, and each of these two events is relevant for the hypotheses of interest. Thus, there are two distinct lines of reasoning upon the hypotheses H. In addition, the events E_1 and E_2 are conditionally dependent given the hypotheses H, which is indicated by an edge connecting the node of E_1 to the node of E_2. Schum calls edges that directly connect two lines of argument *wefts* [125]. The situation depicted in (b') is a special case of (b), in which the two events are conditionally independent, given H, as indicated by the absence of a weft. In situation (a) R_1 and R_2 refer to the same *naturally redundant* event E. In situations (b) and (b') R_1 and R_2 refer to different *naturally nonredundant* events. Situation (a) is a special case of situation (b), in which $E_1 = E_2$ as shown by Schum [125]. For the derivations of the likelihood ratios from argument structures of (a) and (b') from (b) see D.1.

4.1.3 Harmonious and dissonant evidence

Schum identified two distinct forms of evidence in combination: harmonious and dissonant evidence [125]. *Harmonious evidence* refers to situations in which two or more reports support the same hypothesis over the alternative. If all reports refer to the same event, then the reports are said to form an instance of *corroborative evidence*. If, however, the reports refer to naturally nonredundant (i.e., different) events, then the reports are said to form an instance of *converging evidence*.

Unlike harmonious evidence, *dissonant evidence* refers to situations, in which two or more reports support different hypotheses. If the reports refer to naturally redundant events, then the reports form an instance of *contradicting evidence*. If the reports refer to naturally nonredundant events, then the reports form an instance of *conflicting evidence*.

[4] The assessment of observer errors plays an important role in evidence-based inferences. This is reflected by the fact research on that subject are conducted from different perspectives, such as with reference to utility theory [66], to signal detection theory [123], to multiple reports from unreliable sources [129], or false positive probabilities in DNA evidence [140]. All these studies assert that large observer error probabilities can eliminate the inferential force of even highly relevant items of evidence.

Thus, corroboration and contradiction inhabit the same type of argument structure depicted in Figure 4.1 (a). Convergence and conflict also inhabit the same type of argument structure depicted in Figures 4.1 (b) and (b'). Thus, corroboration and contradiction refer to relationships between arguments of credibility, as opposed to convergence and conflict, which refer to relationships between entire lines of reasoning.

4.1.4 Inferential force and weight of evidence

Harmonious and dissonant evidence differ by how the inferential forces of different items of evidence relate to each other. The inferential force of an item of evidence, such as a report, can be measured by the likelihood ratio (LR). The LR is a number that expresses if, and to what extent, an item of evidence supports one hypothesis over another. It appears in the odds form of the Bayes' theorem.

Let r ($r \in R = \{r, \bar{r}\}$) represent a report, $H = \{h, \bar{h}\}$ the hypotheses of interest, and I the background knowledge. The odds form of the Bayes' theorem is

$$\frac{Pr(h \mid r,I)}{Pr(\bar{h} \mid r,I)} = \frac{Pr(r \mid h,I)}{Pr(r \mid \bar{h},I)} \frac{Pr(h \mid I)}{Pr(\bar{h} \mid I)}, \tag{4.1}$$

where $Pr(h \mid r,I)/Pr(\bar{h} \mid r,I)$ and $Pr(h \mid I)/Pr(\bar{h} \mid I)$ are the posterior odds and prior odds respectively (where 'prior' means prior to receiving r and 'posterior' means after having received r). The fraction $Pr(r \mid h,I)/Pr(r \mid \bar{h},I)$ is the LR, the measure of the inferential force that r exerts on H. Following Good's notation, the LR will be written as $F(h : r \mid I)$ [e.g., 58, 64]. The ratios become additive when the logarithm of Equation 4.1 is taken

$$\log \frac{Pr(h \mid r,I)}{Pr(\bar{h} \mid r,I)} = \log F(h : r \mid I) + \log \frac{Pr(h \mid I)}{Pr(\bar{h} \mid I)}. \tag{4.2}$$

Good calls the logarithm of the LR the 'weight of evidence' [e.g., 58, 64]. By adopting Good's notation, the logarithm of the LR can written as $W(h/\bar{h} : r \mid I)$. It expresses the weight of evidence provided by r in favor of h as compared to $\bar{h}$ given I. Since the hypotheses are clear in the present case, one can retain the hypothesis in the numerator for expressing the weight of evidence. That is, one can write $W(h/\bar{h} : r \mid I) = W(h : r \mid I)$. [5]

The joint weight of evidence of n reports that are conditionally independent given H is given by the sum of the weight of evidence of reach report $r_i, i \in \{1, 2, ..., n\}$

$$W(h : r_1, r_2, ..., r_n \mid I) = \sum_{i=1}^{n} W(h : r_i \mid I). \tag{4.3}$$

If, however, the reports are not conditionally independent given H then

$$W(h : r_1, r_2, ..., r_n \mid I) = W(h : r_1 \mid I) + \sum_{i=2}^{n} W(h : r_i \mid r_1, ..., r_{n-1}, I), \tag{4.4}$$

which is a consequence of the *multiplication law of probability*.

The background knowledge I will be not be explicitly represented from now on in order to increase the clarity of the presentation. However, one should keep in mind that all probabilistic terms in this paper are conditioned by I.

[5] The base of the logarithm used in this paper is 10. Thus, the weight of evidence is measured in bans [65]. For example, a report that takes the value of LR of 100 has 2 bans (or short 2 b) of weight of evidence.

4.1.5 Outline of the paper

Section 4.2 introduces the framework and the notation of recurrently used variables. Section 4.3 presents Schum's redundance measure, and explains how it can be derived from the multiplication law of probability. This derivation enables a definition of a general measure of inferential interactions between any number of different items of evidence[6]. Section 4.4 introduces a measure for dissonant and harmonious evidence, and how it can be employed and interpreted. Section 4.5 shows a straightforward way to compare the magnitudes of weight of evidence stemming from different sets of evidence, and discusses evidential subtleties that can be observed on the basis of such a comparison. The discussion and conclusions are presented in the last section.

4.2 Framework for the analysis of masses of evidence

All generic structures for combining evidence considered here, dissonant or harmonious, involve a divergent connection[7], which ties two or more lines of argument to a single inference stage such as an intermediary event E or to ultimate hypotheses of interest H. The situation of Figure 4.1 (a) incorporates a divergent connection of the form $R_1 \leftarrow E \ \rightarrow R_2$, whereas the situations (b) and (b') incorporate divergent connections of the form $E_1 \leftarrow H \rightarrow E_2$. Since the divergent connection is the point where different lines of argument merge, it is natural to suppose that this point plays a crucial role in the emergence of different evidential phenomena such as harmony, dissonance, synergy, redundance, and directional change.

Sections 4.2.1 and 4.2.2 introduce the general representation of a divergent connection depicted in Figure 4.2 and a corresponding notation. This general representation serves as a template for measuring and analyzing different aspects of combined evidence.

4.2.1 General representation of a divergent connection

Figure 4.2 depicts a general representation of a divergent connection in the form of a Bayesian network, where the consideration of wefts is omitted at this juncture and all the variables are binary. Nodes A represent ancestor nodes, as opposed to descendant nodes denoted as D. An index k is assigned to each ancestor node in order to indicate the number of edges separating the ancestor node from the divergent connection. The ancestor node that is directly involved in the formation of a divergent connection has an index $k = 0$. The ancestor node farthest from the divergent connection has an index of $k = a$. Hence, one has A_k, where $k = \{0, 1, ..., a\}$.

Descendant nodes are indexed by i and j. The index i designates the line of argument (i.e. the branch number) to which the descendant node belongs. Thus, if n lines of argument (branches) are merged in a divergent connection, one has $i = \{1, 2, ..., n\}$. Index j indicates the number of edges separating a descendant node from a divergent connection. Descendant nodes that are directly involved in the formation of a divergent connection have the value $j = 0$. The descendant node of the line of argument i that is furthest away from the divergent connection is given the value $j = d_i$. This detailed notation is used in order to account for the fact that the number of edges d_i separating the node furthest away from the divergent connection depends on the line of argument i. Thus, one has $D_{i,j}$, where $i = \{1, 2, ..., n\}$ and $j = \{0, 1, ..., d_i\}$.

For example, applying this notational convention to the argument structure depicted in Figure 4.1 (a) allows one to obtain the following: the node H would be written as A_1, node E as A_0, and the nodes R_1 and R_2 as $D_{1,0}$ and $D_{2,0}$ respectively.

[6]The measure of inferential interaction refers to Schum's descriptions of inferential interaction. It is not to be confounded with interaction of weight of evidence as outlined in [59].

[7]A divergent connection in the context of Bayesian networks refers to an arrangement in which a single node has two or more direct descendants.

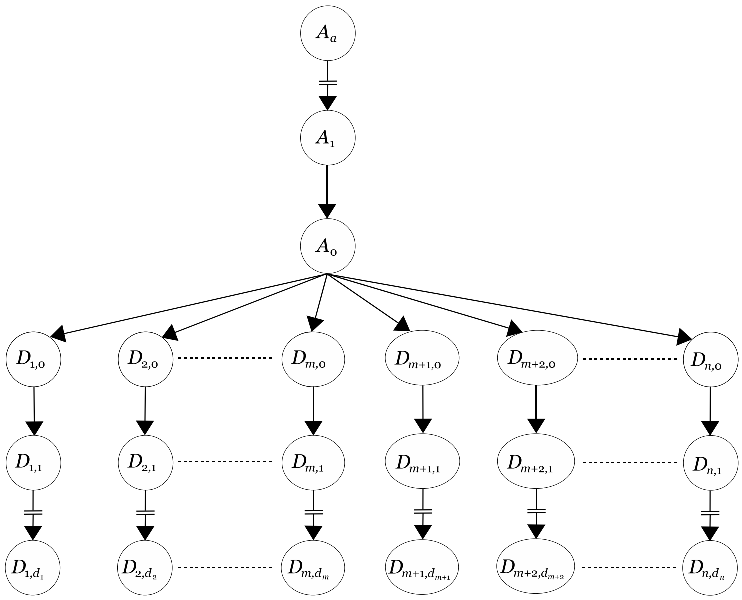


Figure. 4.2 – Bayesian network representing a divergent connection with multiple descendant nodes. Wefts, not shown explicitly here, may be present between any two descendant nodes.

Figure 4.2 further suggests a division of the lines of argument composed by descendant nodes into two groups, namely a group comprising lines of argument with indexes $i \in \{1,2,...,m\}$ and a second group with $i \in \{m+1, m+2,...,n\}$. Such a division facilitates the comparison of one group of lines of argument with another group. For instance, one group could contain lines of argument for which the items of evidence have already been evaluated, while the other group contains the lines of argument for which the items have yet to be evaluated.

4.2.2 Reports and hypothetical evidence

When analyzing an inference structure it is often helpful to enquire about the basis of hypothetical evidence. Consider, for example, Figure 4.1 (b). One can study a conclusion derived from the two reports R_1 and R_2 with regard to H. In contrast to this, one could also ask how a conclusion would change if, instead of R_1 and R_2, one of the events E_1 or E_2, or both were known. Such an analysis can be seen as an instance of evidence sensitivity analyzes [77]. It is driven by some hypothetical evidence claiming that E_1 or E_2 are known, as opposed to the actual evidence referred to as 'report' here. For the remaining parts of this paper, 'evidence' and 'item of evidence' are used as umbrella terms for both hypothetical evidence and actual reports.

All the variables considered in this paper are binary $D_{i,j} = \{d_{i,j}, \overline{d_{i,j}}\}$. Thus, $d_{i,j}$ (or $\overline{d_{i,j}}$) represents a single item of evidence. A set of evidence that considers all lines of argument is denoted $\mathbf{D}$. For example, if there are $n = 4$ lines of argument one could have a set of evidence $\mathbf{D} = \{d_{1,2}, d_{2,0}, d_{3,1}, d_{4,0}\}$. As indicated in Figure 4.2, the arguments are divided into two groups of lines of argument, namely one for which $i \in \{1,2,...,m\}$ and one for which $i \in \{m+1, m+2,...,n\}$. The subset of items of evidence that refers to the former group is denoted as $\mathbf{D}^m$ as it contains m items of evidence. The one referring to the latter group is denoted as $\mathbf{D}^{n-m}$ because it contains $n-m$ items of evidence. Thus, the subset $\mathbf{D}^{n-m}$ is the complement of $\mathbf{D}^m$ in $\mathbf{D}$, so that $\mathbf{D} = \mathbf{D}^m \cup \mathbf{D}^{n-m}$ holds.

If one considers only hypothetical evidence regarding descendants located at the same distance from the divergent connection, such as descendants that are immediately involved in the formation of a divergent connection, then it is instructive to make this apparent by specifying the distance j as a subscript for $\mathbf{D}$, that is $\mathbf{D}_0$, $\mathbf{D}_0^m$, and $\mathbf{D}_0^{n-m}$, where $\mathbf{D}_0 = \mathbf{D}_0^m \cup \mathbf{D}_0^{n-m}$. Similarly, $\mathbf{D}_1$, $\mathbf{D}_1^m$, and $\mathbf{D}_1^{n-m}$, where $\mathbf{D}_1 = \mathbf{D}_1^m \cup \mathbf{D}_1^{n-m}$, denotes the evidence on nodes separated from the converging connection by two edges downwards, and so on. One can indicate sets and subsets of evidence containing actual reports only, namely $\mathbf{D}_{d_i}$, $\mathbf{D}_{d_i}^m$ and $\mathbf{D}_{d_i}^{n-m}$, where $\mathbf{D}_{d_i} = \mathbf{D}_{d_i}^m \cup \mathbf{D}_{d_i}^{n-m}$. However, if a group of evidence stemming from descendants located at different distances is to be defined, such as a mixture of actual reports and hypothetical evidence, then no subscripts are used.

4.3 Measuring inferential interactions

Schum [125] introduced the so-called 'redundance measure' for two events that are conditionally dependent given some hypotheses H (e.g., E_1 and E_2 in Figure 4.1 (b)). This redundance measure $R_{e_2|e_1}$ is given by

$$R_{e_2|e_1} = \frac{W(h:e_2) - W(h:e_2 \mid e_1)}{W(h:e_2)} = 1 - \frac{W(h:e_2 \mid e_1)}{W(h:e_2)}, \tag{4.5}$$

where $W(h:e_2) \neq 0$. Equation 4.5 quantifies the difference between the weight of evidence in favor of h provided by e_2 alone, and the weight of evidence in favor of h provided by e_2 knowing that e_1 occurred, relative to $W(h:e_2)$. Schum showed that different values of $R_{e_2|e_1}$ indicate different types of inferential interactions between the events e_1 and e_2 given H (see D.2).

4.3.1 Explaining inferential interactions

Even though Equation 4.5 is intuitively appealing, it appears to be an *ad hoc* solution for the identification of inferential interactions. It does not provide, for example, an explanation for the different types of inferential interactions. This section captures and expresses inferential interactions based on the concept of weight of evidence.

By applying the multiplication law of probability, the weight of evidence that the events e_1 and e_2 provide in favor of h can be written as

$$W(h:e_1,e_2) = W(h:e_1 \mid e_2) + W(h:e_2) = W(h:e_2 \mid e_1) + W(h:e_1). \tag{4.6}$$

From this it follows that

$$W(h:e_2) - W(h:e_2 \mid e_1) = W(h:e_1) - W(h:e_1 \mid e_2). \tag{4.7}$$

The left side of Equation 4.7 corresponds to the numerator in Equation 4.5. Equation 4.6 stipulates further that $W(h:e_1 \mid e_2) = W(h:e_1,e_2) - W(h:e_2)$. The weight $W(h:e_1 \mid e_2)$ in Equation 4.7 is substituted by the difference $W(h:e_1,e_2) - W(h:e_2)$ to produce

$$W(h:e_1) + W(h:e_2) - W(h:e_1,e_2) = W(h:e_2) - W(h:e_2 \mid e_1). \tag{4.8}$$

Equation 4.8 covers the necessary properties to define types of inferential interactions in terms of weight of evidence in a single identity statement, providing thus a direct explanation for these interactions. In other words, Schum's measure of $R_{e_2|e_1}$ is essentially reformulating the properties of the multiplication law of probability. Table 4.1 compiles the implications of different types of inferential interactions with respect to $W(h:e_1,e_2)$ for cases where $W(h:e_1) > 0$ and $W(h:e_2) > 0$ (see D.3 for the derivations of these relationships from Equation 4.8). The conditional independence between the events e_1 and e_2 given $H = \{h, \bar{h}\}$ prevents e_1 and e_2 to interact inferentially.

Table. 4.1 – Types of inferential interactions in terms of weight of evidence and corresponding values for the degree of inferential interaction for cases where $W(h:e_1)>0$ and $W(h:e_2)>0$

Definition of inferential interaction	Inferential interaction in terms of $W(h:e_1,e_2)$
Synergy	
$W(h:e_2 \mid e_1) > W(h:e_2)$	$W(h:e_1,e_2) > W(h:e_1)+W(h:e_2)$
Conditional independence	
$W(h:e_2 \mid e_1) = W(h:e_2)$	$W(h:e_1,e_2) = W(h:e_1)+W(h:e_2)$
Partial redundance	
$0 < W(h:e_2 \mid e_1) < W(h:e_2)$	$W(h:e_1) < W(h:e_1,e_2) < W(h:e_1)+W(h:e_2)$
Complete redundance	
$W(h:e_2 \mid e_1) = 0$	$W(h:e_1,e_2) = W(h:e_1)$
Directional change	
$W(h:e_2 \mid e_1) < 0$	$W(h:e_1,e_2) < W(h:e_1)$

This property can be used to assess the inferential interaction with respect to H. Namely, one can measure the extent, to which the combined weight of evidence $W(h:e_1,e_2)$ deviates from the sum of the weights of evidence of e_1 and e_2, when they are assumed to be independent given H, that is, $W_{\perp}(h:e_1,e_2) = W(h:e_1)+W(h:e_2)$ (the notation $W_{\perp}(h:e_1,e_2)$ is used as a shorthand for $W(h:e_1,e_2 \mid \{E_1 \perp E_2\},H))$. The larger the deviation, the more important is the influence of the inferential interaction on the weight $W(h:e_1,e_2)$. In more formal terms, measuring the *degree of inferential interaction* between e_1 and e_2 amounts to computing the difference between the weights $W(h:e_1,e_2)$ and $W_{\perp}(h:e_1,e_2)$ so that $W_{\perp}(h:e_1,e_2)-W(h:e_1,e_2)$.

4.3.2 Generalization to more than two lines of argument

To extend the analysis of inferential interaction to several events, Section 4.2.1 is invoked as a framework. Let the set $\mathbf{D}$ denote the entire body of evidence. Thus, $W(a_0:\mathbf{D})$ denotes the joint weight of evidence in favor of a_0 provided by that body of evidence. In turn, let $W_{\perp}(a_0:\mathbf{D})$ denote the weight of evidence when all the events of set $\mathbf{D}$ are assumed to be conditionally independent given the event $A_0 = \{a_0,\bar{a}_0\}$. Further, let the subset $\mathbf{D}^m \subset \mathbf{D}$ subsume all those items of evidence contained in the body of evidence that are conditionally dependent given the proposition A_0 on at least one other item of evidence. Conversely, the subset $\mathbf{D}^{n-m} \subset \mathbf{D}$ represents all the items that are conditionally independent given event A_0. Equation 4.8, thus, becomes

$$W_{\perp}(a_0:\mathbf{D}) - W(a_0:\mathbf{D}) = W_{\perp}(a_0:\mathbf{D}^m) + W_{\perp}(a_0:\mathbf{D}^{n-m}) - [W(a_0:\mathbf{D}^m) + W(a_0:\mathbf{D}^{n-m})].$$

Since by definition $W_{\perp}(a_0:\mathbf{D}^{n-m}) = W(a_0:\mathbf{D}^{n-m})$ applies, one can rewrite

$$W_{\perp}(a_0:\mathbf{D}) - W(a_0:\mathbf{D}) = W_{\perp}(a_0:\mathbf{D}^m) - W(a_0:\mathbf{D}^m). \tag{4.9}$$

The weight $W_{\perp}(a_0:\mathbf{D}^m)$ is computed following Equation 4.3. In contrast, $W(a_0:\mathbf{D}^m$ is computed following Equation 4.4. Equation 4.9 exploits, therefore, again the difference between Equations 4.3 and 4.4, or stated otherwise, the difference in the application of the multiplication law for items conditionally independent and dependent given A_0. Equation 4.9 serves as the basis for the measure of inferential interaction denoted as $ia(a_0:\mathbf{D})$.

The weight $W_{\perp}(a_0 : \mathbf{D}^m)$ is used as a reference weight so that

$$ia(a_0 : \mathbf{D}) = \frac{W_{\perp}(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^m)}{W_{\perp}(a_0 : \mathbf{D}^m)} = 1 - \frac{W(a_0 : \mathbf{D}^m)}{W_{\perp}(a_0 : \mathbf{D}^m)}, \tag{4.10}$$

where $W_{\perp}(a_0 : \mathbf{D}^m) \neq 0$. Interestingly, the concept of evidential interaction as defined by (4.10) depends only on items of evidence conditionally dependent given A_0. Items conditionally independent given A_0 are not part of this expression. In other words, for conditionally independent items of evidence, Equation 4.10 will yield $ia(a_0 : \mathbf{D}^{n-m}) = 0$ (since $W(a_0 : \mathbf{D}^{n-m}) = W_{\perp}(a_0 : \mathbf{D}^{n-m})$ as indicated above).

4.3.3 Breaking up $W(a_0 : \mathbf{D}^m)$ into types of inferential interaction

Considerations on $W(a_0 : \mathbf{D}^m)$ can be taken a step further by breaking the weight up into components of different types of inferential interactions. For that purpose let $\mathbf{D}^m = \mathbf{D}^s \cup \mathbf{D}^{pr} \cup \mathbf{D}^{cr} \cup \mathbf{D}^{dc}$, where $\mathbf{D}^s \subseteq \mathbf{D}^m$ denotes the subset of synergistic evidence, $\mathbf{D}^{cr} \subseteq \mathbf{D}^m$ the subset of completely redundant evidence, $\mathbf{D}^{pr} \subseteq \mathbf{D}^m$ the subset of partially redundant evidence, and $\mathbf{D}^{dc} \subseteq \mathbf{D}^m$ the subset of evidence inducing directional change. By applying this partitioning of subsets to the weight $W(a_0 : \mathbf{D}^m)$, the weight components associated with different types of inferential interactions are rendered explicit as follows

$$W(a_0 : \mathbf{D}^m) = W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{cr}) + W(a_0 : \mathbf{D}^{pr}) + W(a_0 : \mathbf{D}^{dc}).$$

Thus, one can rewrite (4.10) as

$$ia(a_0 : \mathbf{D}) = 1 - \frac{W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}) + W(a_0 : \mathbf{D}^{cr}) + W(a_0 : \mathbf{D}^{dc})}{W_{\perp}(a_0 : \mathbf{D}^m)}, \tag{4.11}$$

where, $W_{\perp}(a_0 : \mathbf{D}^m) = W_{\perp}(a_0 : \mathbf{D}^s) + W_{\perp}(a_0 : \mathbf{D}^{pr}) + W_{\perp}(a_0 : \mathbf{D}^{cr}) + W_{\perp}(a_0 : \mathbf{D}^{dc}) \neq 0$. In order to examine the interaction value for each type of inferential interaction separately, the measure $ia(a_0 : \cdot)$ can be computed for any particular subset. For example, if one were interested in the value of the interaction measure for synergistic evidence, then one can compute

$$ia(a_0 : \mathbf{D}^s) = 1 - \frac{W(a_0 : \mathbf{D}^s)}{W_{\perp}(a_0 : \mathbf{D}^s)},$$

and analogously for the other subsets. Table 4.2 outlines different interaction values for each subset of evidence. A comparison of the former table with Table D.1 reveals that the interaction values and the values of $R_{e_2|e_1}$ are identical for each type of inferential interaction. This result suggests that the inferential interaction measure proposed here is a general form of Schum's measure (see Section 4.3). Note that for completely redundant evidence, one has $W(a_0 : \mathbf{D}^{cr}) = 0$, whereas $W_{\perp}(a_0 : \mathbf{D}^{cr}) \neq 0$.

4.3.4 Example: Interpreting values of $ia(a_0 : \mathbf{D})$.

Consider the Bayesian network in Figure 4.3. Four lines of argument lead to a query event A_0. The items of evidence subjected to the inference are $\mathbf{D} = \{d_{1,0}, d_{2,1}, d_{3,1}, d_{4,2}\}$. Equation 4.9 then becomes

$$\begin{aligned} W_{\perp}(a_0 : \mathbf{D}) - W(a_0 : \mathbf{D}) =& W(a_0 : d_{1,0}) + W(a_0 : d_{2,1}) + W(a_0 : d_{3,1}) + W(a_0 : d_{4,2}) \\ & - \left[W(a_0 : d_{1,0}) + W(a_0 : d_{2,1} \mid d_{1,0}) + W(a_0 : d_{3,1} \mid d_{1,0}, d_{2,1}) + W(a_0 : d_{4,2})\right] \\ =& \underbrace{W(a_0 : d_{2,1}) + W(a_0 : d_{3,1})}_{W_{\perp}(a_0 : \mathbf{D}^m)} - \underbrace{\left[W(a_0 : d_{2,1} \mid d_{1,0}) + W(a_0 : d_{3,1} \mid d_{1,0}, d_{2,1})\right]}_{W(a_0 : \mathbf{D}^m)} \end{aligned}$$

Table. 4.2 – Types of inferential interactions in terms of weight of evidence and corresponding values for the degree of inferential interaction for case where $W(a_0 : d_{i,\cdot}) > 0, d_{i,\cdot} \in \mathbf{D}$

Definition of inferential interaction	$ia(a_0 : \cdot)$
Synergy	
$W(a_0 : \mathbf{D}^s) > W_\perp(a_0 : \mathbf{D}^s)$	$ia(a_0 : \mathbf{D}^s) < 0$
Partial redundance	
$0 < W(a_0 : \mathbf{D}^{pr}) < W_\perp(a_0 : \mathbf{D}^{pr})$	$0 < ia(a_0 : \mathbf{D}^{pr}) < 1$
Complete redundance	
$W(a_0 : \mathbf{D}^{cr}) = 0$	$ia(a_0 : \mathbf{D}^{cr}) = 1$
Directional change	
$W(a_0 : \mathbf{D}^{dc}) < 0$	$ia(a_0 : \mathbf{D}^{dc}) > 1$

and the impact measure is given by

$$ia(a_0 : \mathbf{D}) = \frac{W(a_0 : d_{2,1}) + W(a_0 : d_{3,1}) - [W(a_0 : d_{2,1} \mid d_{1,0}) + W(a_0 : d_{3,1} \mid d_{1,0}, d_{2,1})]}{W(a_0 : d_{2,1}) + W(a_0 : d_{3,1})},$$

where $W(a_0 : d_{2,1}) + W(a_0 : d_{3,1}) \neq 0$. Suppose a case, in which $ia(a_0 : \mathbf{D}) = 0$ resulting, in turn, from $W_\perp(a_0 : \mathbf{D}^m) = W(a_0 : \mathbf{D}^m)$. In other words, the consideration of inferential interactions does not affect the overall weight $W(a_0 : \mathbf{D})$. However, the fact that $W_\perp(a_0 : \mathbf{D}^m)$ and $W(a_0 : \mathbf{D}^m)$ have the same magnitude, does not imply that there are no inferential interactions. A value of $ia(a_0 : \mathbf{D}) = 0$ could result, for example, from a situation, where $W_\perp(a_0 : \mathbf{D}^m) = W(a_0 : d_{2,1}) + W(a_0 : d_{3,1}) = 1 + 1 = 2$ and $W(a_0 : \mathbf{D}^m) = W(a_0 : d_{2,1} \mid d_{1,0}) + W(a_0 : d_{3,1} \mid d_{1,0}, d_{2,1}) = -2 + 4 = 2$. This implies that inferential interactions are present among different lines of argument[8], although the cumulative effect of the inferential interactions is nil. Thus, the measure of the cumulative effect has a reduced interpretative value. A value of $ia(a_0 : \mathbf{D}) = 0$, only indicates that the inferential interactions, if present, have no impact on the joint weight of evidence $W(a_0 : \mathbf{D})$ or $W(a_0 : \mathbf{D}^m)$ respectively.

However, in certain cases Equation 4.11 allows one to deduce the presence of effects of directional change and synergy. In particular, one can show that in cases where $W(a_0 : d_{i,j}) \leq 0, d_{i,j} \in \mathbf{D}^m$, an interaction value of $ia(a_0 : D) > 1$ implies that

$$ia(a_0 : \mathbf{D}) > 1 \iff -W(a_0 : \mathbf{D}^{dc}) > W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}) \Longrightarrow W(a_0 : \mathbf{D}^{dc}) \neq 0. \tag{4.12}$$

Similarly, a value of $ia(a_0 : \mathbf{D}) < 0$ implies that

$$ia(a_0 : \mathbf{D}) < 0 \iff W(a_0 : \mathbf{D}^s) > W_\perp(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^{pr}) - W(a_0 : \mathbf{D}^{dc}) \Longrightarrow W(a_0 : \mathbf{D}^s) \neq 0. \tag{4.13}$$

The proofs for the statements 4.12 and 4.13 are given in D.4.1 and D.4.2 respectively. The statements 4.12 and 4.13 suggest that in cases, where values of $ia(a_0 : \mathbf{D}) > 1$ or $ia(a_0 : \mathbf{D}) < 0$ are obtained, one can be certain of the presence of a directional change or synergy. However, these results also suggest that values other than the ones associated with directional change or synergy, do not allow one to conclude the absence of any inferential interactions (including directional change and synergy). Most of the time one cannot exclude the presence of inferential interactions on the basis of the impact measure unless the values of the weight components for each type of interaction are known. Thus, a given value of the impact measure quantifies only the cumulative impact of inferential interactions on the joint weight of evidence $W(a_0 : \mathbf{D})$.

[8] The inferential interactions involved are a directional change ($ia(a_0 : d_{2,1} \mid d_{1,0}) = (1 - (-2))/1 = 2$) and a synergy ($ia(a_0 : d_{3,1} \mid d_{1,0}, d_{2,1}) = (1-4)/(1) = -3$).

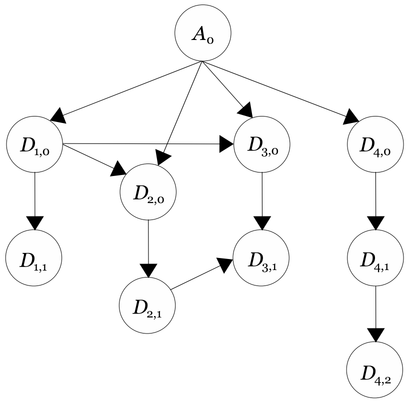


Figure. 4.3 – Bayesian network with four lines of argument leading to query event A_0. Three wefts are involved among the different lines of argument.

4.3.5 Corroborative redundance and the impact measure

Let the set $\mathbf{D} = \{d_{1,\cdot}, d_{2,\cdot}, \ldots, d_{n,\cdot}\}$ contain a series of n corroborative evidence each stemming from n different arguments of credibility. All arguments of credibility are independent given event A_0. Further, let the set $\mathbf{D}$ relate to the hypothesis of interest A_k where $k > 0$ through the intermediary event A_0. The weight of evidence in favor of hypothesis a_k provided by the evidence $\mathbf{D}$ is given by

$$W(a_k : \mathbf{D}) = W(a_k : d_{i,\cdot}) + \sum_{i=2}^{n} W(a_k : d_{1,\cdot} \mid \mathbf{D}^{i-1}),$$

which can be written as

$$W(a_k : \mathbf{D}) = W(a_k : d_{1,\cdot}) + \sum_{i=2}^{n} \log\left(\frac{Pr(a_0 \mid \mathbf{D}^{i-1}, a_k) + [F(a_0 : d_{i,\cdot}) - 1]^{-1}}{Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k}) + [F(a_0 : d_{i,\cdot}) - 1]^{-1}}\right), \tag{4.14}$$

where $F(a_0 : d_{i,\cdot})$ denotes likelihood ratio in favor of a_0 provided by the item of evidence $d_{i,\cdot}$. In turn, let $F(a_0 : \mathbf{D}^{i-1})$ denote the likelihood ratio in favor of a_0 provided by the set of evidence $\mathbf{D}^{i-1}$. By means of the Bayes' theorem, the probabilities $Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)$ and $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k})$ can be written as

$$Pr(a_0 \mid \mathbf{D}^{i-1}, a_k) = \frac{F(a_0 : \mathbf{D}^{i-1}) Pr(a_0 \mid a_k)}{F(a_0 : \mathbf{D}^{i-1}) Pr(a_0 \mid a_k) + Pr(\overline{a_0} \mid a_k)} \tag{4.15}$$

$$Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k}) = \frac{F(a_0 : \mathbf{D}^{i-1}) Pr(a_0 \mid \overline{a_k})}{F(a_0 : \mathbf{D}^{i-1}) Pr(a_0 \mid \overline{a_k}) + Pr(\overline{a_0} \mid \overline{a_k})} \tag{4.16}$$

where

$$F(a_0 : \mathbf{D}^{i-1}) = \prod_{i=1}^{n} \frac{Pr(d_{i,\cdot} \mid a_0)}{Pr(d_{i,\cdot} \mid \overline{a_0})}.$$

For the development see D.1 and D.5. In general, Equation 4.14 suggests that the weight of item $d_{i,\cdot}$ is bounded by the values $W(a_k : d_{i,\cdot} \mid D^{i-1}) = [0, \log \frac{Pr(a_0|\mathbf{D}^{i-1},a_k)}{Pr(a_0|\mathbf{D}^{i-1},\overline{a_1})}]$. More precisely, if item $d_{i,\cdot}$ is conclusive evidence for proposition a_0 (i.e., $Pr(d_{i,\cdot} \mid \overline{a_0}) = 0$), then its weight reduces to $W(a_k : d_{i,\cdot} \mid D^{i-1}) = \log(Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)/Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k}))$. If, however, the item $d_{i,\cdot}$ is completely uninformative (i.e. $Pr(d_{i,\cdot} \mid a_0) = Pr(d_{i,\cdot} \mid \overline{a_0})$), then its weight reduces to $W(a_k : d_{i,\cdot} \mid D^{i-1}) = \log 1 = 0$ (see also [125]).

A closer look at Equations 4.15 and 4.16 reveals an evidential subtlety that Schum named 'corroboratively redundant evidence'. Corroborative redundance refers to the process, in which an accumulation of weight leads to a decrease in weight for any further item of evidence in a corroborative setting. Stated more formally, in situations where $F(a_1, d_{i,\cdot}) > 1$ for all $d_{i,\cdot} \in \mathbf{D}$, the relationship $W(a_k : d_{i,\cdot} \mid \mathbf{D}^{i-1}) > W(a_k : d_{i,\cdot}) > 0$ holds. This implies that $W_\perp(a_k : \mathbf{D}) > W(a_k : \mathbf{D}) > 0$. Hence, the impact measure will yield $1 > ia(a_k : \mathbf{D}) > 0$. In turn, in situations where $F(a_1, d_{i,\cdot}) < 1$ for all $d_{i,\cdot} \in \mathbf{D}$, the relationship $W(a_k : d_{i,\cdot} \mid \mathbf{D}^{i-1}) < W(a_k : d_{i,\cdot}) < 0$ holds, and the value for impact measure is negative.

Note further that in the case, where the reduncance is measured on the level of the event A_0, that is, $ia(a_k : a_0)$ one obtains

$$ia(a_k : a_0) = \frac{W_\perp(a_k : a_0) - W(a_k : a_0)}{W_\perp(a_k : a_0)} = 1 - \frac{W(a_k : a_0)}{W_\perp(a_k : a_0)},$$

where $W(a_k : a_0) = W_\perp(a_k : a_0) \neq 0$. The measure $ia(a_k : a_0)$, thus, produces the value of exactly 1 indicating complete redundance (see Table 4.2). In all cases, in which the inferential interaction of corroboration is measured on the level of a_0, the measure of inferential interaction must take a value of 1, because a_0 is necessarily redundant with itself. Hence, if a value other than 1 is produced, then the arguments of credibility in $\mathbf{D}$ do not relate to the same event a_0, and one is not in the presence of a corroboration. This result is in agreement with Schum's findings on corroborative redundance [127].

4.4 Describing dissonance

For establishing a measure of dissonance and harmony, the authors chose to compare the 'expressed weight' against the 'weight potential' of the evidence. The expressed weight is the total amount of weight that the body of evidence provides in favor of an event. The weight potential is the total amount of weight generated by the body of evidence, regardless of which event is favored. Thus, if a body of evidence is enabled to express its entire weight potential, then the evidence is necessarily completely harmonious. However, if it can only express some fraction of its weight potential, then the evidence is dissonant. If no weight is expressed, although the evidence has a weight potential (i.e., the items of evidence are not entirely irrelevant), then the dissonance is maximal in a body of evidence. Note that the presence of conclusive evidence is not considered here. The involvement conclusive evidence produces a situation that cannot be logically reconciled under the given assumptions. The authors believe that such situations describe a paradoxical model, rather than dissonant evidence. A discussion on paradoxical models, however, is not an aim in this paper.

Recall that the weight of evidence is positive, if a given event or report supports the proposition in the numerator of the likelihood ratio (say a_k), and negative if it supports the proposition in the denominator (say $\overline{a}_k$). This allows for straightforward formulations of expressed weight and weight potential. Let $W_{\text{ex}}(a_k : \mathbf{D})$ denote the weight expressed by the evidence, and $W_{\text{pot}}(a_k : \mathbf{D})$ the weight potential of the body of evidence $\mathbf{D}$ so that

$$W_{\text{ex}}(a_k : \mathbf{D}) = |W(a_k : \mathbf{D})|, \tag{4.17}$$

$$W_{\text{pot}}(a_k : \mathbf{D}) = |W(a_k : d_{1,\cdot})| + \sum_{i=2}^{n} |W(a_k : d_{i,\cdot} \mid d_{1,\cdot}, \ldots, d_{n-1,\cdot})|. \tag{4.18}$$

Table. 4.3 – Different values of $W_{\text{diss}}(a_k : \mathbf{D})$, their implication on the relationship between $W_{\text{pot}}(a_k : \mathbf{D})$ and $W_{\text{ex}}(a_k : \mathbf{D})$, and on the characteristics of the body of evidence

$W_{\text{diss}}(a_k : \mathbf{D})$	$W_{\text{pot}}(a_k : \mathbf{D})$ and $W_{\text{ex}}(a_k : \mathbf{D})$	Characteristics
$= W_{\text{pot}}(a_k : \mathbf{D})$	$W_{\text{ex}}(a_k : \mathbf{D}) = 0$	Maximum dissonance
$< W_{\text{pot}}(a_k : \mathbf{D})$	$W_{\text{pot}}(a_k : \mathbf{D}) > W_{\text{ex}}(a_k : \mathbf{D})$	Dissonant evidence
$= 0$	$W_{\text{pot}}(a_k : \mathbf{D}) = W_{\text{ex}}(a_k : \mathbf{D})$	Harmonious evidence

Thus, the absolute values of weight provided by the evidence, is the amount of weight expressed (Equation 4.17). Conversely, the sum of the absolute values of the weights provided by each item is the potential amount of weight provided by the evidence (Equation 4.18). Now, if the set $\mathbf{D}$ relates to a body of completely dissonant evidence, then one has $W_{\text{ex}}(a_k : \mathbf{D}) = 0$. If the set $\mathbf{D}$ relates to a body of dissonant evidence, then one has $W_{\text{ex}}(a_k : \mathbf{D}) < W_{\text{pot}}(a_k : \mathbf{D})$. If, however, the set $\mathbf{D}$ relates to a body of harmonious evidence, then the two quantities are equal, that is, $W_{\text{ex}}(a_k : \mathbf{D}) = W_{\text{pot}}(a_k : \mathbf{D})$. Hence, the following relationship holds

$$W_{\text{ex}}(a_k : \mathbf{D}) \leq W_{\text{pot}}(a_k : \mathbf{D}), \tag{4.19}$$

which corresponds to the triangle inequality of real numbers.

A comparison can be achieved by formulating the measure of the amount of dissonance in the weight $W(a_k : \mathbf{D})$ as the difference between the weight potential and the expressed weight, that is,

$$W_{\text{diss}}(a_k : \mathbf{D}) = W_{\text{pot}}(a_k : \mathbf{D}) - W_{\text{ex}}(a_k : \mathbf{D}). \tag{4.20}$$

Equation 4.20 measures, therefore, the dissonance as the amount of potential weight lost due to dissonance in the body of evidence. From the Inequality 4.19, it follows directly that $0 \leq W_{\text{diss}}(a_k : \mathbf{D})$. Note also that if the items provide no weight at all (i.e., $W_{\text{pot}}(a_k : \mathbf{D}) = W_{\text{ex}}(a_k : \mathbf{D}) = 0$), then the weight lost due to dissonance is zero. Table 4.3 summarizes the relationship between a values of $W_{\text{diss}}(a_k : \mathbf{D})$ and the characteristics of the evidence in terms of harmony and dissonance.

Further, let $\mathbf{D}^m$ denote the subset of $\mathbf{D}$ that contains the items providing a positive weights, and $\mathbf{D}^{n-m}$ the subset containing the items providing negative weights. Thus, one has $W(a_k : \mathbf{D}^m) > 0$ and $W(a_k : \mathbf{D}^{n-m}) < 0$, which reduces Equation 4.20 to

$$W_{\text{diss}}(a_k : \mathbf{D}) = \begin{cases} 2W(a_k : \mathbf{D}^m), & \text{for } W(a_k : \mathbf{D}^m) < -W(a_k : \mathbf{D}^{n-m}), \\ W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m}), & \text{for } W(a_k : \mathbf{D}^m) = -W(a_k : \mathbf{D}^{n-m}), \\ -2W(a_k : \mathbf{D}^{n-m}), & \text{for } W(a_k : \mathbf{D}^m) > -W(a_k : \mathbf{D}^{n-m}). \end{cases} \tag{4.21}$$

A formal development leading to Equation 4.21 is given in D.6.

The measures of dissonance represent at the same time measures of contradiction or conflict (see Section 4.1.3). More precisely, if the argument structure under examination corresponds to the one of corroboration and contradiction, then naturally the measure of dissonance is a measure of contradiction. If the argument structure corresponds to the convergence and conflict model, however, then the measure of dissonance is a measure of conflict.

4.4.1 Comparing dissonances

When analyzing the effect of different probability values or different items of evidence in a given model, it might be more informative, to examine the weight lost through $W_{\text{diss}}(a_k : \mathbf{D})$ relative to the corresponding weight potential

$W_{\text{pot}}(a_k : \mathbf{D})$, rather than $W_{\text{diss}}(a_k : \mathbf{D})$ alone:

$$\frac{W_{\text{diss}}(a_k : \mathbf{D})}{W_{\text{pot}}(a_k : \mathbf{D})} = 1 - \frac{W_{\text{ex}}(a_k : \mathbf{D})}{W_{\text{pot}}(a_k : \mathbf{D})}.$$

Consider the example, in which one obtains $W_{\text{diss}}(a_k : \mathbf{D}) = 1$. There are different values for $W_{\text{pot}}(a_k : \mathbf{D})$ and $W_{\text{ex}}(a_k : \mathbf{D})$ to produce this amount of weight lost through. One could have $W_{\text{pot}}(a_k : \mathbf{D}) = 1.5$ and $W_{\text{ex}}(a_k : \mathbf{D}) = 0.5$, just as well as $W_{\text{pot}}(a_k : \mathbf{D}) = 8$ and $W_{\text{ex}}(a_k : \mathbf{D}) = 7$. In the former case the weight lost through relative to its weight potential is $W_{\text{diss}}(a_k : \mathbf{D})/W_{\text{pot}}(a_k : \mathbf{D}) \approx 0.67$. In the latter case, however, one obtains a quite different value: $W_{\text{diss}}(a_k : \mathbf{D})/W_{\text{pot}}(a_k : \mathbf{D}) \approx 0.13$. In other words, although the weight lost through is identical, the evidential situations producing the respective dissonances are different. Hence, the value of $W_{\text{diss}}(a_k : \mathbf{D})$ alone might not provide a meaningful comparison for dissonances in evidence.

4.4.2 Example: Directional change is a form of dissonant evidence

As pointed out in Section 4.3, an effect of directional change entails the combined weight of evidence of two items $d_{1,\cdot}$ and $d_{2,\cdot}$ to be smaller than that of $d_{1,\cdot}$ alone in situations, where all the items provide weight in favor of, say, a_0 ($W(a_0 : d_{1,\cdot}, d_{2,\cdot}) < W(a_0 : d_{1,\cdot})$). This a consequence of the fact that a directional change is defined as $W(a_0 : d_{2,\cdot} \mid d_{1,\cdot}) < 0$ (see D.3). It follows that the joint weight of the items $d_{1,\cdot}$ and $d_{2,\cdot}$ is necessarily smaller than the weight these items have in absolute terms ($|W(a_0 : d_{1,\cdot}) + W(a_0 : d_{2,\cdot} \mid d_{1,\cdot})| < |W(a_0 : d_{1,\cdot})| + |W(a_0 : d_{2,\cdot} \mid d_{1,\cdot})|$). By applying Equation 4.20 one obtains a weight lost due to dissonance larger than zero ($0 < W_{\text{diss}}(a_k : \mathbf{D})$). Hence, evidence involving directional change is necessarily dissonant to a certain degree. Cases of directional change where $W(a_0 : d_{1,\cdot}) < 0$ and $W(a_0 : d_{2,\cdot}) < 0$ lead to the same conclusion.

4.4.3 Example: Propagation of dissonance along two lines of reasoning

Let $d_{1,1} \in D_{1,1} = \{d_{1,1}, \overline{d_{1,1}}\}$ and $d_{2,1} \in D_{2,1} = \{d_{2,1}, \overline{d_{2,1}}\}$ designate the reports on the occurrence of the events $d_{1,0} \in D_{1,0} = \{d_{1,0}, \overline{d_{1,0}}\}$ and $d_{2,0} \in D_{2,0} = \{d_{2,0}, \overline{d_{2,0}}\}$ respectively. Both events are relevant for the hypotheses $A_0 = \{a_0, \overline{a_0}\}$. The two reports and the two events are independent given the hypotheses. The structure of the reasoning model corresponds, therefore, to the one depicted in Figure 4.1 (b'). Let $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ denote the amount of weight lost due to dissonance among the arguments of relevance. Conversely, let $W_{\text{diss}_1}(a_k : d_{1,1}, d_{2,1})$ refer to the amount of weight lost due to dissonance among the arguments of credibility. The weight lost due to dissonance among the arguments of evidence (i.e., among the two lines of reasoning) is given by the sum of the weight lost due do dissonance among each sub-argument. That is,

$$W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) = W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) + W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}), \tag{4.22}$$

where $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ and $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ are given by

$$\begin{aligned} W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) &= W_{\text{pot}}(a_0 : d_{1,1}, d_{2,1}) - W_{\text{ex}}(a_0 : d_{1,1}, d_{2,1}), \\ W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) &= W_{\text{pot}}(a_0 : d_{1,0}, d_{2,0}) - W_{\text{ex}}(a_0 : d_{1,0}, d_{2,0}). \end{aligned}$$

Note that $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ considers the potential weight and the expressed weight of the events that are reported, which is not necessarily the same to events favored by the weights of the reports. In order to study dissonances among descendant variables at varying distance from A_0, assume that the argument of evidence, relevance, and credibility of the first line of reasoning provide positive weight as shown in Table 4.4. Given such an initial situation, there are only three possible configurations for the second line of argument to produce a dissonance. These three configurations designated by (i)-(iii) are outlined in Table 4.4, and studied case by case (or configuration by configuration).

Table. 4.4 – Three possible configurations of the second line of argument ($d_{2,\cdot}$) allowing for dissonance given a fixed first line of argument ($d_{1,\cdot}$)

	$d_{1,\cdot}$	$d_{2,\cdot}$		
		(i)	(ii)	(iii)
Argument of evidence	$W(a_0 : d_{1,1}) > 0$	$W(a_0 : d_{2,1}) < 0$	$W(a_0 : d_{2,1}) < 0$	$W(a_0 : d_{2,1}) > 0$
Argument of relevance	$W(a_0 : d_{1,0}) > 0$	$W(a_0 : d_{2,0}) < 0$	$W(a_0 : d_{2,0}) > 0$	$W(a_0 : d_{2,0}) < 0$
Argument of credibility	$W(d_{1,0} : d_{1,1}) > 0$	$W(d_{2,0} : d_{2,1}) > 0$	$W(d_{2,0} : d_{2,1}) < 0$	$W(d_{2,0} : d_{2,1}) < 0$

Configuration (i) represents a situation, in which the arguments of relevance produce a dissonance. The dissonance is passed on the arguments of evidence. The amount of weight lost due to dissonance among the arguments of credibility is given by Equation 4.22. It can be shown that the weight lost due to dissonance $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1})$ can never be larger than zero (see D.6.1 for proof), that is,

$$W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) = W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) - W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) \not> 0.$$

The weight lost due to dissonance $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1})$ takes a value of exactly zero in most cases where $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) = W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$. In all the other cases $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ is smaller than $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$. This is a direct consequence of the fact that the weight of evidence for any $d_{\cdot,i}$ is bounded by the weight provided by preceding event $d_{\cdot,i-1}$ in a given line of reasoning (see D.6.1).

Configuration (ii) represents a situation, in which there is no dissonance between the arguments of relevance, but only between the arguments of evidence. Stated otherwise, one has $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = 0$, and Equation 4.22 becomes $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) = W_{\text{diss}_1}(a_k : d_{1,1}, d_{2,1}) > 0$.

Configuration (iii) represents a situation, in which the arguments of relevance produce a dissonance that is nullified by the arguments of credibility. Thus, in (iii) the arguments of evidence are harmonious. In other words, one has $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) = 0$, and therefore, $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = -W_{\text{diss}_1}(a_k : d_{1,1}, d_{2,1})$. As can be seen, the whole amount of weight lost due to dissonance between the arguments of relevance is reclaimed by the arguments of credibility. Hence, the arguments of credibility produce a negative loss of weight.

The present example shows that a dissonance between two lines of reasoning decreases as it propagates away from A_0 (configuration (i) and (iii)) and increases only if a dissonance emerges along two lines of reasoning (configuration (ii)). These properties of dissonance propagation remain valid irrespective of the length of the lines of reasoning in an inference task.

4.5 Measuring the relative contribution of evidence items

The authors deem the examination of the contribution of items of evidence relative to some other evidence to be informative in the context of harmonious evidence, but less so for dissonant evidence. The dissonance measure, discussed in the previous section, is better suited to examine dissonant evidence. Thus, the discussion regarding the relative contribution of evidence is confined to harmonious evidence.

Consider the framework of Section 4.2.1, where the set $\mathbf{D}$ denotes the entire body of evidence, and $\mathbf{D}^m$ and $\mathbf{D}^{n-m}$ two arbitrary subsets of $\mathbf{D}$. The weight provided by the body of evidence in favor of some event a_k can be written as

$$W(a_k : \mathbf{D}) = W(a_k : \mathbf{D}^m) + W(a_k : \mathbf{D}^{n-m}). \tag{4.23}$$

By expressing Equation 4.23 relative to $W(a_k : \mathbf{D})$ one can produce measures of contribution of weight of evidence in the form of weight ratios

$$1 = \frac{W(a_k : \mathbf{D}^m) + W(a_k : \mathbf{D}^{n-m})}{W(a_k : \mathbf{D})} = rc(a_k : \mathbf{D}^m/\mathbf{D}) + rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}), \tag{4.24}$$

where $W(a_k : \mathbf{D}) \neq 0$ because the evidence is harmonious. The relative contribution measure of the weight provided by $W(a_k : \mathbf{D}^m)$ relative to the contribution provided by the entire body of evidence $W(a_k : \mathbf{D})$ (i.e., $W(a_k : \mathbf{D}^m)/W(a_k : \mathbf{D})$) is denoted as $rc(a_k : \mathbf{D}^m/\mathbf{D})$. In turn, $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D})$ stands for the contribution of the weight $W(a_k : \mathbf{D}^{n-m})$ relative to $W(a_k : \mathbf{D})$. Finally, a third measure is considered, namely $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}^m)$, which is obtained by dividing the weight $W(a_k : \mathbf{D}^{n-m})$ by $W(a_k : \mathbf{D}^m)$ (or by dividing $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D})$ by $rc(a_k : \mathbf{D}^m/\mathbf{D})$). Therefore, the measures of relative contribution are ratios of different weights.

The measures $rc(a_k : \mathbf{D}^m/\mathbf{D})$ and $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D})$ take values between zero and one. That is, they become zero if either subset $\mathbf{D}^m$ or $\mathbf{D}^{n-m}$ produces no weight respectively (i.e., either $W(a_k : \mathbf{D}^m) = 0$ or $W(a_k : \mathbf{D}^{n-m}) = 0$). The fact that the weight of the body of evidence $W(a_k : \mathbf{D})$ cannot be zero means that, if $W(a_k : \mathbf{D}^m) = 0$, then one necessarily has $W(a_k : \mathbf{D}^{n-m}) = W(a_k : \mathbf{D})$. Conversely, if $W(a_k : \mathbf{D}^{n-m}) = 0$, then $W(a_k : \mathbf{D}^m) = W(a_k : \mathbf{D})$. Hence, in terms of relative contribution one asserts that a value of $rc(a_k : \mathbf{D}^m/\mathbf{D}) = 0$ implies $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}) = 1$ and vice versa. Finally, relative contribution values of $rc(a_k : \mathbf{D}^m/\mathbf{D}) = rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}) = 1/2$ are obtained in cases, where the two subsets provide the same amount of weight in favor of a_k (i.e., $W(a_k : \mathbf{D}^m) = W(a_k : \mathbf{D}^{n-m}) > 0$).

The measure $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$, however, takes a value between zero and infinity. The relative contribution value $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}^m) = 0$ applies in cases, in which the subset $\mathbf{D}^{n-m}$ provides no weight at all (i.e., $W(a_k : \mathbf{D}^{n-m}) = 0$). In turn, a value of $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}^m) = \infty$ is obtained if either the subset $\mathbf{D}^{n-m}$ provides conclusive evidence for event a_k ($W(a_k : \mathbf{D}^{n-m}) = \infty$), or if the subset $\mathbf{D}^m$ provides no weight at all ($W(a_k : \mathbf{D}^m) = 0$). Note also that a relative contribution value of $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}^m) = 1$ implies that both subsets provide equal amounts of weight ($W(a_k : \mathbf{D}^m) = W(a_0 : \mathbf{D}^{n-m}) > 0$).

4.5.1 The effect of diminishing relative contribution of additional items of evidence

The relative contribution measure reveals a subtlety that is less obvious, when dealing with few items of evidence, but becomes important as one is confronted with masses of evidence. The authors call this evidential subtlety 'the effect of diminishing relative contribution'.

Consider the extreme case, where $\mathbf{D}^m$ contains all the items of evidence but one, and $\mathbf{D}^{n-m}$ incorporates the remaining item of evidence $d_{i,n}$ among a total of n items of evidence. Moreover, assume that all lines of reasoning are independent given event A_k. Equation 4.23 suggests that weight provided by the body of evidence in favor of a_k is given by

$$\sum_{i=1}^{n} W(a_k : d_{i,j}) = \sum_{i=1}^{n-1} W(a_k : d_{i,j}) + W(a_k : d_{i,n}),$$

which in turn can be expressed in terms of relative contributions according to Equation 4.24

$$rc(a_k : \mathbf{D}^m/\mathbf{D}) = 1 - rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}).$$

Next, consider what happens if n tends towards infinity. In cases, where all the items of evidence provide weight of evidence in favor of a_k (i.e. $W(a_k : d_{i,j}) > 0$ for all $d_{i,j} \in \mathbf{D}$), this leads to the conclusion that the last item of evidence makes no contribution relative to the body of evidence

$$\underbrace{\lim_{n\to\infty^+} \frac{\sum_{i=1}^{n-1} W(a_k : d_{i,j})}{\sum_{i=1}^{n} W(a_k : d_{i,j})}}_{=1} = 1 - \underbrace{\lim_{n\to\infty^+} \frac{W(a_k : d_{i,n})}{\sum_{i=1}^{n} W(a_k : d_{i,j})}}_{=0}.$$

The same conclusion is obtained for $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$, that is,

$$rc(\mathbf{D}^{n-m}/\mathbf{D}^m) = \frac{rc(\mathbf{D}^{n-m}/\mathbf{D})}{rc(\mathbf{D}^m/\mathbf{D})} = \lim_{n\to\infty^+} \frac{W(a_k : d_{i,n})}{\sum_{i=1}^{n-1} W(a_k : d_{i,j})} = 0.$$

Identical conclusions can be made for cases in which all the items of evidence provide weight of evidence in favor of $\overline{a}^k$ ($W(a_k : d_{i,j}) < 0$ for all $d_{i,j} \in \mathbf{D}$).

As was observed, the contribution of item $d_{i,n}$ relative to $\mathbf{D}^m$ becomes zero, as n approaches infinity. Even in less extreme cases, the following general tendency can be anticipated: When the relative contribution measures are viewed as a function of m, the measure $rc(\mathbf{D}^{n-m}/\mathbf{D})$ decreases in proportion to the increase of $rc(\mathbf{D}^m/\mathbf{D})$ as m approaches n. The measure $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$, in turn, is initially larger than one. As m approaches n, it will eventually take a value smaller than one, as the weight provided by the subset $\mathbf{D}^{n-m}$ becomes smaller than the weight provided by the subset $\mathbf{D}^m$ ($W(a_k : \mathbf{D}^{n-m}) < W(a_k : \mathbf{D}^m)$). Finally, the measure $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$ takes a value of zero when m equates n.

4.5.2 Example: Relative contribution for items providing the same amount of weight and items providing different amounts of weight.

Assume that all the items in a body of evidence $\mathbf{D}$ are conditionally independent given event a_k. Assume further that each of these items provide the same amount of weight. In such a case the relative contribution becomes a simple ratio of the number of items in the corresponding set and subsets:

$$rc(\mathbf{D}^{n-m}/\mathbf{D}^m) = \frac{W(a_k : \mathbf{D}^{n-m})}{W(a_k : \mathbf{D}^m)} = \frac{\sum_{i=m+1}^{n} W(a_k : d_{i,j})}{\sum_{i=1}^{m} W(a_k : d_{i,j})} = \frac{(m-n)W(a_k : d_{i,j})}{mW(a_k : d_{i,j})},$$

and because $W(a_k : d_{i,\cdot}) = W(a_k : d_{i,\cdot})$ one obtains

$$rc(\mathbf{D}^{n-m}/\mathbf{D}^m) = \frac{n-m}{m} = \frac{n}{m} - 1.$$

Analogously one obtains $rc(\mathbf{D}^m/\mathbf{D}) = m/n$ and $rc(\mathbf{D}^{n-m}/\mathbf{D}) = (n-m)/n = -m/n + 1$. Hence, the measure $rc(\mathbf{D}^m/\mathbf{D})$ is an increasing linear function of m, $rc(\mathbf{D}^{n-m}/\mathbf{D})$ is a decreasing linear function of m, and $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$ is a decreasing exponential function of m. The two graphs in first column of Figure 4.4 depict the curves of the relative contributions $rc(\mathbf{D}^m/\mathbf{D})$ (in grey), $rc(\mathbf{D}^{n-m}/\mathbf{D})$ (in black), and $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$ respectively for $n = 10$ items of evidence providing the same weight.

These properties of the relative contribution measures can be observed to a certain extent, even when the items of evidence do not provide the same weight. Take a body of evidence $\mathbf{D}$ containing $n = 10$ items. The weight of the items is $W(a_k : d_{i,\cdot})_{i\in\{1,2,\ldots,n\}} = \{0.8, 1.1, 1.7, 2.7, 0.6, 2.7, 2.8, 2.0, 1.9, 0.3\}$. Again, all these ten items are assumed to be independent given event a_k. The division into subsets $\mathbf{D}^m$ and $\mathbf{D}^{n-m}$ is applied. The two graphs in the second column of Figure 4.4 depict the curves of the relative contributions $rc(\mathbf{D}^m/\mathbf{D})$ (in grey), $rc(\mathbf{D}^{n-m}/\mathbf{D})$ (in black), and $rc(\mathbf{D}^{n-m}/\mathbf{D}^m)$ respectively.

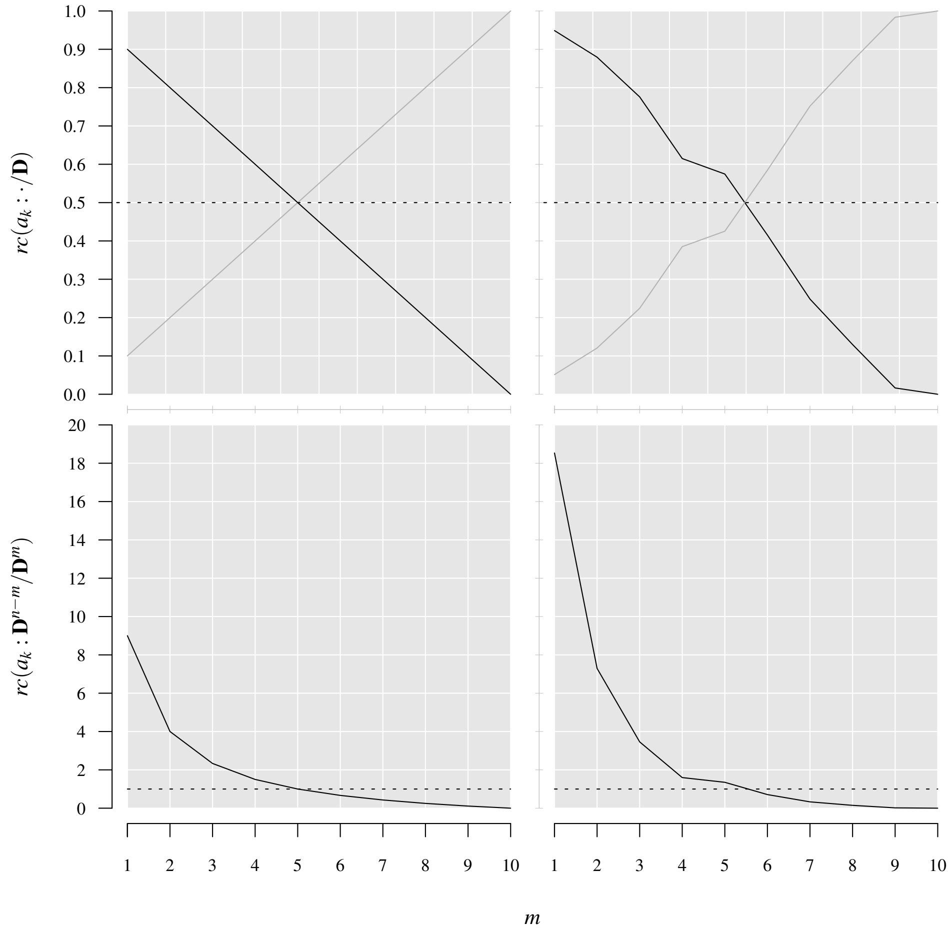


Figure. 4.4 – Graphs of relative contribution measures as m progresses towards n. The graphs in the first row side represent $rc(a_k : \mathbf{D}^m/\mathbf{D})$ (in grey) and $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D})$ (in black). The graphs in the second row represent $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}^m)$. The dotted lines indicate where $W(a_k : \mathbf{D}^m) = W(a_k : \mathbf{D}^{n-m})$, i.e. $rc(a_k : \mathbf{D}^m/\mathbf{D}) = rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}) = 0.5$, and $rc(a_k : \mathbf{D}^{n-m}/\mathbf{D}) = 1$.

4.6 Diminishing relative contribution and corroborative redundance

Section 4.3.5 pointed out that an accumulation of weight of evidence leads to a decrease in weight for any further item of evidence in a body of corroborative evidence. This effect was called corroborative redundance. Such a decrease in weight is not to be confounded with the effect of diminishing relative contribution of additional items of evidence in a body of harmonious evidence. The name already indicates: corroborative redundance is a form of redundance and leads to a decrease in weight. However, this not the case for the effect of diminishing relative contribution. This effect leads to a decrease in relative contribution, but not to a decrease in weight. However, these two effects are closely related. Imagine one were to examine only the arguments of credibility in a corroborating body of evidence, then it is not an effect of corroborative redundance that emerges, but an effect of diminishing contribution relative to the event A_0. Yet, if the examination of the same body of evidence extends to the argument of evidence, that is, from $\mathbf{D}$ to, say, A_1, then an instance of corroborative redundance will emerge. It appears that these two evidential subtleties differ by the perspective taken in situations of corroborative evidence. Nonetheless, the two effects are quite different because a different perspective implies the consideration of a different conditional dependence relationship in the reasoning pattern. In fact it is even possible to trace the origin of each effect, that is, diminishing relative contribution and corroborative redundance, to specific terms that the weight of corroborative evidence incorporates.

4.6.1 The origin of the effects of diminishing relative contribution and corroborative redundance in corroborative evidence

Consider first the relative contribution of items of evidence that are assumed to be independent given a_1, that is, a body of corroborative evidence is taken as a body of convergent evidence. The contribution of the nth item relative to the entire body of evidence is

$$rc_{\perp}(a_1 : d_{n,\cdot}/\mathbf{D}^{n-1}) = \frac{W_{\perp}(a_1 : d_{n,\cdot})}{W_{\perp}(a_1 : \mathbf{D})}.$$

The weight $W_{\perp}(a_1 : \mathbf{D}^{i-1})$ cannot be zero since the body of evidence is corroborative or ‘convergent’. Section 4.5.1 pointed out that if all the items of evidence provide the same weight, then the relative contribution becomes the ratio of the number of items in each the subsets. In the present case this is that $rc_{\perp}(a_1 : d_{n,\cdot}/\mathbf{D}^{n-1}) = 1/n$. However, the relative contribution of corroborative evidence is

$$rc(a_1 : d_{n,\cdot}/\mathbf{D}) = \frac{W(a_1 : d_{n,\cdot} \mid \mathbf{D}^{n-1})}{W(a_1 : \mathbf{D})} = \frac{\log \frac{Pr(a_0|\mathbf{D}^{i-1},a_1)+\left[F(a_0:d_{i,\cdot})-1\right]^{-1}}{Pr(a_0|\mathbf{D}^{i-1},\overline{a_1})+\left[F(a_0:d_{i,\cdot})-1\right]^{-1}}}{\log \frac{Pr(a_0|a_1)+[F(a_0:\mathbf{D})-1]^{-1}}{Pr(a_0|\overline{a_1})+[F(a_0:\mathbf{D})-1]^{-1}}}.$$

More precisely, in corroborative evidence the items $d_{i,\cdot}$ are conditionally dependent given event A_1, while this is not the case if A_0 is given. Thus, if the arguments from $\mathbf{D}$ to A_0 are considered, then the weight of each additional evidence is bounded by $W(a_0 : d_{i,\cdot}) =]0,\infty]$. [9] However, in situations of corroboration the weight of each additional evidence $d_{i,\cdot}$ is bounded by the values $W(a_1 : d_{i,\cdot} \mid D^{i-1}) = [0, \log \frac{Pr(a_0|\mathbf{D}^{i-1},a_1)}{Pr(a_0|\mathbf{D}^{i-1},\overline{a_1})}]$ (see Section 4.3.5). Consider the ratio

$$\frac{Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)}{Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k})} = \frac{\frac{F(a_0:\mathbf{D}^{i-1})Pr(a_0|a_k)}{F(a_0:\mathbf{D}^{i-1})Pr(a_0|a_k)+Pr(\overline{a_0}|a_k)}}{\frac{F(a_0:\mathbf{D}^{i-1})Pr(a_0|\overline{a_k})}{F(a_0:\mathbf{D}^{i-1})Pr(a_0|\overline{a_k})+Pr(\overline{a_0}|\overline{a_k})}}. \tag{4.25}$$

[9] Note that since the evidence is corroborative it is assumed that the weight that each items provides in favor of a_0 is larger than zero.

As can be seen, two types of ingredients determine the ratio in question. These are, notably, probabilities and likelihood ratios. The influence of the probabilities is considered first, and the influence of the likelihood ratios afterwards.

If the probabilities take values of $Pr(\overline{a_0} \mid a_k) = 0$ and $Pr(a_0 \mid \overline{a_k}) = 0$ (stated otherwise $Pr(a_0 \mid a_k) = Pr(\overline{a_0} \mid \overline{a_k}) = 1$), then one obtains $Pr(a_0 \mid \mathbf{D}^{i-1}, a_k) = 1$ and $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k}) = 0$. In such cases, the limit of the ratio in Equation 4.25 is infinity. Hence, one has $W(a_1 : d_{i,\cdot} \mid D^{i-1}) = W(a_1 : d_{i,\cdot}) =]0,\infty]$. This implies that the situation is essentially one of convergence and not of corroboration. Thus, as $Pr(\overline{a_0} \mid a_k)$ and $Pr(a_0 \mid \overline{a_k})$ approach zero, the weight $W(a_1 : d_{i,\cdot} \mid D^{i-1})$ approaches that of $W(a_1 : d_{i,\cdot})$. Stated otherwise, the evidence transitions from a situation of corroboration towards a situation of convergence. At the same time the effect of corroborative redundance diminishes, whereas that of diminishing relative contribution amplifies. The extent to which the latter effect takes place in corroborative evidence is, therefore, ultimately defined by the probabilities $Pr(\overline{a_0} \mid a_k)$ and $Pr(a_0 \mid \overline{a_k})$.

The combined likelihood ratio $F(a_0 : \mathbf{D}^{i-1})$, however, weakens the influence of the probabilities in Equation 4.25. That is, the stronger the inferential force conveyed by the likelihood ratios, the closer the values taken by probabilities $Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)$ and $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k})$ are to one, and conversely, the smaller the weight provided by item $d_{n,\cdot}$ (i.e., $W(a_1 : d_{n,\cdot} \mid \mathbf{D}^{n-1})$). In the extreme case, where $F(a_0 : \mathbf{D}^{i-1})$ tends towards infinity, the limit of those probabilities is exactly one, and the weight provided by item $d_{n,\cdot}$ is zero. This suggests that the combined likelihood ratio $F(a_0 : \mathbf{D}^{i-1})$ reduces the weight provided by item $d_{n,\cdot}$. Thus, the effect of corroborative redundance is instigated by likelihood ratio $F(a_0 : \mathbf{D}^{i-1})$. In fact, the stronger the inferential force provided by $F(a_0 : \mathbf{D}^{i-1})$, the stronger the effect of corroborative redundance on item $d_{n,\cdot}$.

The interplay between the said probabilities and the likelihood ratios, and therefore, the interplay between the two effects in question, appear rather intricate. Nonetheless, one can anticipate that the effect of corroborative redundance will quickly and strikingly gain the upper hand as the items of evidence accumulate. Consider again the probabilities $Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)$ and $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k})$ in ratio in Equation 4.25. The combined likelihood ratio $F(a_0 : \mathbf{D}^{i-1})$ assumes a value in the range of $]1,\infty]$ in corroborative evidence [10], whereas probabilities assume values between zero and one by definition. Given that $F(a_0 : \mathbf{D}^{i-1})$ appears in the numerator and denominator of the probabilities $Pr(\overline{a_0} \mid a_k)$ and $Pr(a_0 \mid \overline{a_k})$, its influence will rapidly overpower the influence of the probabilities $Pr(\overline{a_0} \mid a_k) = 0$, $Pr(a_0 \mid \overline{a_k}) = 0$, and their corresponding complements. The following example illustrates the interplay between the effect of diminishing relative contribution and that of corroborative redundance.

4.6.2 Example: The effect of corroborative redundance on relative contribution

The present example compares the relative contribution measure and the inferential interaction measure as functions of the number of items of evidence n that constitute a body of evidence $\mathbf{D}$. Each item of evidence produces the same inferential force in favor of a_0. Stated more formally, one has $F(a_0 : d_{i-1,\cdot}) = F(a_0 : d_{i,\cdot}) > 1, i = 2,3,...,20$. Figure 4.5 depicts the result in four graphs.

The graphs in the upper row show the changes in relative contribution as the number of items n constituting the body of evidence increases. The graphs in the lower row show the corresponding changes in the inferential interaction measure. The graphs in the left column show the impact of varying values of the inferential force produced by the argument of relevance, that is, $F(a_1 : a_0) = Pr(a_0 \mid a_1)/Pr(a_0 \mid \overline{a_1})$. As indicated in the previous section, the probabilities $Pr(a_0 \mid a_1)$ and $Pr(a_0 \mid \overline{a_1})$ define the extent to which the effect of diminishing relative contribution takes place. The graphs in the right column show the impact of varying values of the inferential force of the arguments of credibility ($F(a_0 : d_{n,\cdot})$). Remember that $F(a_0 : d_{n,\cdot})$ defines the strength of the effect of corroborative redundance in the evidence. The black curves in the four graphs represent identical situations defined by the values $F(a_0 : d_{n,\cdot}) = 10$ and $F(a_1 : a_0) = 10^5$. The dashed curves in the upper row show the contribution of

[10] Note that cases, where each item supports the opposite event $\overline{a_0}$ so that $F(a_0 : \mathbf{D}^{i-1}) = [0,1[$, lead to the same conclusion. In fact the closer the combined likelihood ratio takes a value to zero, the closer the probabilities $Pr(a_0 \mid \mathbf{D}^{i-1}, a_k)$ and $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_k})$ take values to zero, and the closer the weight provided item $d_{n,\cdot}$ is to zero.

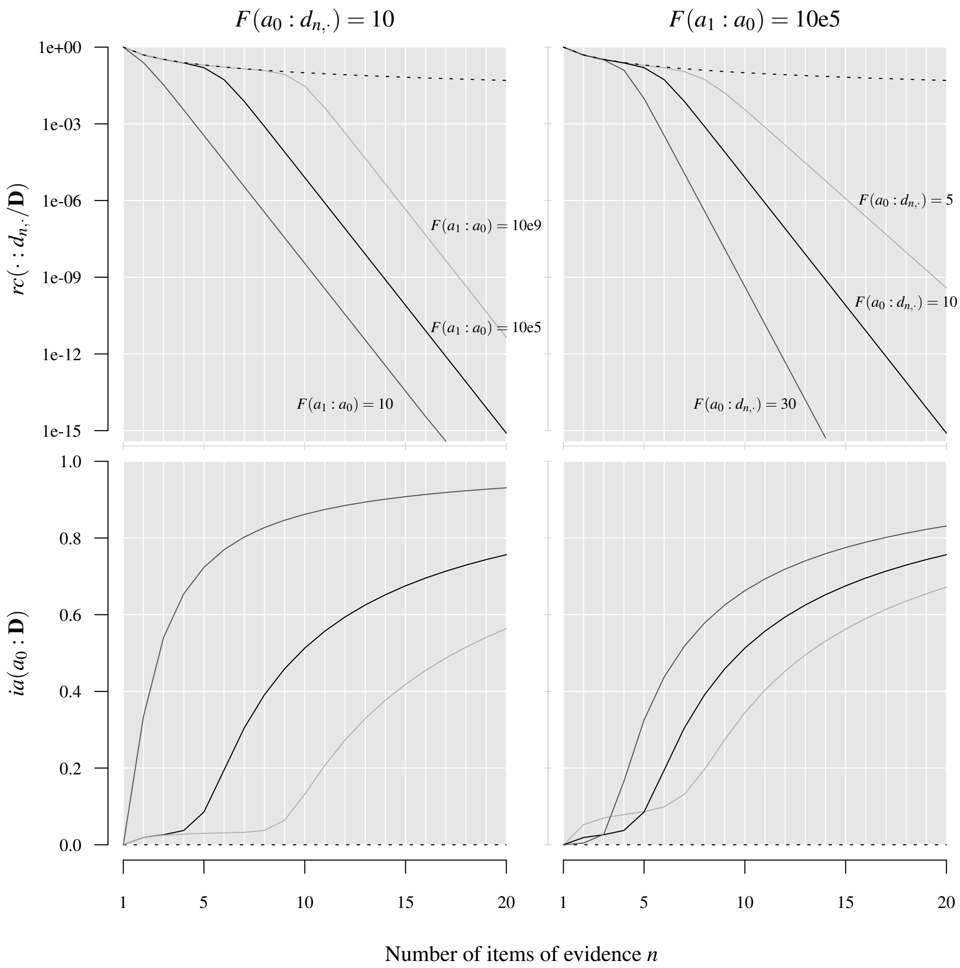


Figure. 4.5 – Relative contribution measure (upper row) and the corresponding inferential interaction measure (lower row) as a function of number of items n in a body of evidence

the nth item relative to $\mathbf{D}$ regarding event a_0 (i.e., $rc(a_0 : d_{n,\cdot}/\mathbf{D})$). Since all the items are independent given A_0, one has $rc(a_0 : d_{n,\cdot}/\mathbf{D}) = 1/n$. Hence, the decrease in relative contribution depicted by this curve is exclusively due to the effect of diminishing relative contribution. The corresponding inferential interaction measure (graphs in the lower row), in turn, remains constant at zero indicating conditional independence among the items given A_0.

As can be seen in the graphs of the upper row, the inferential force of the argument of relevance $F(a_1 : a_0)$ defines degree, to which the items are subjected to the effect of diminishing relative contribution before being entirely subjected to the effect of corroborative redundance. However, the strength of the effect of corroborative redundance itself is defined by the strength of the inferential force of the argument of credibility $F(a_0 : d_{m,\cdot})$. As can be seen in the upper right graph, the stronger the inferential force of the arguments of credibility the more accentuated is the relative contribution of the nth item.

The graphs in the lower row show the inferential interaction among the items of evidence. As expected, values of $ia(a_0 : \mathbf{D})$ for corroborative redundant evidence take values in the interval $]0, 1]$, indicating the presence of redundance (see Section 4.3.5). In the lower left graph it can be seen that the stronger $F(a_1 : a_0)$ is, the stronger the tendency of items to adhere to the effect of diminishing relative return, before giving in entirely to the effect of corroborative redundance, which is marked by a sharp increase in the value of $ia(a_0 : \mathbf{D})$. The lower right graph shows that the stronger $F(a_0 : d_{m,\cdot})$ is, the earlier the onset of the effect of corroborative redundance.

The example further suggests that the inferential interaction measure is sensitive enough to capture even weak tendencies of redundance, and to distinguish between the effects of corroborative redundance and diminishing relative contribution.

4.7 Conclusion

The importance of understanding evidential phenomena occurring in combined evidence is well known by practitioners of law, who are involved in evidence-based reasoning on a daily basis ([e.g., 146]), and statisticians, who are engaged in the subject of evidence interpretation (see for example Lindley's Foreword in [4, p. xxiv]). However, the scientific literature on these subjects remains sparse even today.

The paper proposes descriptions and methods of quantification for evidential phenomena recurrently occurring in combined evidence and masses of evidence based on the logic of probability theory. The notion of Irving J. Good's weight of evidence turned out to be extremely useful for capturing and clarifying such evidential phenomena [e.g., 58, 64]. Many of these phenomena were first described by David A. Schum from a probabilistic point of view [125]. These phenomena include different types of inferential interactions, dissonances, and the relative contribution of items of evidence. The present research provides an intuitive and succinct description of these phenomena by exploiting of the additive property of the weight of evidence. The measures developed on that basis are convenient methods to analyze diverse properties of combined evidence. The general formulations of these measures widen the scope of application to masses of evidence. A recursive and complementary application of these measurements to sets and subsets of evidence can be envisaged in order to clarify and deepen the understanding of large bodies of evidence not only as a whole, but also of its particular details. That is, they allow not only to identify, but also to quantify particular relationships among items of evidence, such as synergy, directional change, and dissonance, or general tendencies and regularities underlying the process of evidence accumulation regarding the strength of items of evidence, such as the effect of diminishing relative contribution and different forms of redundance.

Such measures cannot only improve the understanding of combined evidence and masses of evidence in particular cases, but also help investigating and handling combined evidence from a more general perspective. Consider, for example, the following questions that might be pursued in the future: Does the human mind recognize inferential interactions among items of evidence? If not, what systems and processes can be established to remediate this problem? If yes, what are the conditions that allow the human mind to do so? How, can these conditions be optimally exploited in probabilistic expert systems? Or: How often can dissonance be observed in combined evidence? Is dissonance a common feature of all types of evidence, or is it more common in one type of evidence, but less common in other types of evidence (such as DNA evidence and fingermark evidence)? If dissonance is more common in some types of evidence, then why is that so, and what are the implications for the use of this type of evidence?

Hence, the authors believe that the measures and methods proposed in this paper can be of great help to a reasoner engaged in evidence-based reasoning. The simplicity of the proposed measures enables a straightforward implementation in automated expert systems and other software environments, making the present research readily applicable to real life problems.

Acknowledgements

This work has beed supported by the Swiss National Science Foundation [grant number 100014_150276].

5 Aggregating DNA evidence: Probabilistic analysis of the DNA evidence in *State of Texas vs Josiah Sutton*

ABSTRACT. The present paper proposes a joint analysis of four DNA typing results stemming from the case of *State of Texas vs Josiah Sutton*. The case involves a sexual assault that took place in 1998. The forensic expert, who examined the DNA typing results at the time, stated in court that a suspect named Josiah Sutton could not be excluded as a donor. However, the forensic expert misinterpreted the DNA evidence. The paper demonstrates that the different typing results produce an evidential phenomena that is called 'directional change'. A directional change, as the name already indicates, describes an interaction among items of evidence, where the items support a different hypothesis when evaluated jointly or in isolation. Unfortunately, even today almost nothing is known in the forensic discipline on how to combine items of DNA evidence, let alone about the detection of directional changes. This poses a non-negligible risk of misinterpretations of forensic evidence, which may contribute wrongful convictions in court. The authors show how Bayesian networks can be used to combine items of DNA evidence and how to detect directional changes in the evidence.

Keywords. Bayesian networks · DNA · combining evidence · scenarios · directional change

5.1 Introduction

The present paper aggregates multiple items of DNA evidence in a single Bayesian network (BN) and presents the inferences produced by the BN. The DNA evidence stems from a real case described and examined by W.C. Thompson in [139]. The case in question is that of *State of Texas vs Josiah Sutton*. It involved a sexual assault, and led the wrongful conviction of a person named J. Sutton. The conclusions reached by W.C. Thompson are compared to the conclusions produced by the present model in order to validate the approach taken, and the BN thereby produced.

The propositions of interest are formulated on the so-called 'crime-level' following the hierarchy of propositions established by R. Cook and his fellow researchers in 1998 [23]. Crime-level propositions or hypotheses assign a crime to an authorship. Common formulations of such hypotheses are, for example, 'Mr. X is the assailant of the crime' and 'An unknown person is the assailant'. The hypotheses used in this case correspond to the positions put forth by the defense and the prosecution in the court.

To the best knowledge of the authors, no publication exists on aggregation of multiple DNA typing results under crime-level hypotheses. The interpretation of combined evidence is generally scarce in forensic literature, although, items of evidence occur concurrently most of the time. The authors found that a final reasoning structure for the BN could not be established from the start. In fact, the number of variables to consider turned out to be immense. Thus, in a first step, the authors pursued the reduction of the number of variables. Such a reduction was accomplished by a concise formulation of the problem at hand. Namely, what is the relevance of each specimen?, what number of contributors should be considered for the typing result from a given specimen?, and what are the prerequisites for the existence of the specimens at the crime scene? A separate BN was produced for each of these questions. They served as cornerstones for the reasoning structure of the final BN tackling the main goal of the study: the interpretation of multiple DNA typing results given crime-level propositions. The number of contributors was assessed with the help of a BN proposed in [12]. Meanwhile, observational errors regarding the detection of alleles in the crime profile were included at every step. The consideration of observational errors in probabilistic models is well known, although rarely applied in forensic practice. For a general account on the assessment of observational errors see [125], for the effect of observational errors on DNA evidence see [140], and for its impact in diagnostic processes see [66]. For the evaluation of the mixture profiles, the authors modified a BN first proposed in [104], and combined the latter BN with an original BN that governs well defined contributor scenarios. The BN itself was constructed in Hugin Researcher 8.0.[1] All the computations and analyses were realized in R [25] in conjunction with the RHugin package [85]. The RHugin package serves as an interface between the Hugin Decision Engine (HDE) and the software environment R.

5.1.1 Case description

The case is portrayed following the description given by W. C. Thompson in [139]. The crime took place at the parking lot of the victim's apartment on October 25, 1998. The victim was forced into her car at gunpoint by two men and abducted. The car, a Ford Expedition, has three rows of seats. The victim was forced to perform oral copulation and sexual intercourse on the middle row seats. After she was set free by the assailants, she immediately reported the crime to the police and gave a description of the offenders. The victim stated further that her last sexual contact occurred six days before the crime. The HPDLC Serology/DNA Unit analyzed vaginal swabs and pubic hair combings from a rape kit, a stain found on the victim's jeans, a stain detected on the middle row seat of the Ford Expedition. Acid Phosphatase and P-30 tests on all of theses samples returned a positive result, which is indicative for the presence of sperm. A differential extraction was applied to the material from the vaginal swap. This produced two fractions: the vaginal sperm fraction (containing the male DNA) and the vaginal epithelial fraction (containing the female DNA). The DNA profiles were established using two commercial kits: the PM/DQ-alpha and D1S80 test kits, which were marketed by Perkin Elmer and popular at the time.

Five days after the incident, the victim returned to her apartment. On the way she saw three men on the street, whom she believed to recognize as her assailants. The victim reported her sighting to the police, who arrested the three men thereafter. The three men were placed in the backseat of a patrol car and taken to a parking lot by the police officers. At the parking lot, the victim identified two of the three men as her assailants, still sitting in the patrol car and wearing their hats.

The scientific expert concluded in court that the DNA typing results were DNA mixtures consistent with Sutton, the victim, and at least one other unknown donor for the vaginal swabs, the pubic hair combings, the jeans sample, and the carseat sample. The expert indicated, further, that the DNA profile of J. Sutton had a relative frequency of 1 in 694'000 among the black population. The prosecution dropped all charges against Adams. Sutton, on the other hand, was convicted.

5.1.2 DNA evidence

The DNA typing results and reference profiles are shown in Table 5.1. A box $\boxed{T}$ indicates that the corresponding allele was observed (observation of a given allele is *true*). In turn, a box $\boxed{F}$ indicates the non-observation of an allele (i.e., observation of a given allele is *false*). In the BNs the alleles are named by capital letters of the alphabet in order to have a unified denomination of alleles. For instance, allele 1.1 of marker DQA (as a short for

[1] www.hugin.com

HLA-DQA1) is named A, allele 1.2 is named B, and so on[2]. The relative allelic frequencies were published in [109]. They correspond to the Afro-American subpopulation. Note that the allelic frequencies for the marker DAQ do not sum to 1 for unknown reasons. However, these values are automatically normalized when the BN is compiled. Hence, this incoherence does not constitute a problem for the envisaged evaluations.

Each of the four specimens represents an item of evidence E. Accordingly, the specimen taken from the sperm fraction[3] of the vaginal swab denoted as E_1, the one from debris of the pubic hair combings as E_2, the one from the stain on the victim's the jeans as E_3, and the specimen extracted from a stain on the middle-row carseat as E_4. The typing results for E_1 and E_2 are identical. However, for reasons unknown to the authors, the locus D1S80 was not examined for the specimens E_2 to E_4. In contrast to W.C. Thompson, who had the opportunity to inspect the original the typing results, the HPDLC Serology/DNA Unit was unable to observe allele 3 of DQA (i.e., allele D) in specimen E_3 [138]. The authors examine the DNA results from the perspective that the allele 3 of DQA was observed in E_4. It seems, therefore, reasonable to take into account observational errors regarding the detection of alleles in the crime profiles.

5.1.3 Pre-assessment of the DNA evidence

The forensic expert stated in court that the DNA types of E_1, E_2, E_3, and E_4 were 'consistent' with a DNA mixture of Sutton, the victim, and at least one other unknown donor. If that were true, then one would expect to observe a similar or identical allelic configuration across all specimens. However, this applies only for E_1 and E_2. The configurations in E_3 and E_4 are quite different. Moreover, the question of how a DNA mixture 'consistent' with the DNA profiles of some persons pertains to the identification of assailants is not addressed. Other inconsistencies are found, when the profiles of Adams, Sutton, and the victim are compared to the DNA results of the specimens. Consider the following assessments on the DNA contribution by Adams, Sutton, and the victim:

Adams. Adams appears to possess several alleles that cannot be observed in any of the crime profiles, namely 1.2 DQA, B HBGG, A Gc, and 34 D1S80. Unless the analysis failed to detect these four alleles, it is not possible that Adams contributed to any of the specimens. Even by accounting for observational errors, one expects the probability that Adams is an assailant to be very small given the DNA evidence. In any case, it seems very unlikely[4] for Adams to be a DNA contributor to any of the specimens.

Sutton. All the alleles that Sutton seems to possess can be found in the crime profiles of E_1 and E_2. In contrast, allele 1.1 DQA cannot be observed in E_3 and E_4, although it is found in Sutton's profile. Conversely, the alleles B GYPA and C HBGG were found in E_3 and E_4 but not in Sutton's profile. In E_3 one can further observe allele B D7S8, which is also absent in Sutton's profile. Thus, Sutton can be a contributor to E_1 and E_2. However, for E_3 and E_4 it is unlikely for Sutton to be a contributor, unless one (for E_4) or both (for E_3) alleles of marker DQA were missed during the DNA analysis.

Victim. Allele 4.2/4.3 DQA found in the victim's profile, is not present in any of the crime profiles. Thus, unless this allele was missed during the analysis, the victim cannot be a contributor to any of the specimens. The remaining alleles possessed by the victim are also present in the crime profiles.

It appears that Sutton is unlikely to be a contributor to specimens E_3 and E_4. Thus, E_3 and E_4 must come from at least one other unknown assailant. At this point W.C. Thompson made the following crucial observation: All the

[2] The words 'marker' and 'loci' are used interchangeably.

[3] As expected, the vaginal epithelial fraction produced results that correspond to the victim's profile. Thus, in the present analysis only the vaginal sperm fraction is examined.

[4] Throughout this paper, the words 'probably', 'likely', 'improbable', and 'unlikely' are understood as in the everyday use of these terms (i.e., non-statistical use). Moreover, the terms 'likely' and 'probable' (also 'unlikely' and 'improbable') are used interchangeably.

Table. 5.1 – Allele frequencies and findings of the DNA typing results

Marker	Allele (Name in BN)	Frequency	Specimens				References		
			E_1	E_2	E_3	E_4	Sutton	Adams	Victim
HLA-DQA1	1.1 (A)	0.125	T	T	F	F	T	F	F
	1.2 (B)	0.330	F	F	F	F	F	T	F
	2 (C)	0.130	T	T	F	T	T	F	F
	3 (D)	0.090	T	T	T	T	F	F	T
	4.1 (E)	0.185	T	T	F	F	F	F	F
	4.2/4.3 (F)	0.083	F	F	F	F	F	F	T
	other (other)	0.058	F	F	F	F	F	F	F
LDLR	A	0.194	T	T	T	T	T	T	F
	B	0.806	T	T	T	T	T	T	T
GYPA	A	0.476	T	T	T	T	T	T	T
	B	0.524	T	T	T	T	F	F	T
HBGG	A	0.429	T	T	T	T	T	F	T
	B	0.238	F	F	F	F	F	T	F
	C	0.333	T	T	T	T	F	T	T
D7S8	A	0.663	T	T	T	T	T	T	T
	B	0.337	T	T	T	F	F	F	F
Gc	A	0.087	F	F	F	F	F	T	F
	B	0.750	T	T	T	T	T	T	T
	C	0.163	F	F	F	F	F	F	F
D1S80	20 (A)	0.034	T	-	-	-	F	F	F
	21 (B)	0.135	T	-	-	-	F	F	T
	24 (C)	0.199	T	-	-	-	F	T	F
	25 (D)	0.057	T	-	-	-	T	F	F
	28 (E)	0.148	T	-	-	-	T	F	T
	34 (F)	0.220	F	-	-	-	F	T	F
	other (other)	0.207	F	-	-	-	F	F	F

alleles of E_3 and E_4 are included in the allelic configurations of E_1 and E_2. If, therefore, E_3 or E_4 are connected to the rape, then it must contain the DNA of at least one rapist (under the exclusion of the victim's and Sutton's DNA, which do not correspond). In this case, the combination of alleles from E_3 or E_4, together with the alleles of the victim and Sutton, should produce profiles corresponding to E_1 and E_2. This, however, is not the case since allele 4.1 DQA in E_1 and E_2 cannot be accounted for. In short, when all the specimens are considered together then it is unlikely for Sutton to be a contributor. As can be seen, W.C. Thompson's reasoning does not merely involve considerations on possible contributors. The reasoning extends to the question of how the specimens are related to each other.

There is, however, another problem. Consider E_3 and E_4. It is very likely that both specimens are connected to the crime. However, the profiles of E_3 and E_4 are not the same. Moreover, their profiles do not amount to the allelic configuration of E_1 and E_2 even if considered together with the alleles of the victim. There are two explanations, one of which seems much more likely than the other. The first and more likely explanation is that E_3 and E_4 stem from the same person, and that the difference is due to observational errors (e.g., allele 2 DQA missed in E_3 and allele B D7S8 missed in E_4). The second, much less likely, explanation is that either E_3 or E_4 or both are not connected to the crime in question. Based on the evaluation in Sections 5.2 and 5.3 the second explanation is rejected and not considered in the BN. This implies that the DNA evidence can only be analyzed on an observational level in the BN. Otherwise, one would enter inconsistent evidence into the model.

The pre-assessment shows that the simultaneous examination of multiple typing results is not a trivial task.

5.1.4 Hypotheses of interest

The victim stated that she was assaulted by two persons. An hypothesis must, therefore, involve two persons. The investigation suggested the two suspects, Sutton and Adams, as the assailants. If, however, Sutton and Adams are not the assailants, then two other persons must be the assailants. Moreover, it is equally reasonable to entertain the possibility that either Sutton or Adams could have committed the crime along with another person. This implies that the hypotheses involve four individuals: Sutton, Adams, and two unknown persons.

Let person P_1 denote Sutton, P_2 Adams, and P_3 and P_4 each unknown assailants. On that basis, four hypotheses $H = \{h_1, h_2, h_3, h_4\}$ are formulated, where each hypothesis identifies a pair among P_1, P_2, P_3, and P_4 as the assailants. Table 5.2 presents each hypothesis as one of four possible assailant pairings.

The two hypotheses, h_1 and h_2, identify P_1 as an assailant. In turn, the hypotheses h_3 and h_4 identify persons different from Sutton. The question of whether P_1 is an assailant or not, is formulated by the two composite hypotheses $h_p = \{h_1, h_2\}$ and $h_d = \{h_3, h_4\}$.

Table. 5.2 – Crime-level hypotheses for the specimens

Assailant pairs	H	Hypotheses involving Sutton
$\{P_1, P_2\}$	h_1	h_p
$\{P_1, P_3\}$	h_2	
$\{P_2, P_3\}$	h_3	h_d
$\{P_3, P_4\}$	h_4	

5.1.5 Overview of the paper

Sections 5.2 and 5.3 establish the logical framework for the aggregation of the DNA evidence under the hypotheses of H. Section 5.2 deals with the construction of the most likely worlds, in which the specimens could have been created. The number of contributors is discussed in Section 5.3. Based on the hypotheses, the number of contributors, and the worlds, the different contributor scenarios are established and a corresponding BN is created. This is discussed in Section 5.4. The BN for the typing results themselves is considered in Section 5.5. The BN for the typing results and the BN for the contributor scenarios are assembled to form the final BN. Section 5.6 presents the results obtained from computations based on the final BN. Finally, Section 5.7 outlines the main findings and insights gained from the research, and concludes the paper.

5.2 Creating worlds for the specimens

The creation of worlds centers around the question of how the stains (from which the specimens were extracted) came into existence. This involves the question about the relevance of the specimens for the crime (Subsection 5.2.1) and the possible involvement of another person unrelated to the crime (Subsections 5.2.2). A given world is based on the consideration of these two questions (Subsection 5.2.3). Subsection 5.2.4 refines the relevance considerations assimilated so far. The probability of each world is then computed by using a BN designed for that purpose (Subsection 5.2.5).

5.2.1 Assessing the relevance of the specimens

A trace or mark is usually considered to be left either by the assailant or by some other person, who was not involved in the crime. In such situations, these two possibilities are mutually exclusive and exhaustive. However, biological fluids (and the DNA it contains) can mix so that some part of a trace can come from an assailant and the remaining part from the other person. In such cases, the two aforementioned possibilities are not mutually exclusive and exhaustive. The possibility that a trace was left by the assailant is, therefore, evaluated separately from the possibility that it was left by another person.

Let $G_i = g_i$ denote the fact that the assailants left the trace from which specimen $E_i, i \in \{1,2,3,4\}$ was taken and $G_i = \overline{g_i}$ that the assailant(s) left no such trace. If g_i applies, then the specimen is relevant for the crime [130]. Conversely, let $O_i = o_i$ denote the fact that a person other than the assailants left the trace from which E_i was extracted, and $O_i = \overline{o_i}$ that no other person left such a trace. Clearly, if a specimen was left by some other person, then it has no inferential value regarding the authorship of the rape. In such cases, the specimen is said to be *irrelevant* for the crime in question. Note that this 'other person' is the unknown contributor but not an assailant.

Probability of relevance $Pr(g_i)$

Consider the probability of g_i of specimen $E_i, \in \{1,2,3,4\}$. In order for an assailant to have left a trace of sperm it is necessary that (a) the assailant ejaculated, and (b) that he did not use a condom.[5] Points (a) and (b) are assessed by using the criminological study of O'Neal et al. from 2013 [107]. The relevant pieces of information from this study are denoted by I_1 and I_2.

> I_1: The study suggests that the prevalence rate for the use of condoms during sexual assaults ranges from 11.7% to 15.6%. This study also indicates that younger suspects, and suspects who used a weapon during the rape are more likely to use a condom. In the present case, the suspects were young and used a weapon during the rape.
>
> I_2: The same study suggests that suspects who used a condom ejaculated in 83.3% to 90.5% of the examined assault cases. In the present case it is assumed that the rate of ejaculation is comparable to cases, in which condoms are not used.

[5] The possibility of azoospermia is not taken into account. In fact, azoospermia is a rare condition especially among young adults, since they are very unlikely to have undergone a vasectomy. The probability that the assailants were azoospermic at the time of the rape is, thus, considered negligible.

Let $C_1 = \{c_1, \overline{c_1}\}$ denote the use of condom by the first assailant during the assault, which can be true (c_1) or false ($\overline{c_1}$). Let $C_2 = \{c_2, \overline{c_2}\}$ denote the same but for the second assailant. Let $J_1 = \{j_1, \overline{j_1}\}$ refer to the first assailant ejaculating during the assault, which can be true (j_1) or false ($\overline{j_1}$). Again $J_2 = \{j_2, \overline{j_2}\}$ refers to the same but for the second assailant.

The variables J_1, and J_2 are considered to be independent given I_2 (i.e., $Pr(j_1, j_2 \mid I_2) = Pr(j_1 \mid I_2)Pr(j_2 \mid I_2)$). The same applies to C_i and J_i given I_1 and I_2 (i.e., $Pr(C_i, J_i \mid I_1, I_2) = Pr(C_i \mid I_1)Pr(J_i \mid I_2)$). However, C_1 and C_2 are not considered independent because the assailants are more likely share the same mindset, rather than disparate ones. Notably, if one assailant uses a condom, then the other is more likely to use one as well. Conversely, if one assailant does not use a condom, then the other assailant is less likely to use a condom. Hence, C_2 is conditioned by C_2 so that $Pr(c_2 \mid c_1, I_1) > Pr(c_2 \mid \overline{c_1}, I_1)$.

To the probability that an assailant used a condom during the assault given information I_1 the probability assignments $Pr(c_1 \mid I_1) = 0.16$, $Pr(c_2 \mid c_1, I_1) = 0.8$, and $Pr(c_1 \mid \overline{c_1}, I_1) = 0.12$ are made. To the probability that the assailant ejaculated during the assault, given information I_2 the value $Pr(j_1 \mid I_2) = Pr(j_2 \mid I_2) = 0.833$ is assigned.

For a given specimen to be relevant, it is necessary that at least one assailant ejaculated and did not wear a condom during the assault. In other words, the probability that a stain (from which the specimen E_i was extracted) was left by one or both assailants is given by the sum of the probabilities of each combination of C_1, C_2, J_1, and J_2 that can produce a semen stain. In other words, one has

$$Pr(g_i \mid I_1, I_2) = \sum_{C_1}\sum_{J_1}\sum_{C_2}\sum_{J_2} Pr(g_i \mid C_1, J_1, C_2, J_2, I_1, I_2)Pr(C_1, J_1, C_2, J_2 \mid I_1, I_2). \tag{5.1}$$

BN for computing the probability of g_i

Throughout this paper node names and states are written in teletype font. Query nodes are colored in grey and their node names written in white. Evidence nodes are colored in light grey while their node names are written in black. All the nodes are Boolean, unless indicated otherwise.

The probability of g_i of specimen $E_i, , \in \{1,2,3,4\}$ is computed with the BN in Figure 5.1. The probabilities of j_i are assigned to the `true`-states of the root nodes[6] with the corresponding names. Conversely, the probabilities of $\overline{c_i}$ are assigned to the `false`-states of the corresponding nodes. As discussed previously, an assailant can only produce a stain if he ejaculated, *and* did not wear a condom during the assault. Hence, the nodes `J1C1` and `J2C2` are defined as logical conjunctions of `J1` and `C1`, and `J2` and `C2` respectively (e.g., `J1C1 = J1` $\wedge$ `C1`). Finally, a relevant stain exists, if one *or* both of the assailants produced a stain. Node `G` is, therefore, defined as a logical disjunction of its parent nodes (i.e., `G = J1C1` $\vee$ `J2C2`). The probability of g_i is retrieved from node `G` after compiling the network. The value $Pr(g_i \mid I_1, I_2) \approx 0.83$ is retained from the BN.

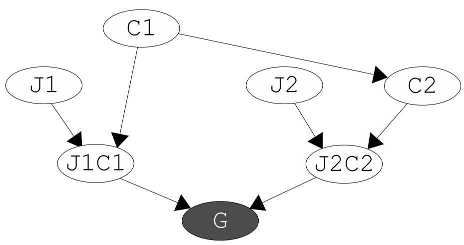


Figure. 5.1 – BN for computing the probability $Pr(g_i \mid I_1, I_2)$ according to Equation 5.1. Once the BN is compiled, the value for $Pr(g_i \mid I_1, I_2)$ can be retrieved from the state `true` of node `G`.

[6]A node without an ancestor is called 'root node'.

5.2.2 Probability that another person unrelated to the crime left the trace

Next, consider the probability of o_i, that is, a person unrelated to the crime left the trace from which the specimen $E_i, \in \{1,2,3,4\}$ was taken. Consider the two pieces of information derived from common sense and from the case description:

> I_3: It is extremely unlikely to find semen stains on a female's body, clothing, or in her car, unless, the female herself was engaged in a sexual activity producing male ejaculate.

> I_4: An event unrelated to the rape must have occurred before or after the rape itself, but not simultaneously to the rape. The victim stated not to have engaged in any sexual activities six days before the rape, and to have sought out to the police immediately after the rape.

For specimen E_3 and E_4 this probability is judged to be extremely small. Therefore, given I_3 and I_4 the following probability values are assigned: $Pr(o_3 \mid I_3, I_4) = Pr(o_4 \mid I_3, I_4) = 0.001$. In contrast, the value zero is assigned to the probability $Pr(o_2 \mid I_3, I_4)$. It is assumed that the victim washed herself several times during these six days. Under this assumption, it is impossible to detect seminal fluid (and DNA from semen) in the pubic area after six days.

Finally, consider the specimen E_1. The case description provides the following piece of information:

> I_5: All the DNA typing results were produced by the PM/DQ-alpha and D1S80 test kits.

In general, the probability of obtaining a DNA profile decreases rapidly, when sperm is exposed to the vaginal environment. In particular, [102] Mayntz-Press et al. state that:

> I_6: The sperm fraction in samples with extended post-coital interval may contain less than 10 sperm cells. Thus, the quantity of DNA may be below the detection limit of autosomal STR systems. [7]

Given I_5 and I_6 the authors consider it to be virtually impossible that the tests kits used in the present case could detect DNA from sperm six days after its deposition in the victims vagina, that is, $Pr(o_1 \mid I_3, I_4, I_5, I_6) = 0$.

5.2.3 Identifying possible worlds that can generate the specimens

Each specimen stems either from a trace left exclusively by the rapist(s), exclusively by some other person(s), a mixture of both, or none of the above. More formally, one can associate a specimen $E_i, \in \{1,2,3,4\}$ to one of four possible subsets: $\{g_i, \overline{o_i}\}$, $\{\overline{g_i}, o_i\}$, $\{g_i, o_i\}$, or $\{\overline{g_i}, \overline{o_i}\}$. For four specimens this means, that there are a total of $n = 4^4 = 256$ combinations of subsets. A given combination is called a 'world' w_k and is, thus, a set of four subsets. Let $W = \{w_1, w_2, ..., w_n\}$ denote the set of possible worlds, and w_1 represent a world in which all the specimens derive from traces exclusively left by the rapist(s). Such a world can be represented by the set $w_1 = \{\{g_1, \overline{o_1}\}, \{g_2, \overline{o_2}\}, \{g_3, \overline{o_3}\}, \{g_4, \overline{o_4}\}\}$. The authors use the notation $w_k \sim E_i$ to represent the subset of w_k associated with E_i. Thus, $w_1 \sim E_1$, for example, corresponds to $\{g_1, \overline{o_1}\}$. The temporal dimension is not addressed here. For example, the interest is not whether g_i occurred before or after o_i, when $\{g_i, o_i\}$ applies, but only that both occurred during the relevant timespan. Hence, the subsets as well as the worlds themselves are represented as sets and not tuples.[8]

As can be seen, the number of worlds increases exponentially with the number of specimens. Thus, worlds that are found to be impossible in the present case are excluded from consideration, before establishing an exhaustive

[7] In rare cases a partial DNA profile from sperm can be detected after a post-coital interval of up to seven days [102]. However, the methods applied in this study involved more modern Y-STR systems and more sophisticated sample treatments (i.e., post-PCR purification). In an early study from 2003 that also analyzed Y-STR markers, the DNA from the sperm donor could not be detected after 72 hours [70].

[8] It can readily be seen, that addressing the temporal dimension leads to an exponential increase in the complexity of the reasoning problem. For instance, let the double (g_i, o_i) refer to the sequence 'g_i before o_i' and (o_i, g_i) to the sequence 'o_i before g_i'. Note that g_i and o_i cannot occur simultaneously given that I_4 applies. As a result, one has: $\{g_i, o_i\} \leftrightarrow (g_i, o_i) \veebar (o_i, g_i)$, where $\veebar$ denotes an exclusive or. Hence, by including the temporal consideration one has doubled the possible relationships between g_i and o_i to examine (and tripled in the absence of I_4).

list of possible worlds. The reduction of the number of worlds is accomplished by three means. First, two further pieces of information are invoked. Second, the consequence of a probability assignment is exploited. Third, an item of evidence is included in the consideration of possible worlds.

Reducing the number of possible worlds by using circumstantial information

Consider the pieces of information denoted by I_7 and I_8:

> I_7: All the specimens tested positive for the p30 and the PSA tests.
>
> I_8: All the specimens were retrieved from a small, coherent geographical area, namely, from the victim's genital area, her jeans, and the car's middle row seat (i.e., where the assault took place).

Information I_7 suggests that all the specimens contain seminal fluid rather than some other substance testing positive. This in turn, implies that there is a coherence in substance among the four specimens. Similarly, I_8 suggests a strong coherence in space. A joint consideration of these two pieces of information leads to the conclusion that all these specimens are more likely to be the consequence of a single event (such as the assault in question), rather than two or more events. For E_1 and E_2, it seems even reasonable to assume their unity in substance and space. Stated otherwise, it appears inconceivable for these two specimens to be associated with different subsets of G_i and O_i (for instance, $\{g_1, \overline{o_1}\}$ and $\{\overline{g_2}, o_2\}$). As a consequence, these specimens are better described as a single specimen. Let, therefore, $E_{1'} = \{E_1, E_2\}$ denote the specimen, under which E_1 and E_2 are subsumed. The indexes for the other specimens remain unchanged. The number of possible worlds is reduced to $n = 4^3 = 64$.

Reducing the number of possible worlds by examining the consequence of a probability assessment

In the previous section (Section 5.2.2), it was judged that $Pr(o_1 \mid I_3, I_4, I_5, I_6) = Pr(o_2 \mid I_3, I_4) = 0$. Hence, by including the considerations of the previous paragraph one has $Pr(o_{1'} \mid I_3, I_4, I_5, I_6, I_7, I_8) = 0$. This, in turn, implies that the propositions $g_{1'}$ and $\overline{g_{1'}}$ become exhaustive. Thus, for any given world, specimen $E_{1'}$ is described either by the subset $\{g_{1'}\}$ or $\{\overline{g_{1'}}\}$ (instead of $\{g_{1'}, \overline{o_{1'}}\}$, $\{\overline{g_{1'}}, o_{1'}\}$, $\{g_{1'}, o_{1'}\}$, or $\{\overline{g_{1'}}, \overline{o_{1'}}\}$). This reduces the number of possible worlds by a half, notably, $n = 2 \times 4^2 = 32$.

Reducing the number of possible worlds by introducing an item of evidence

Consider the fact that each specimen yielded a DNA profile. This fact is not a piece of circumstantial information, but part of the results from the DNA analysis. It is, therefore, an item of evidence. This item enables the exclusion of all the worlds containing $\{\overline{g_{1'}}\}$, $\{\overline{g_3}, \overline{o_3}\}$, or $\{\overline{g_3}, \overline{o_3}\}$. The number of possible worlds now is $n = 1 \times 3^2 = 9$.

5.2.4 Refining considerations on G_3 and G_4

Section 5.2.1 concluded that $Pr(g_i \mid I_1, I_2)$ expresses the probability that at least one of the two assailants ejaculated and did not use a condom during the assault. As discussed in Section 5.2.3, the pieces of circumstantial information I_7 and I_8 suggest a strong coherence in substance and space among all the specimens. In consideration of these facts, the most likely course of events is that one assailant produced a sperm stain during the assault, and that this stain was subsequently scattered across the crime scene creating thereby the remaining stains[9]. A reasoning about G_i

[9] Such a generation of evidence by scattering of an object is also described by the fundamental forensic principle of divisible matter: 'Matter divides into smaller component parts when sufficient force is applied. The component parts will acquire characteristics created by the process of division itself and retain physico-chemical properties of the larger piece' [73, p.12]. In the present case, the chemical properties are the DNA and other materials contained in semen.

must address this course of events, but also allow for the less likely possibility that a stain of sperm was generated (i.e., j_i and $\overline{c_i}$) more than once during the assault.

Next consider the geographical pattern of how the stains were scattered across the crime scene. Recall that specimen $E_{1'}$ stems from the victim's genital area, E_3 from the victims jeans, and E_4 from the carseat where the assault took place. It seems safe to assume that the victim's genital area coincides with the location at which a stain was generated. This leads to the possibility that this stain dispersed from the genital area to the remaining locations by subsequent actions of the assailants or the victim herself. Conversely, the course of events that the stain was initially produced at the extraction point of either E_3, E_4, or some other place, and came to be located afterwards at the victim's genital area (including the victim's vagina) is not convincing. Such a course of events is not taken into account. Thus, the stains located at the extraction points of specimens E_3 and E_4 may be either derived from a stain that stems from the victims genital area, or from a stain generated somewhere else. That is, they may have dropped to their corresponding extraction points from an assailant's penis, or from the victim's mouth after her having been forced to perform oral copulation. Finally, the stain associated with E_3 may come from the stain associated with E_4, and vice versa. Now, one must assess how likely each of these possible courses of events are. The case description does not provide any further information on that matter. However, general literature provides this pieces of information [24]:

> I_9: It is very common in rape cases that sperm stains are found at places such as the crime scene or on the victim's clothing aside from the victim's natural orifices.

Further, one can invoke common sense:

> I_{10}: In general, it is more likely for two stains to emerge, if one stain already exists; it is more likely for three stains to emerge, if one stain already exists; and that is even more likely, if two stains exist; and so on...

In order to accomplish a reasoning pattern that embraces all these possibilities, G_3 and G_4 are conditioned on $G_{1'}$, and G_4 additionally on G_3. The probabilities $Pr(g_{1'} \mid I_1, I_2, I_7, I_8)$, $Pr(g_3 \mid \overline{g_{1'}}, I_1, I_2, I_7, I_8, I_{10})$ and $Pr(g_4 \mid \overline{g_{1'}}, \overline{g_3}, I_1, I_2, I_7, I_8, I_{10})$ are then given by Equation 5.1. Further, one has the relationships $Pr(g_3 \mid g_{1'}, I_1, I_2, I_7, I_8, I_{10}) > Pr(g_3 \mid \overline{g_{1'}}, I_1, I_2, I_7, I_8, I_{10})$ and $Pr(g_4 \mid g_{1'}, g_3, I_1, I_2, I_7, I_8, I_{10}) > Pr(g_4 \mid \overline{g_{1'}}, g_3, I_1, I_2, I_7, I_8, I_{10}) \approx Pr(g_4 \mid g_{1'}, \overline{g_3}, I_1, I_2, I_7, I_8, I_{10}) > Pr(g_4 \mid \overline{g_{1'}}, \overline{g_3}, I_1, I_2, I_7, I_8, I_{10})$. Hence, the lower bound is defined by Equation 5.1 for the probability of any given $g_i, \in \{1', 2, 4\}$.

The probability of a world is, therefore, conditioned by $G_{1'}$ to G_4, O_3 and O_4, and all the pieces of information these variables invoke. From now on, the entire body of information is denoted by the set $\mathbf{I} = \{I_1, I_2, ..., I_n\}$ (at this point one has $n = 10$). Any further information used later is added to $\mathbf{I}$. The probability of each world is computed by using the BN presented in the following section. The probability values discussed are outlined in Table 5.3.

5.2.5 Bayesian network for computing the probability of each world

Figure 5.2 depicts the BN established for computing the probabilities of the different worlds. Note that node `G1` represents variable $G_{1'}$. The probabilities of Table 5.3 are assigned to the `true`-states of the nodes `G1`, `G3`, `G4`, `O3`, and `O4`. The nodes `GO3` and `GO4` are numbered. They possess the states 1, 2, 3, 4. State 1 corresponds to the subset $\{g_i, \overline{o_i}\}$, state 2 to $\{g_i, \overline{o_i}\}$, state 3 to $\{\overline{g_i}, o_i\}$, and state 4 to $\{\overline{g_i}, \overline{o_i}\}$, where $i = \{3, 4\}$. Therefore, the probability value of 1 is assigned to a given state, if it is in correspondence with the state configuration given by its parental nodes, and 0 otherwise. The design of node `W` (worlds) is guided by an analogous reasoning. It is numbered and possesses ten states 1, 2, 3, ..., 9, 99. Each state from 1 to 9 corresponds to one of the nine worlds (see Section 5.2.3). The probability value 1 is given to a state, if it corresponds to the state configurations of its parental nodes, and 0 otherwise. All the other worlds are pooled in state 99, which receives the probability value 1, unless, the parental configuration corresponds to one of the nine worlds. In the latter case, the value 0 is assigned. Node `R1` sets state 99 of `W` to zero. This operation rescales the probabilities exclusively over the worlds 1 to 9. Node `R1` is defined

Table. 5.3 – Probability values assigned to each event $g_{1'}, g_3, g_4, o_3$, and o_4 given the corresponding pieces of information, and the possible configurations of conditioning events

Event	Conditioned by	Information	$Pr(\cdot)$
$g_{1'}$		I_1, I_2, I_7, I_8	0.83
g_3	$g_{1'}$	$I_1, I_2, I_7, I_8, I_{10}$	0.90
	$\overline{g_{1'}}$		0.83
g_4	$g_{1'}, g_3$	$I_1, I_2, I_7, I_8, I_{10}$	0.92
	$g_{1'}, \overline{g_3}$		0.90
	$\overline{g_{1'}}, g_3$		0.90
	$\overline{g_{1'}}, \overline{g_3}$		0.83
o_3		I_3, I_4	0.001
o_4		I_3, I_4	0.001

as a negation of state 99, that is, `W = ¬99`, and serves to enter the evidence that sperm stains were found at each extraction point. Node `R2` rescales the probabilities over an arbitrary number of chosen worlds. It is defined as a logical disjunction of the worlds, over which the probabilities are to be rescaled. In the present case one has `W=1` ∨ `W=2` ∨ `W=3`. Its use is clarified shortly. In the meantime, consider the probabilities obtained for each world in the middle column of Table 5.4. As can be seen, the world w_1 (where each specimen is directly linked to an assailant) is by far the most probable world. The next probable worlds are w_2 and w_3, where either specimen E_3 or E_4 is associated with a mixed stain from an assailant, and from some other person not involved in the crime. In the following examination of the DNA profiles, only the worlds w_1 to w_3 are considered. In order to do so, the probabilities of the worlds are rescaled over w_1, w_2 and w_3. This is accomplished by instantiating `R2 = true`. The rescaled probabilities are listed in the outer right column.

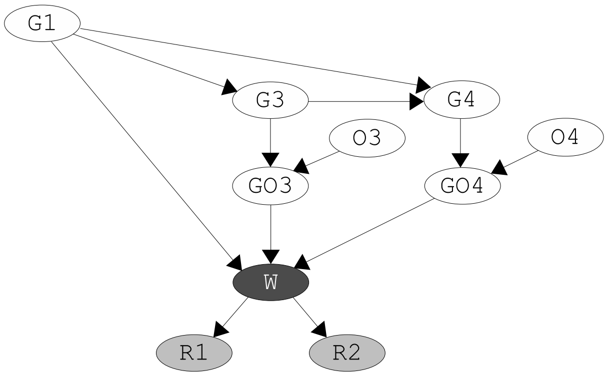


Figure. 5.2 – BN for computing the probability of each possible world

Table. 5.4 – Marginal probabilities of each world w_k. Event e represents the evidence that sperm stains were found at the extraction points of each specimen.

W	$Pr(w_k \mid e,\mathbf{I})$	$Pr(w_k \mid k<4,e,\mathbf{I})$
$w_1 = \{\{g_{1'}\},\{g_3,\overline{o_3}\},\{g_4,\overline{o_4}\}\}$	0.99781	0.998
$w_2 = \{\{g_{1'}\},\{g_3,\overline{o_3}\},\{g_4,o_4\}\}$	0.00100	0.001
$w_3 = \{\{g_{1'}\},\{g_3,o_3\},\{g_4,\overline{o_4}\}\}$	0.00100	0.001
$w_4 = \{\{g_{1'}\},\{\overline{g_3},o_3\},\{g_4,\overline{o_4}\}\}$	0.00011	
$w_5 = \{\{g_{1'}\},\{g_3,\overline{o_3}\},\{\overline{g_4},o_4\}\}$	0.00009	
$w_6 = \{\{g_{1'}\},\{g_4,o_4\},\{g_4,o_4\}\}$	0.9998e-6	
$w_7 = \{\{g_{1'}\},\{\overline{g_3},o_3\},\{g_4,o_4\}\}$	0.1087e-6	
$w_8 = \{\{g_{1'}\},\{g_4,o_4\},\{\overline{g_4},o_4\}\}$	0.8694e-7	
$w_9 = \{\{g_{1'}\},\{\overline{g_3},o_3\},\{\overline{g_4},o_4\}\}$	0.1207e-7	

5.3 Assessing the number of contributors for each specimen

The number of DNA contributors in a crime stain can rarely be known with certainty. Moreover, the inferential force that a profile generates on these hypotheses may vary greatly depending on the number of contributors considered in the hypotheses. Methods that apply probability theory to address this problem have been known for quite some time in forensic science [17, 91, e.g.]. The present study employs the BN proposed in [12]. This BN is tailored for inferring the number of contributors in a DNA profile, and is presented later in Subsection 5.3.2. This BN assesses the number of contributors in terms of probabilities on the basis of the alleles observed in a DNA specimen, and the circumstantial information available. An inference executed by the model is discussed in Subsection 5.3.1. The results obtained for the present case are presented in Subsection 5.3.3.

5.3.1 Inferring the number of contributors from DNA typing results

Let $N = \{1,2,...,n\}$ denote the number of contributors and $\mathbf{A}_i$ the set of alleles analyzed in specimen $E_i, i \in \{1',3,4\}$. The underlying inference is given by the Bayes' theorem

$$Pr(N \mid \mathbf{A}_i,\mathbf{I}) = \frac{Pr(\mathbf{A}_i \mid N,\mathbf{I})Pr(N \mid \mathbf{I})}{Pr(\mathbf{A}_i \mid \mathbf{I})},$$

where $Pr(\mathbf{A}_i \mid \mathbf{I}) = \sum_N Pr(\mathbf{A}_i \mid N,\mathbf{I})Pr(N \mid \mathbf{I})$. The probability of interest is $Pr(N \mid \mathbf{A}_i,\mathbf{I})$.

Let P_j denote a possible DNA contributor, and $\mathbf{P} = \{P_1,P_2,...,P_n\}$ the set of DNA contributors considered. In the present case a mixture with a maximum of five contributors per specimen was considered. In other words, one has $N = \{1,2,3,...,5\}$ and, therefore, $P_n = P_5$.

Further let $\mathbf{L} = \{L_1,L_2,...,L_n\}$ denote the set of all the markers analyzed by the kits, and L_l a single marker. In the present case there are $n = 7$ markers, where $L_1,L_2,...,L_7$ represent the markers HLA-DQA, LDLR, ..., D1S80 respectively (see Table 5.1).

A single allele is denoted as $A_{k,i}$, that is, the kth allele examined in the ith specimen. An allele can be present $A_{k,i} = a_{k,i}$ or absent $A_{k,i} = \overline{a_{k,i}}$ in a specimen DNA. The observation of the presence or absence of an allele is denoted as $A^*_{k,i} = \{a^*_{k,i},\overline{a^*_{k,i}}\}$, and the set of all observations as $\mathbf{A}^*_i$. As will be seen later, two distinct observations are envisaged for a single allele for the markers L_1 to L_6 of specimen $E_{1'}$ (i.e., $k \leq 19$). If the observations themselves are to be specified, then $A^*_{1,k,1'}$ and $A^{2*}_{2,k,1'}$ are used to denote the first and the second observation respectively.

The variable $A^*_{o,k,1'}, k \leq 19$ is written when the observation is not intended to be specified. When the variable $A^*_{k,1'}, k \leq 19$ is used in a statement, then this is meant to imply both observations equally (i.e., $A^*_{1,k,1'}$ and $A^{2*}_{2,k,1'}$).

In the present case, the inference is not based on the actual presence or absence of an allele, but the *observation* of its presence or absence. One computes, therefore, the probabilities $Pr(N \mid \mathbf{A}^*_i, \mathbf{I})$ rather than $Pr(N \mid \mathbf{A}_i, \mathbf{I})$.

5.3.2 BN for computing the probability of the number of contributors.

The BN is shown in Figure 5.3 by using the plate notation[10]. The subgraph in the upper plate is replicated for every possible contributor P_j considered (i.e., $\forall P_j \in \mathbf{P}$). The subgraph in the lower plate is replicated for every marker examined (i.e., $\forall L_l \in \mathbf{L}$).

The nodes `PjLlpg` and `PjLlmg` represent the paternal genotype (pg) and maternal genotype (pg) respectively that a person P_j possesses at locus L_l. These nodes are labelled by the letter of the alphabet. Each state represents an allele of the marker considered (i.e., $A_{k,i}$). For example, node `PjL1pg` has seven (A to F) and `PjL2pg` two states (A and B). To each allele state its corresponding allele frequency is assigned. The allele frequencies were listen in Table 5.1.

Node `N` represents the number of persons that contributed to a DNA. The node is numbered, and each state corresponds to the number of contributors. In the present case, there are states numbered from 1 to 5.

The node `Nj` inquires whether person P_j is a DNA contributor to the specimen. Hence, node `Nj` is defined by the function `N` $\geq$ `j`. The nodes `PjA1l` to `PjAnl` refer to whether person P_j possesses an allele $A_{k,i}$ and take the value `true` if `PjLlpg` *or* `PjLlmg` contain the allele $A_{k,i}$, *and* if further `Nj` holds. These nodes are, thus, defined by the logical expression {`PjLlpg` $\vee$ `PjLlmg`} $\wedge$ `Nj`.

The nodes `isA1l` to `isAnl` enquire whether the allele $A_{k,i}$ is contained in the DNA of the specimen. Hence, a profile possesses $A_{k,i}$, if at least one $P_j \in \mathbf{P}$ possesses this allele. These nodes are, therefore, defined as logical disjunctions of `PjA1l`, that is, $\vee_{\mathbf{P}}$ `PjA1l`.

The nodes `obsA1l` to `obsAnl` enquire whether the allele $A_{k,i}$ contained in the DNA, was actually observed or not ($A^*_{k,i}$). These nodes codify, therefore, observational errors. The values $Pr(\overline{a^*_{k,i}} \mid a_{k,i}, \mathbf{I}) = 0.02$ and $Pr(a^*_{k,i} \mid \overline{a_{k,i}}, \mathbf{I}) = 0.001$ are retained for the probabilities of false negative and false positive respectively (i.e., probabilities for the errors of the first and second kind).[11] Recall that specimen $E_{1'}$ subsumes E_1 and E_2. Even though E_1 and E_2 are considered to be a single specimen, each observation of an allele is still taken into account. In other words, the alleles of the markers L_1 to L_6 were observed twice (i.e., E_1 and E_2), whereas loci L_7 only once (i.e., E_1). In the present BN, this circumstance is translated by creating two observational nodes for each allele of loci L_1 to L_6 when examining specimen $E_{1'}$. For example, the nodes `obsA11_1` and `obsA11_2` are the observational nodes for `isA11`. These two observational nodes correspond to the variables $A^*_{1,1,1'}$ and $A^*_{2,1,1'}$ respectively.

Finally, the node `R` is used to rescale the probabilities of the number of contributors $Pr(N \mid \mathbf{A}^*_i, \mathbf{I})$ to desired maximum and/or minimum number of contributors.

5.3.3 Results for the probabilities of the number of contributors

The prior probabilities of N ($Pr(N \mid \mathbf{I})$) used, and the posterior probabilities ($Pr(N \mid \mathbf{A}_i, \mathbf{I})$) calculated are outlined Table 5.5. Note that different prior probabilities are used for specimen $E_{1'}$, and E_3 and E_4. This is because given I_3, I_4, I_5,and I_6, one does not expect to find DNA from another person not involved in the assault (see Section 5.2.2). Hence, a number of contributors larger than N =3 was considered impossible. For specimens E_3 and E_4 the situation is different. Namely, given that these specimens were taken from stains openly exposed to the environment, the possibility of residual DNA from a person not involved in the assault is judged to be possible but unlikely.

[10] Repetitive elements of a BN are represented as subgraphs in rectangles so called 'plates'. The number of times a subgraph is replicated (including arcs that cross plate edges) is indicated in the corner of each plate [20]

[11] The notation $Pr(\overline{a^*_{k,i}} \mid a_{k,i}, \mathbf{I})$ is used as a shorthand for $Pr(A^*_{k,i} = \overline{a^*_{k,i}} \mid A_{k,i} = a_{k,i}, \mathbf{I})$ throughout this paper.

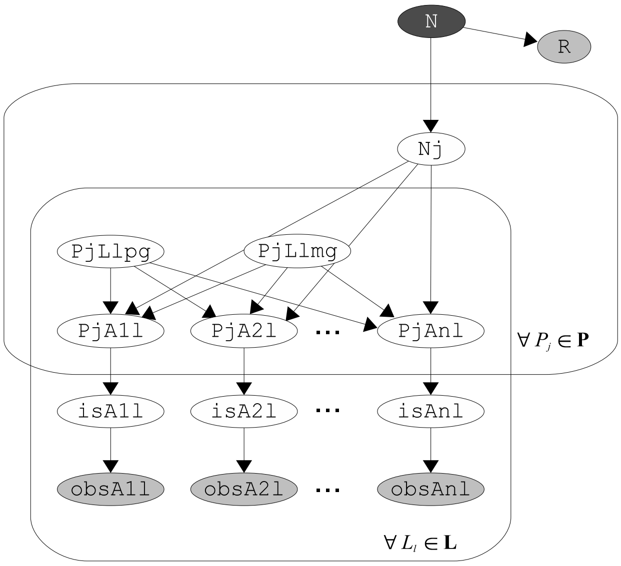


Figure. 5.3 – Generic BN for computing the probability of the number of contributors in a given profile. After the compilation of the BN, the observations are entered in the nodes `obsA1l` to `obsAnl`. The probabilities for the number of contributors are retrieved from node `N`.

For each specimen, the same BN was used for computing the posterior probabilities of N. More precisely, nodes `obsLlA1` were instantiated to `true` if the allele $A_{k,i}$ was observed in specimen E_i, and `false` otherwise. The posterior probabilities were then extracted from node `N`. This process was repeated for each specimen, that is, for each $i \in \{1', 3, 4\}$.

As the resulting posterior values indicate, the most likely number of contributors for specimen $E_{1'}$ is $N = 3$, and for specimens E_3 and E_4 each $N = 1$. For a further evaluation of the profiles only the most likely, and the second most likely number of contributors were retained for each specimen. The probabilities obtained after rescaling are given in Table 5.6.

Table. 5.5 – Prior and posterior probabilities for the number of contributors

N	$Pr(N \mid \mathbf{I})$	$Pr(N \mid \mathbf{A}_{1'}, \mathbf{I})$	$Pr(N \mid \mathbf{I})$	$Pr(N \mid \mathbf{A}_3, \mathbf{I})$	$Pr(N \mid \mathbf{A}_4, \mathbf{I})$
1	0.7	1.682e-15	0.7	9.3992e-1	9.677e-1
2	0.2	2.216e-2	0.2	5.7116e-2	3.1621e-2
3	0.1	9.778e-1	0.09	2.9298e-3	6.7627e-4
4	0.0	0	0.009	3.1186e-5	5.0233e-6
5	0.0	0	0.001	3.9789e-7	5.6392e-8

Table. 5.6 – Posterior probabilities retained after rescaling

N	$Pr(N \mid 1 < N < 4, \mathbf{A}_{1'}, \mathbf{I})$	$Pr(N \mid N < 3, \mathbf{A}_3, \mathbf{I})$	$Pr(N \mid N < 3, \mathbf{A}_4, \mathbf{I})$
1	0	0.94	0.97
2	0.02	0.06	0.03
3	0.98	0	0

5.4 Possible contributor scenarios for aggregating multiple DNA typing results

The variables W and N set the corenerstones for the compilation of possible contributor scenarios under H (see Tables 5.4 and 5.6). The role of these two variables is twofold. On the one hand, they help to identify the possible contributor scenarios. On the other hand, they are directly included in the definition of a possible contributor scenario. In other words, they not only serve as a stepping stone for reasoning about those scenarios, but they also play an active role in the final BN. A general account on the specification of the scenarios is given Subsection 5.4.1. The role played by W and N in the scenarios is portrayed in detail in Subsection 5.4.2. Subsection 5.4.3 addresses the distinction between 'actual contributor' and 'potential contributor', as well as the relationships between the specimens. This subsection also explains how an actual contributor relates to the hypotheses H. Subsection 5.4.4 deals with the derivation of a reasoning pattern that includes all the considerations discussed in the previous subsection. The BN created on the basis of that reasoning pattern is presented in Subsection 5.4.5.

5.4.1 Specifying scenarios

Let N_i refer to the number of contributors for specimen $E_i, i \in \{1', 3, 4\}$. The set of possible contributors considered for specimen E_i is denoted as $\mathbf{P}_i$. Let P_1 to P_4 represent the four potential assailants, P_5 the victim and P_6 the individual not involved in the assault (i.e., the 'other person' from Section 5.2.1). A scenario represents a donor configuration, and corresponds itself to a set of persons P_j's involved in that configuration. One has, for example, $S_{1,1} = \{P_1, P_2\}$ or $S_{2,1} = \{P_1, P_3\}$. Again, the temporal dimension (i.e., the sequence according to which the different persons contributed to a stain) is not taken into account. Note also that for a given number of contributors, and for a given world $w_k \in W$, each specimen can only have four possible scenarios. Remember that there are only two actual assailants and four potential assailants. Moreover, in any given world $w_k \in W$ the assailant(s) sperm is present in a specimen. In other words, (a) a specimen contains the sperm of either both assailants, in which case one has a pair among four potential assailants, and therefore, $n = \binom{4}{2} = 4$ possible scenarios; or (b) a single assailant, in which case one has $n = \binom{4}{1} = 4$ possible scenarios. All the scenarios are listed in Table 5.7 and Table 5.8 for specimens $E_{1'}$, and E_3 and E_4 respectively. The scenarios are explained step by step in the subsequent subsections. The discourse follows the order of the columns in the Tables 5.7 and 5.8 from left to right.

5.4.2 Number of contributors and worlds

For specimen E_1 there is only one variant of subset associated with each world, that is, $\{g_{1'}\}$. Thus, if $N = 2$ applies, then the contributor configurations are given by the combinations of two assailants out of the four possible individuals P_1 to P_4 ($S_{1 \le s \le 4,1}$). If one has $N = 3$, then the possible scenarios are essentially the same, except that this time, the victim P_5 is included in every scenario ($S_{5 \le s \le 8,1}$). Given the fact that $\{g_{1'}\}$ applies, a person unrelated to the assault cannot have contributed to the stain. Thus, person P_6 does not figure in any scenario of specimen E_1. Therefore, for E_1 one has a total of $|\mathbf{S}_1| = 8$ scenarios.

For specimens E_3 and E_4 there are two variants of subsets associated with each world, which are $\{g_i, \overline{o_i}\}, i \in \{3, 4\}$ and $\{g_i, o_i\}, i \in \{3, 4\}$. The difference between E_3 and E_4, is that the variant $\{g_i, o_i\}$ appears in w_3 for E_3, Table. 5.7 – Possible contributor scenarios for $E_{1'}$

$N_{1'}$	$\mathbf{P}_{1'} \setminus \{\cdot\}$	H	$S_{s,1'}$	
2	$\mathbf{P}_{1'} \setminus \{P_3, P_4, P_5\}$	h_1	$\{P_1, P_2\}$	$S_{1 \leq s \leq 4,1}$
	$\mathbf{P}_{1'} \setminus \{P_2, P_4, P_5\}$	h_2	$\{P_1, P_3\}$	
	$\mathbf{P}_{1'} \setminus \{P_1, P_4, P_5\}$	h_3	$\{P_2, P_3\}$	
	$\mathbf{P}_{1'} \setminus \{P_1, P_2, P_5\}$	h_3	$\{P_3, P_4\}$	
3	$\mathbf{P}_{1'} \setminus \{P_3, P_4\}$	h_1	$\{P_1, P_2, P_5\}$	$S_{5 \leq s \leq 8,1}$
	$\mathbf{P}_{1'} \setminus \{P_2, P_4\}$	h_2	$\{P_1, P_3, P_5\}$	
	$\mathbf{P}_{1'} \setminus \{P_1, P_4\}$	h_3	$\{P_2, P_3, P_5\}$	
	$\mathbf{P}_{1'} \setminus \{P_1, P_2\}$	h_3	$\{P_3, P_4, P_5\}$	

Table. 5.8 – Possible contributor scenarios for $E_i, i \in \{3, 4\}$

W	N_i	$\mathbf{P}_i \setminus \{\cdot\}$	H	$S_{s,1'}$	$S_{s,i}$	
$\neg w_k, k \in \{3, 2\}$	1	$\mathbf{P}_i \setminus \{P_2, P_3, P_4, P_5, P_6\}$	h_1, h_2	$S_{1,1}, S_{2,1}, S_{5,1}, S_{6,1}$	$\{P_1\}$	$S_{1 \leq s \leq 4,i}$
		$\mathbf{P}_i \setminus \{P_1, P_3, P_4, P_5, P_6\}$	h_1, h_3	$S_{1,1'}, S_{3,1'}, S_{5,1'}, S_{7,1'}$	$\{P_2\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_5, P_6\}$	h_2, h_3, h_4	$S_{2 \leq s \leq 4,1'}, S_{6 \leq s \leq 8,1'}$	$\{P_3\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_5, P_6\}$	h_4	$S_{4,1'}, S_{8,1'}$	$\{P_4\}$	
	2	$\mathbf{P}_i \setminus \{P_3, P_4, P_5, P_6\}$	h_1	$S_{1,1'}, S_{5,1'}$	$\{P_1, P_2\}$	$S_{5 \leq s \leq 8,i}$
		$\mathbf{P}_i \setminus \{P_2, P_4, P_5, P_6\}$	h_2	$S_{2,1'}, S_{6,1'}$	$\{P_1, P_3\}$	
		$\mathbf{P}_i \setminus \{P_1, P_4, P_5, P_6\}$	h_3	$S_{3,1'}, S_{7,1'}$	$\{P_2, P_3\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_5, P_6\}$	h_4	$S_{4,1'}, S_{8,1'}$	$\{P_3, P_4\}$	
		$\mathbf{P}_i \setminus \{P_2, P_3, P_4, P_6\}$	h_1, h_2	$S_{1,1'}, S_{2,1'}, S_{5,1'}, S_{6,1'}$	$\{P_1, P_5\}$	$S_{9 \leq s \leq 12,i}$
		$\mathbf{P}_i \setminus \{P_1, P_3, P_4, P_6\}$	h_1, h_3	$S_{1,1'}, S_{3,1'}, S_{5,1'}, S_{7,1'}$	$\{P_2, P_5\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_6\}$	h_2, h_3, h_4	$S_{2 \leq s \leq 4,1'}, S_{6 \leq s \leq 8,1'}$	$\{P_3, P_5\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_6\}$	h_4	$S_{4,1'}, S_{8,1'}$	$\{P_4, P_5\}$	
$w_k, k \in \{3, 2\}$	1	-	-	-	-	-
	2	$\mathbf{P}_i \setminus \{P_2, P_3, P_4, P_5\}$	h_1, h_2	$S_{1,1'}, S_{2,1'}, S_{5,1'}, S_{6,1'}$	$\{P_1, P_6\}$	$S_{13 \leq s \leq 16,i}$
		$\mathbf{P}_i \setminus \{P_1, P_3, P_4, P_5\}$	h_1, h_3	$S_{1,1'}, S_{3,1'}, S_{5,1'}, S_{7,1'}$	$\{P_2, P_6\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_5\}$	h_2, h_3, h_4	$S_{2 \leq s \leq 4,1'}, S_{6 \leq s \leq 8,1'}$	$\{P_3, P_6\}$	
		$\mathbf{P}_i \setminus \{P_1, P_2, P_4, P_5\}$	h_4	$S_{4,1'}, S_{8,1'}$	$\{P_4, P_6\}$	

but in w_2 for E_4. Otherwise, all the other subsets associated with the worlds correspond to the variant $\{g_i, \overline{o_i}\}$. Thus, if $\{g_i, \overline{o_i}\}$ and $N = 1$ hold, then the scenarios correspond to each potential assailant P_1 to P_4 ($S_{1 \leq s \leq 4, i}$). If $\{g_i, \overline{o_i}\}$ and $N = 2$ apply, then they refer to either all possible a pairings of possible assailants ($S_{5 \leq s \leq 8, i}$), or each possible assailant along with P_5 ($S_{9 \leq s \leq 12, i}$). Now, consider $\{g_i, o_i\}$. Logically $\{g_i, o_i\}$ and $N = 1$ are incompatible. Thus no scenario exists for such a conjunction. Finally, if $\{g_i, o_i\}$ and $N = 2$ apply, then the scenarios correspond to a pairing of a possible assailant with P_6 ($S_{13 \leq s \leq 16, i}$). Hence, an individual not involved in the assault P_6 can only be a contributor in $S_{13 \leq s \leq 16, i}$. Each of the specimens E_3 and E_4 have a total of $|\mathbf{S}_3| = |\mathbf{S}_4| = 16$ scenarios.

5.4.3 Adding considerations on the contributors, the hypotheses, and the scenarios of $E_{1'}$

One has to enquire whether or not a given person P_j is a contributor to a specimen E_i. Clearly, if there is no contributor, there can be no specimen in the first place. For this purpose, let $\mathbf{P}_i$ denote the set of *possible* contributors specified for specimen E_i, that is $\mathbf{P}_{1'} = \{P_1, P_2, ..., P_5\}$ or $\mathbf{P}_i = \{P_1, P_2, ..., P_6\}, i \in \{3, 4\}$. Next, if a specimen is composed of a mixture of DNA from, say, P_1 and P_2 exclusively, then this is written as $\mathbf{P}_{1'} \setminus \{P_3, P_4, P_5\}$ or $\mathbf{P}_i \setminus \{P_3, P_4, P_5, P_6\}, i \in \{3, 4\}$. Thus, $\mathbf{P}_{1'} \setminus \{P_3, P_4, P_5\}$ or $\mathbf{P}_i \setminus \{P_3, P_4, P_5, P_6\}, i \in \{3, 4\}$ state $\{P_1, P_2\}$ as the *actual* contributors out of the possible contributors $\mathbf{P}_{1'}$ or $\mathbf{P}_i$ respectively. Stated otherwise, it is not only defined whose DNA is (actually) in the specimen, but also whose DNA is (actually) *not* in the specimen.

For a definition of each scenario it is necessary − but not sufficient − to consider who actually contributed to a given stain or not. Consider, for example, scenario $S_{5,1'} = \{P_1, P_2, P_5\}$. It can only occur if $g_{1'}$ applies, that is, if the corresponding stain was left by the assailants. However, being a contributor does not qualify as being an assailant, otherwise P_5 could also be an assailant, which she is obviously not. Consequently, after having specified the actual contributors, one has to further qualify which of the actual contributors are the assailants. Such a qualification of an actual contributor being an assailant is established by invoking the hypotheses H.

Finally, consider the scenarios of $E_{1'}$. If, say, it contains P_1 as an actual contributor, then one can directly exclude all the scenarios that contain P_4 in $E_i, i \in \{3, 4\}$ (and vice versa) because there is no scenario in which P_1 and P_4 appear together. Similarly, if it contains P_1 and P_2 then one can exclude all the scenarios that contain P_3 and P_4 (and vice versa). Therefore, the scenarios of $E_{1'}$ directly influence the scenarios in E_3 and E_4, and vice versa.

5.4.4 Identifying the reasoning structure for a BN

As can be seen from Tables 5.7 and 5.8, each scenario $S_{s,1'}$ is defined by a unique configuration of $N_{1'}$, $\mathbf{P}_{1'} \setminus \{\cdot\}$, and H; the scenarios $S_{s,3}$ and $S_{s,4}$ further include W and $S_{s,1'}$. Thus, the probability of each scenario $S_{s,1'}$ is conditioned by $N_{1'}$, $\mathbf{P}_i \setminus \{\cdot\}$, H, and additionally by $W, N_{1'}, \mathbf{P}_i \setminus \{\cdot\}$, and H for the scenarios $S_{s,i}, i \in \{3, 4\}$. Moreover, as was discussed in the previous section a given $S_{s,i}, i \in \{3, 4\}$ can only occur if certain scenarios $S_{s,1'}$ apply. This fact can be accounted for by conditioning a given $S_{s,i}, i \in \{3, 4\}$ on the corresponding scenarios $S_{s,1'}$, as outlined in Table 5.8. Moreover, since one, and only one scenario $S_{s,i}, i \in \{3, 4\}$ must have occurred for each specimen, one has $Pr(\mathbf{S}_{1'} \mid \mathbf{I}) = Pr(\mathbf{S}_i \mid \mathbf{I}) = 1$. Further, N_3 and N_4 are conditioned on W. By doing so one can establish that $Pr(N_3 = 2 \mid w_3, \mathbf{I}) = Pr(N_2 = 2 \mid w_2, \mathbf{I}) = 0$.

Note, the variables N_i, and additionally W for $S_{s,3}$ and $S_{s,4}$, provide no discrimination among scenarios in the groups listed in the outer right column of Tables 5.7 and 5.8. Stated otherwise, if a uniform (prior) probability distribution is assigned to the hypotheses H, and to the actual contributors to a specimen, then each scenario within such a group has the same probability value. A discrimination among these scenarios can only be achieved either on the basis of (a) the DNA profiles influencing the probability that a possible contributor is an actual contributor, or (b) some other evidence that allows to discriminate among the states of H.

5.4.5 BN for the contributor scenarios

Figure 5.4 shows a BN handling the scenarios. All the nodes are Boolean, except for `H`, `W`, and `N1` to `N4`. These are numbered. Node `H` possesses the states from 1 to 4, where each number corresponds to the hypothesis h_1 to h_4. A uniform distribution is assigned to the hypotheses, that is, $Pr(h_1 \mid \mathbf{I}) = Pr(h_2 \mid \mathbf{I}) = ... = 0.25$. Node `R` represents the hypotheses $\{h_p, h_d\}$, and is defined as a logical disjunction of $H = 1$ and $H = 2$. Thus, h_p is `true`, whenever an `H` involves P_1 as an assailant, but `false` otherwise. Node `W` stands for the worlds. It possesses the states 1 to 3, each representing the world w_1 to w_3 respectively. The probability values in the third column of Table 5.4 are assigned to each corresponding state. The nodes `N1` to `N4` refer to the number of contributors and each state to the actual number of contributors. Therefore, `N1` possesses the states 2 and 3, whereas `N3` and `N4` possess the states 1 and 2. The posterior probabilities of Table 5.6 are assigned to each corresponding node and state. However, for `N3` and `N4`, which are conditioned on `W`, these posterior probabilities are only valid given certain worlds. That is, one has to account for the fact that $Pr(N_3 = 2 \mid w_3, \mathbf{I}) = Pr(N_2 = 2 \mid w_2, \mathbf{I}) = 0$. The nodes `Ss1` to `Ss4` represent the scenarios. Each scenario node converges towards the evidence nodes `xorS1` to `xorS4`. Each is defined as an exclusive or over all the scenarios of a given specimen $E_i, i \in \{1', 3, 4\}$ (i.e., $\veebar_{\mathbf{S}_i} S_{s,i}$). By instantiating these nodes as `true`, one assures that all the scenarios of a specimen E_i are mutually exclusive so that $Pr(\mathbf{S}_i \mid \mathbf{I}) = 1$ ($i \in \{1', 3, 4\}$). The nodes `PjinE1` to `PjinE4` represent the possible contributors. For each of the latter nodes, a uniform prior distribution is assumed, because their probability is already captured by the variables W, and $N_{1'}$ to N_4 through the scenarios. The possible contributor nodes are then connected to the BN for the evaluation of the DNA profiles, which is discussed later in Section 5.5.

A scenario node is defined as a logical conjunction of its parent nodes, so that the node represents a conclusive formulation of the scenario it refers to according to Tables 5.7 and 5.8. For instance, scenario $S_{1,1'}$ refers to P_1 and P_2 as the actual contributors. The corresponding node `S11` negates all the nodes `PjinE1` except for `P1inE1` and `P2inE1`. As a consequence, P_1 and P_2 are qualified as actual contributors to E_1. Moreover, `H=1` (i.e., $H = h_1$) and `N1=2` (i.e., $N_{1'} = 2$) must cumulatively apply. Thus, one has `S11` = (`H=1` $\wedge$ `N1=2` $\wedge$ `P1inE1` $\wedge$ `P2inE1` $\wedge_{\mathbf{P}_{1'} \setminus P_1, P_2} \neg$ `PjinE1`).

The scenarios of specimens E_3 and E_4 are further conditioned by scenarios of E_1. This is indicated by the two dashed arcs in Figure 5.4.[12] For example, take scenario $S_{1,3}$ which can only occur if one scenario among $S_{1,1}, S_{2,1}, S_{5,1}, S_{6,1}$ applies. Therefore, the node for this scenario is defined as `S31` = ((`S11` $\vee$ `S21` $\vee$ `S51` $\vee$ `S61`) $\wedge$ (`H = 1` $\vee$ `H=2`) $\wedge \neg$`W=3` $\wedge$ `N3=1` $\wedge$ `P1inE3` $\wedge_{\mathbf{P}_3 \setminus P_1} \neg$ `PjinE3`). All the other scenario nodes can be defined following an analogous procedure.

At this point it is important to state that P_1 to P_4 cannot have left a semen stain at the corresponding extraction points, unless, they were the assailants. The only person who could have left a semen stain for reasons unrelated to the crime is P_6, who intervenes among the scenarios of E_3 and E_4. In other words, for a complete stranger, such as P_1 to P_4, to have left the semen stains at these locations for reasons unrelated to the crime is deemed impossible.

This BN is revealing on how to construct a framework of scenarios for the aggregation of different DNA typing results. However, it is extremely inefficient from a computational point of view. This model was broken down into its logical components in order to render it more efficient for calcaulations. This less revealing but more efficient model is shown in Appendix E.1.

5.5 The evaluation of the DNA typing results

The BN for the evaluation of the DNA typing results is essentially based on a model for DNA mixture profiles proposed in [104]. A general account of that model is given in Subsection 5.5.1. The reader interested in a detailed

[12]The authors are aware that the two arcs in the figure are a notational abuse, and that they do not represent the actual dependence relationship among the scenario variables. However, they were added for instructive purposes. That is, as a reminder that further dependence relationships have to be considered among these variables. It is for this reason that only dashed arcs were used.

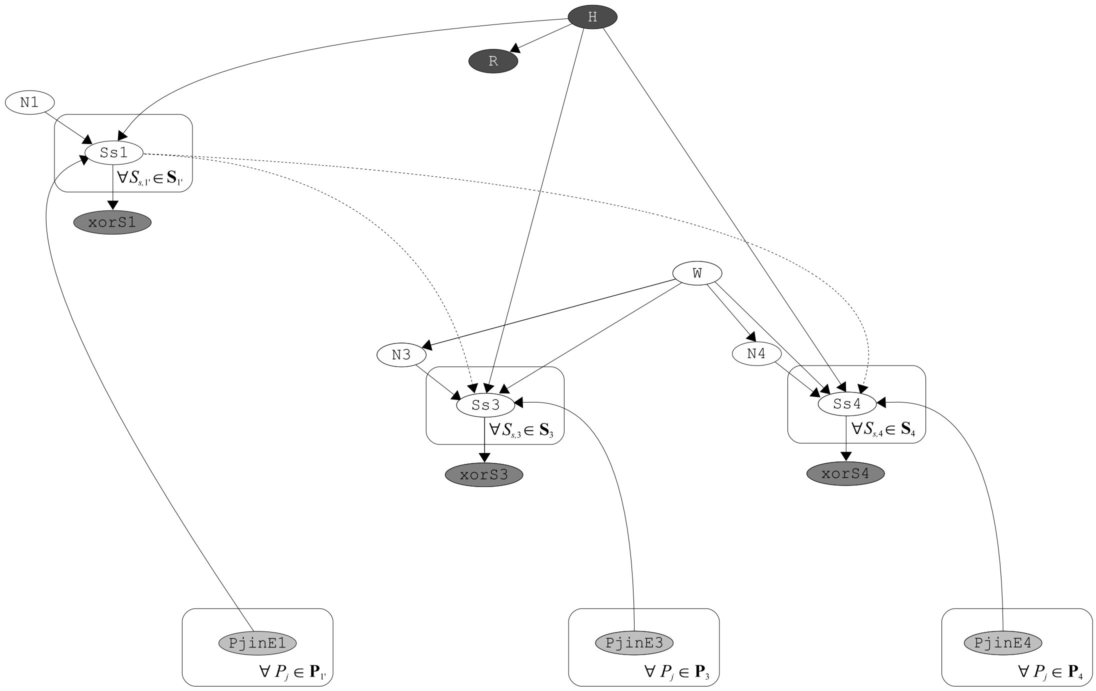


Figure. 5.4 – BN for handling the scenarios of all the specimen

discussion on this model is requested to consult the said paper. The main purpose of this section is to explain the adjustments carried out on Mortera et al.'s model for multiple specimens. The latter adjustments are discussed in Subsection 5.5.2.

5.5.1 A generic BN for mixture profiles

Figure 5.5 shows a generic BN for the evaluation of multiple typing results from different specimens. As can be seen, the BN is very similar to the one used for inferring the number of contributors in Section 5.3.2. It is represented by using three plates. The subgraph in the first plate is replicated for each possible contributor considered ($\forall P_j \in \mathbf{P}$). It is referred to as the 'person-subgraph'. The nodes in this plate are discussed first. The second plate instructs us to replicate the corresponding subgraph for each marker L_l that was examined ($\forall L_l \in \mathbf{L}$). It is referred to as the 'locus-subgraph'. Finally, the last plate is the specimen-subgraph. It is replicated for each specimen that is analyzed ($\forall E_i, i = 1, 2, ..., n$).

The nodes `PjLlpg`, `PjLlmg`, and `PjA1l` to `PjAnl` are exactly the same as in Section 5.3.2. Node `PjLlgt` refers to person P_j's genotype at locus L_l. This node is labelled. That is, each state is represented by a string. In the present case the string stands for a possible genotype (e.g., 'AA', 'AB', or 'BB'). The probability value 1 is assigned for state configurations of `PjLlpg` and `PjLlmg` corresponding to the state of `PjLlgt`, and 0 otherwise (e.g., `PjLlgt` = AA if `PjLlpg` = A and `PjLlmg` = A). However, for markers that contain numerous alleles, it might be useful not to consider all the genotypes but only the most likely ones. In such cases, a state `other` can be added to pool all the genotype states that one does not wish to consider explicitly. The nodes `PjinEi` are the same as previously discussed. They are also connected to the BN for the contributor scenarios (see Section 5.4.5). The nodes `PjA1linEi` to `PjAnlinEi` are defined as logical conjunctions of their parent nodes (e.g., `PjA1linEi` = `PjinEi` ∧ `PjA1l`). In other words, person P_j can contribute an allele, say, $A_{1,l,i}$ to E_i only if he possesses this allele, and if he actually contributed to E_i.

The nodes `A1lEi` to `AnlEi` refer to the whether an allele $A_{k,l,i}$ is present at the locus L_l in the profile from specimen E_i. Naturally, such an allele is present if at least one P_j possesses such an allele. Hence, these nodes are defined as logical disjunctions (e.g., `A1lEi` = $\vee_{\mathbf{P}}$ `PjA1linEi`). The nodes `obsA1lEi` to `obsAnlEi` include considerations on observational errors. They are the same as discussed in Section 5.3.2.

As indicated by the plates, the genotype of a person at a given marker (represented by `PjLlgt`, `PjLlpg`, and `PjLlmg`) as well as the presence of a certain allele in that person's genotype (represented by `PjA1l` to `PjAnl`) is *not* replicated for each specimen E_i. This describes the belief that the genotype of a person − and, therefore, the alleles it contains − remains stable from one specimen to another.

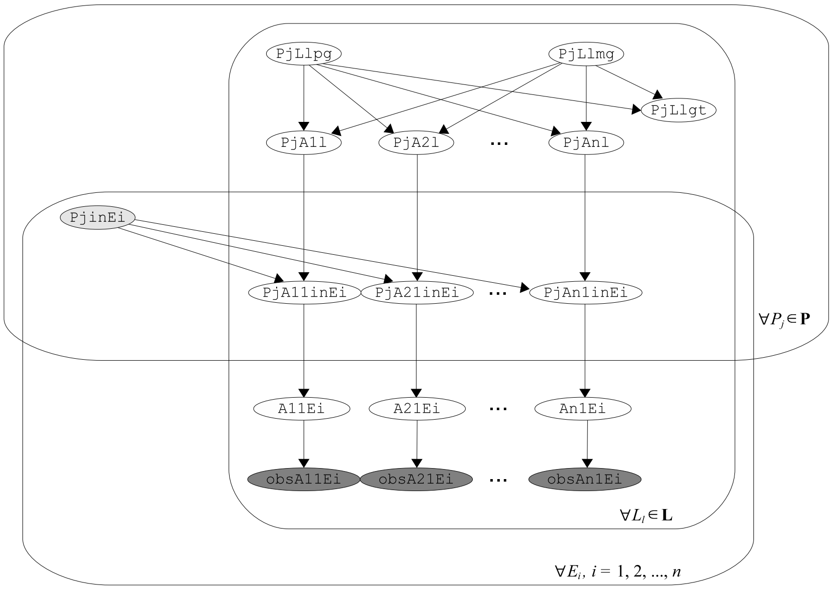


Figure. 5.5 – Generic BN for evaluating multiple typing results of mixture profiles. The observations are instantiated at the observational nodes colored in dark grey. The nodes colored in light grey are exactly the same as in the BN of Figure 5.4. These nodes connect the BN for the scenario to the BN for the typing results.

5.5.2 Adapting the generic BN to the case

The BN presented in the previous section was adapted to the case in question. The first and third modifications are due to the particularity of case itself. The second modification is performed in order to reduce the size of the BN.

First, the number of possible contributors is not the same for all the specimens. Five persons are considered for the first specimen $E_{1'}$ but six persons for the other two specimens. As a result, the person-subgraph for specimen $E_{1'}$ is established for five possible contributors P_1 to P_5, but for six possible contributors P_1 to P_6 for specimens E_3 and E_4. Note that in the present case, the graph for the genotype and its alleles (comprising the nodes `P6Llgt`, P6L1pg, and P6L1mg, and P6A1l to P6Anl) was created once for both specimens E_3 and E_4. This, stems from the assumptions that P_6 is the same person in E_3 and E_4. If assumed otherwise, then one such graph must be provided separately for each specimen.

Second, observational errors are not considered for the reference profiles. Stated otherwise, if for example, the genotype AB was observed for P_1 at L_2 (LDLR), then it is assumed that this genotype is certain[13]. This decision allows for a considerable reduction in size of the BN. Namely, one is not held to model the genotype and its alleles for all the persons whose genotypes are known (i.e., P_1, P_2 and P_3). Whether a specimen E_i contains a given allele $A_{k,i}$ stemming from these individuals can be entirely governed by the question of whether a contribution occurred or not (i.e., by the nodes PjinEi).

Third, for the same reasons as was done with the BN for the computation of the probability of the number of contributors (see Section 5.3.2), the observational nodes are duplicated for each allele in loci L_1 to L_6 of $E_{1'}$. An example of how the BN in Figure 5.5 was adapted is shown in Appendix E.2.

5.6 Inferences and analyses

The question raised by the case is whether Sutton is an assailant or not. This is the question that is investigated here. Remember that $h_p = h_1 \vee h_2$ and $h_d = h_3 \vee h_4$ denote the hypotheses of the prosecution and the defense respectively. The present study focuses on the inferential force that the typing results exert on the hypotheses $\{h_p, h_d\}$, and the inferential interactions among the typing results from the different specimens. However, the analysis of each allele and each locus is omitted here. The inferential force is measured by the weight of evidence (WoE) and the likelihood ratio (LR). The present discourse uses the notation suggested by I.J. Good [58, 64, e.g.]. A short explanation on how to compute the LR and WoE of the typing results is given in Subsection 5.6.1. Subsection 5.6.2 explains how the observations were instantiated, and how the relevant probabilities were retrieved. The WoE and LR obtained from the computations with the final BN are presented and discussed in Subsection 5.6.3. The analysis of the inferential interactions among the specimens, as well as the inferential dissonance among the specimens are analyzed in Subsections 5.6.4 and 5.6.5 respectively. Some measurements used for this analysis are published elsewhere by the present authors. Other measurements were published in [125]. Subsection 5.6.6 examines the posterior probabilities of the hypotheses and the scenarios of each specimen.

5.6.1 Likelihood ratio and weight of evidence for the present case

In order to produce an LR or WoE, the probabilities $Pr(A^*_{k,i} \mid h_p, \mathbf{I})$ and $Pr(A^*_{k,i} \mid h_d, \mathbf{I})$ must be retrieved.[14] Moreover, The alleles, and by extension the loci, inform H through the scenarios. The scenarios of each specimen are mutually exclusive and interconnected with the scenarios from other specimens. The alleles and the loci are not independent given H. The LR for the typing results of all the specimens is, therefore, given by

$$F(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = F(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) F(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I}) F(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}), \tag{5.2}$$

[13] The reasoning behind this choice is that DNA from living individuals can be retested if necessary. However, this might not always be possible for DNA from a crime stain due to its strictly limited quantity.

[14] Strictly speaking, the alleles are not only conditioned by the hypothesis and the circumstantial information, but also by the known genotypes of persons P_1, P_2, and P_5. However, for a better visibility the genotypes are not explicitly listed after the conditioning bar.

where

$$
\begin{aligned}
&F(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) \\
&= F(h_p : A^{1*}_{1,1,1'} \mid \mathbf{I}) \prod_{\mathbf{A}^*_{1'} \setminus A^{1*}_{1,1,1'}} F(h_p : A^*_{k,l,1'} \mid obsA_{1,l,1'}, A^*_{2,l,1'}, \ldots, A^*_{n-1,l,1'}, \mathbf{I}) \\
&= \frac{Pr(A^{1*}_{1,1,1'} \mid h_p, \mathbf{I})}{Pr(A^{1*}_{1,1,1'} \mid h_d, \mathbf{I})} \prod_{\mathbf{A}^*_{1'} \setminus A^{1*}_{1,1,1'}} \frac{Pr(A^*_{k,l,1'} \mid A^*_{1,l,1'}, A^*_{2,l,1'}, \ldots, A^*_{n-1,l,1'}, h_p, \mathbf{I})}{Pr(A^*_{k,l,1'} \mid A^*_{1,l,1'}, A^*_{2,l,1'}, \ldots, A^*_{n-1,l,1'}, h_d, \mathbf{I})},
\end{aligned}
\tag{5.3}
$$

$$
\begin{aligned}
&F(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I}) \\
&= F(h_p : A^*_{1,1,3} \mid \mathbf{A}^*_{1'}, \mathbf{I}) \prod_{\mathbf{A}^*_3 \setminus A^*_{1,1,3}} F(h_p : A^*_{k,l,3} \mid A^*_{1,l,1'}, A^*_{2,l,3}, \ldots, A^*_{n-1,l,3}, \mathbf{A}^*_{1'}, \mathbf{I}) \\
&= \frac{Pr(A^*_{1,1,3} \mid \mathbf{A}^*_{1'}, h_p, \mathbf{I})}{Pr(A^*_{1,1,3} \mid \mathbf{A}^*_{1'}, h_d, \mathbf{I})} \prod_{\mathbf{A}^*_3 \setminus A^*_{1,1,3}} \frac{Pr(A^*_{k,l,3} \mid A^*_{1,l,3}, A^*_{2,l,3}, \ldots, A^*_{n-1,l,3}, \mathbf{A}^*_{1'}, h_p, \mathbf{I})}{Pr(A^*_{k,l,3} \mid A^*_{1,l,3}, A^*_{2,l,3}, \ldots, A^*_{n-1,l,3}, \mathbf{A}^*_{1'}, h_d, \mathbf{I})},
\end{aligned}
\tag{5.4}
$$

$$
\begin{aligned}
&F(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \\
&= F(h_p : A^*_{1,1,4} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \prod_{\mathbf{A}^*_4 \setminus A^*_{1,1,4}} F(h_p : A^*_{k,l,4} \mid A^*_{1,l,4}, A^*_{2,l,4}, \ldots, A^*_{n-1,l,4}, \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \\
&= \frac{Pr(A^*_{1,1,4} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, h_p, \mathbf{I})}{Pr(A^*_{1,1,4} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, h_d, \mathbf{I})} \prod_{\mathbf{A}^*_4 \setminus A^*_{1,1,4}} \frac{Pr(A^*_{k,l,4} \mid A^*_{1,l,4}, A^*_{2,l,4}, \ldots, A^*_{n-1,l,4}, \mathbf{A}^*_{1'}, \mathbf{A}^*_3, h_p, \mathbf{I})}{Pr(A^*_{k,l,4} \mid A^*_{1,l,4}, A^*_{2,l,4}, \ldots, A^*_{n-1,l,4}, \mathbf{A}^*_{1'}, \mathbf{A}^*_3, h_d, \mathbf{I})}.
\end{aligned}
\tag{5.5}
$$

The WoE of all the specimens is given by taking the logarithm of the LRs (for which the base of 10 is used in this paper). In that way the quantities become additive

$$
W(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) + W(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I}) + W(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \tag{5.6}
$$

where

$$
\begin{aligned}
&W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) \\
&= W(h_p : A^*_{1,1,1'} \mid \mathbf{I}) \sum_{\mathbf{A}^*_{1'} \setminus A^*_{1,1,1'}} W(h_p : A^*_{k,l,1'} \mid obsA_{1,l,1'}, A^*_{2,l,1'}, \ldots, A^*_{n-1,l,1'}, \mathbf{I}),
\end{aligned}
\tag{5.7}
$$

and similarly

$$
\begin{aligned}
&W(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I}) \\
&= F(h_p : A^*_{1,1,3} \mid \mathbf{A}^*_{1'}, \mathbf{I}) \sum_{\mathbf{A}^*_3 \setminus A^*_{1,1,3}} W(h_p : A^*_{k,l,3} \mid A^*_{1,l,1'}, A^*_{2,l,3}, \ldots, A^*_{n-1,l,3}, \mathbf{A}^*_{1'}, \mathbf{I}),
\end{aligned}
\tag{5.8}
$$

and

$$
\begin{aligned}
&W(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \\
&= W(h_p : A^*_{1,1,4} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}) \sum_{\mathbf{A}^*_4 \setminus A^*_{1,1,4}} W(h_p : A^*_{k,l,4} \mid A^*_{1,l,4}, A^*_{2,l,4}, \ldots, A^*_{n-1,l,4}, \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I}).
\end{aligned}
\tag{5.9}
$$

Equations 5.2 to 5.9 produce LRs and WoEs following a joint consideration of all the specimens. However, it is also possible to compute the LR and WoE of each specimen in isolation. That is, without taking into account the reciprocal influences among the different specimens. This implies that the observed alleles of one specimen are never conditioned by the observed alleles of another specimen. The LR obtained under these settings is denoted as $F_{\perp}(\cdot)$, and given by

$$F_{\perp}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = F(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) F(h_p : \mathbf{A}^*_3 \mid \mathbf{I}) F(h_p : \mathbf{A}^*_4 \mid \mathbf{I}). \tag{5.10}$$

Analogously, the WoE is denoted as $W_{\perp}(\cdot)$, and given by

$$W_{\perp}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) + W(h_p : \mathbf{A}^*_3 \mid \mathbf{I}) + W(h_p : \mathbf{A}^*_4 \mid \mathbf{I}). \tag{5.11}$$

Finally, recall that the typing results can only be explained by taking into account observational errors, or by considering highly improbable worlds and number of contributors. However, since the present research only considers the most likely worlds and number of contributors, this means that the DNA typing results can only be examined based on $obs\mathbf{A}_i$. An evaluation of the results based on $\mathbf{A}_i$ leads to the propagation of inconsistent evidence in the network. More precisely, the typing results are impossible under any given hypothesis and scenario.

5.6.2 Node instantiations and probability retrievals

All the observations on the DNA profiles made are entered into the BN by an instantiation of the corresponding observational node. For the present BN, the authors used two different sets of values for the probabilities of observational errors for the computation of the LRs and WoEs. The first set consists of the same probability values used for the computation of the probability of the number of contributors, namely $Pr(\overline{a^*_{k,i}} \mid a_{k,i}, \mathbf{I}) = 0.02$ and $Pr(a^*_{k,i} \mid \overline{a_{k,i}}, \mathbf{I}) = 0.001$ ('error 1' as short for 'observational errors 1'). The second set of values is $Pr(\overline{a^*_{k,i}} \mid a_{k,i}, \mathbf{I}) = 0.001$ and $Pr(a^*_{k,i} \mid \overline{a_{k,i}}, \mathbf{I}) = 0.001$ ('error 2' as short for 'observational errors 2'). Hence, the error probabilities are smaller for the set 'error 2' than 'error 1'.

In order to retrieve the probabilities for the *joint evaluation of the specimens*, one has to instantiate `xorS1 = true`, `xorS3 = true`, and `xorS4 = true`. This operation d-connects[15] all the scenarios within a specimen, and as a consequence, also between the specimens. Next, the probabilities at the nodes `obsAklEi` need to be retrieved for each observation made on the typing results (see Table 5.1).

The probability retrieval for the *evaluation of the specimens in isolation* is slightly different. Namely, `xorS1 = true`, `xorS3 = true`, or `xorS4 = true` is instantiated exclusively for the specimen considered in isolation. That is, if one were to focus on specimen, say $E_{1'}$, then `xorS1 = true` is entered but no instantiation is specified for the nodes `xorS3` and `xorS4`. This prevents the communication among the subgraphs associated with each specimen. Naturally, the probability retrievals and instantiations at the observational nodes were realized exclusively for the targeted specimen. The probabilities at the observational nodes are retrieved by following the same procedure used for the joint evaluation of the specimens. In other words, one is following the computation shown in Equations 5.3 or 5.7 analogously for all the specimens.

In general, all observation regarding an allele $A^*_{k,l,i}$ are conditionally dependent given H. Thus, in any case, a given observation must be instantiated after its probability is retrieved, and remain so for the probability retrieval at the next observational node of the specimen. If the probabilities of the observational nodes should be further conditioned by observations from other specimens, then the corresponding instantiations have the to be carried out beforehand (this includes the instantiations at the nodes `xorS1 = true`, `xorS3 = true`, and `xorS4 = true`). For instance, the probability retrieval for the computation of the WoE $W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$ requires instantiations

[15] D-connection refers to the setup of a communication pathway between two or more parts of a graph. In contrast, d-separation refers to the transection of such a communication path [108].

at the observational nodes of specimen E_3 and E_4 corresponding to the findings $\mathbf{A}_3^*$ and $\mathbf{A}_4^*$. Next the probabilities at the observational nodes have to be retrieved. For example, in specimen $E_{1'}$ the allele 1.1 DQA was observed twice: once for E_1, and once for E_2. After instantiating `R = true` (i.e., $h_p = \{h_1, h_2\}$) the marginal probability is retrieved at the `true`-state (i.e., allele 1.1 DQA was observed) of the first observational node `obsA11E1_1`. This marginal probability corresponds to $Pr(a^*_{1,1,1'} \mid h_p, \mathbf{I})$. The same procedure is repeated with `R = false` (i.e., $h_d = \{h_3, h_4\}$) producing $Pr(a^*_{1,1,1'} \mid h_d, \mathbf{I})$. Node `obsA11E1_1 = true` is instantiated before the probabilities at the second observational node `obsA11E1_2` are retrieved. By instantiating alternatively `R = true` and `R = false` the probabilities $Pr(a^*_{2,1,1'} \mid a^*_{1,1,1'}, h_p, \mathbf{I})$ and $Pr(a^*_{2,1,1'} \mid a^*_{1,1,1'}, h_d, \mathbf{I})$ can be retrieved respectively. Next, `obsA11E1_2 = true` is instantiated. The next observation is the absence of 1.2 DQA in the typing results. Thus, by instantiating alternatively `R = true` and `R = false` the marginal probabilities at the `false` state of node `obsA21E1_1` correspond to $Pr(\overline{a^*_{1,2,1'}} \mid a^*_{1,1,1'}, a^*_{2,1,1'}, h_p, \mathbf{I})$ and $Pr(\overline{a^*_{1,2,1'}} \mid a^*_{1,1,1'}, a^*_{2,1,1'}, h_d, \mathbf{I})$ respectively.

In short, the marginal probabilities are retrieved at the relevant state of the observational node once under `R=true` and once under `R=true`. If the allele was observed in the DNA typing results, then the relevant state at the observational node is `true` and `false` otherwise. Before retrieving the probabilities at the subsequent observational node the present observational node is instantiated according to the observation made (i.e., `true` if the allele was observed in the DNA typing results and `false` otherwise). As a consequence, the observational nodes are cumulatively instantiated one after another across all loci of the DNA typing results of the specimen.

5.6.3 LR and WoE of the specimens

The LRs and WoEs produced by the model are shown in Table 5.9. The row for the joint inferential force correspond to the computations following Equations 5.2 and 5.6 respectively. Conversely, the row for the inferential force of the specimens in isolation follows the computation according to Equations 5.10 and 5.11. As can be seen from the results, the joint evaluation of the specimens produces inferential forces against the hypotheses that Sutton is an assailant ($h_p = \{h_1, h_2\}$) and supporting, in turn, the alternative hypotheses $h_d = \{h_3, h_4\}$. However, if the specimens are evaluated in isolation h_d is favored, rather than h_p. This suggests the presence of inferential interaction between the typing results from the different specimens. The subject of inferential interaction is examined later. Note also, the inferential forces produced by the joint consideration vary significantly for different observational error probabilities. In contrast, only small variations can can be observed for the inferential forces produced by an isolated evaluation of the specimens. This suggests that smaller observational error probabilities have a stronger effect on the inferential forces in the joint evaluation setting than in the isolated setting. An exhaustive list of the individual WoEs of each allele is given in Appendix E.3.

Table. 5.9 – LRs and WoEs of the DNA typing results

		Joint	Isolated
Error 1	LR	0.185	10.197
	WoE	-0.733	1.009
Error 2	LR	0.007	9.333
	WoE	-2.190	0.970

5.6.4 Inferential interaction among the specimens

A previous article by the authors showed how Schum's redundance measure [125] can be generalized for an arbitrary number of items of evidence (see article in Chapter 4 of Part III). This extended measure was named 'inferential interaction measure', and is based on the WoE-metric for the inferential force. The inferential interaction measure portraits the difference arising between items of evidence that are independent given some hypotheses and those that are not independent, when the multiplication rule of probability is applied. In other words, it unveils the difference between Equation 5.6 and 5.11, and is given by

$$ia(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = \frac{W_\perp(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) - W(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I})}{W_\perp(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I})}, \tag{5.12}$$

where $W_\perp(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) \neq 0$. For a single specimen, say, $E_{1'}$ one has

$$ia(h_p : \mathbf{A}^*_{1'} \mid \mathbf{A}^*_3, \mathbf{A}^*_4 \mathbf{I}) = \frac{W_\perp(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) - W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})}{W_\perp(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I})}, \tag{5.13}$$

where $W_\perp(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) \neq 0$.

As was shown in the article reproduced in III.4, an interaction value $ia[\cdot] < 0$ demonstrates the presence of a synergistic relationship among the examined items of evidence. Stated otherwise, the inferential force of an item of evidence increases upon the knowledge on some other item of evidence. A value $ia[\cdot] > 0$ establishes the presence of a directional change among the items. That is, the inferential force changes its direction (i.e., supporting the opposite hypothesis) upon the knowledge on some other item of evidence. Table 5.10 outlines the interaction values obtained according to Equations 5.12 and 5.13, and their corresponding WoEs.

The results show that the directional change is the prevalent inferential interaction among the three typing results. Specimen $E_{1'}$ supports h_p over h_d, when considered in isolation, but it supports hypothesis h_d over h_p, when considered together with the other specimens. The situation is reversed for the typing results of E_3. Specimen E_4 is an exception as it exhibits a synergistic relationship with the other specimens. The dominating inferential interaction is that of a directional change as can be seen from the last row of Table 5.10. It is possible to retrace the origin of these inferential interaction values by examining the WoEs of each allele in each specimen. Such an examination, however, would exceed the scope of this paper, and is omitted. The most important finding is the presence of a directional change among the specimens. Namely, an evaluation of each specimen in isolation provides weight in favor of the hypothesis that Sutton is an assailant, whereas the opposite is true, when the specimens are evaluated jointly. This finding is in agreement with Thompson's observations in [139].

Table. 5.10 – Inferential interaction among the typing results of the three specimens

				$ia(h_p : \cdot \mid \mathbf{I})$		Interaction type
		Error 1	Error 2	Error 1	Error 2	
$E_{1'}$	$W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$	-0.118	-0.116	1.06	1.06	Directional change
	$W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I})$	1.923	1.925			
E_3	$W(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_4, \mathbf{I})$	0.450	1.645	1.94	4.44	Directional change
	$W(h_p : \mathbf{A}^*_3 \mid \mathbf{I})$	-0.479	-0.478			
E_4	$W(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I})$	-2.717	-4.182	-5.24	-7.76	Synergy
	$W(h_p : \mathbf{A}^*_4 \mid \mathbf{I})$	-0.435	-0.477			
All	WoE according to Equation 5.6	-0.733	-2.190	1.73	3.26	Directional change
	WoE according to Equation 5.11	1.009	0.970			

5.6.5 Dissonance among the specimens

A directional change always stems from a dissonance among the inferential forces of different items of evidence (see article in Chapter 4 of Part III). A dissonance among items of evidence implies that some amount of inferential force is lost, since forces pointing in opposite directions compensate each other by nullification. The amount of inferential force lost due to a dissonance can be measured based on the WoE. The potential weight produced by a collection of items is the total amount WoE produced, irrespective of directionality of the inferential force of each item. It is denoted as $W_{\text{pot}}(\cdot)$. The WoE expressed by a collection of items of evidence is the totalWoE produced by accounting for the directionality of the inferential force of each item. It is denoted as $W_{\text{ex}}(\cdot)$. The WoE lost due to dissonance $W_{\text{diss}}(\cdot)$ is given by the difference between the potential weight and the weight expressed so that

$$W_{\text{diss}}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) = W_{\text{pot}}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) - W_{\text{ex}}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}), \tag{5.14}$$

where

$$\begin{aligned} W_{\text{pot}}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) &= |W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I})| + |W(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I})| + |W(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I})|, \\ W_{\text{ex}}(h_p : \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4 \mid \mathbf{I}) &= |W(h_p : \mathbf{A}^*_{1'} \mid \mathbf{I}) + W(h_p : \mathbf{A}^*_3 \mid \mathbf{A}^*_{1'}, \mathbf{I}) + W(h_p : \mathbf{A}^*_4 \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{I})|. \end{aligned} \tag{5.15}$$

The amount of WoE lost through can never be negative since $W_{\text{pot}}(\cdot) \geq W_{\text{ex}}(\cdot)$, and it is zero only if the items of evidence are harmonious (i.e., absence of dissonance). The WoE lost through, the corresponding potential WoE and expressed WoE are outlined in Table 5.11. As expected, the weight lost through is larger than zero in any case, indicating the presence of dissonance among the specimens. Notice that the actual amount of weight lost through is almost the same between 'error 1' and 'error 2'. However, by considering the weight lost through relative to their potential weights ($W_{\text{diss}}(\cdot)/W_{\text{pot}}(\cdot)$) the difference becomes remarkable. For 'error 1' the weight lost relative to the potential weight is roughly 85% (i.e., 3.97/4.7). For 'error 2', however, one has 65% (i.e., 3.98/6.17), that is, a relative loss smaller by 20%. Stated otherwise, even though the actual amount of weight lost through is almost the same between 'error 1' and 'error 2', in proportion to the potential weight produced by the specimens, there is a reduction in dissonance. It appears that the smaller the observational errors, the less dissonant the typing results of the specimen.

Table. 5.11 – Dissonance among the typing results of the three specimens

	WoE potential $W_{\text{pot}}(h_p : \cdot \mid \mathbf{I})$	WoE expressed $W_{\text{ex}}(h_p : \cdot \mid \mathbf{I})$	WoE lost through $W_{\text{diss}}(h_p : \cdot \mid \mathbf{I})$
Error 1	4.70	0.73	3.97
Error 2	6.17	2.19	3.98

5.6.6 Scenarios and posterior probabilities of H

Figure 5.6 outlines the posterior probabilities of H, and of the scenarios of each specimen, that is, after instantiating all observations in the BN. As can be seen, irrespective of the observational error probabilities applied, the observations provoke the same tendencies in the probability distributions of the hypotheses and the scenarios. Consider the posterior probabilities of the hypotheses. The most likely hypothesis is h_4 irrespective of whether probabilities of 'error 1' or 'error 2' were applied. For 'error 1' this posterior probability amounts to 84%, and for 'error 2' even to 99% (see Appendix E.4). Thus, given the DNA typing results it is far more likely that two unknown persons are the assailants, rather than any other hypothesis considered in this paper and by the court. The most likely contributor scenario for specimen $E_{1'}$ is the one involving two unknown assailants ($S_{4,1'}$), followed by the much less likely scenario of Sutton and an unknown assailant ($S_{2,1'}$). The same pattern is proportionally repeated among the scenarios $S_{5,1'}$ to $S_{8,1'}$, however, on a much smaller range of probabilities. For the specimens E_3 and E_4 none of the most likely contributor scenarios involve Sutton or Adams as assailants. The most likely scenarios involve either unknown assailant P_3 (i.e., $S_{3,\cdot}$) or P_4 (i.e., S_4). This is expected, since the typing results of these specimens do not correspond to Sutton's or Adams' DNA profiles. Note also, that for both, 'error 1' and 'error 2', the probabilities of scenarios $S_{3,\cdot}$ and $S_{4,\cdot}$ seem to gravitate around the probability value of 50%. This result coincides with an intuitive reasoning approach. Namely, since Sutton's and Adams' DNA profile do not correspond, some other persons must be the contributors. If it is only one contributor (the most likely number of contributors for these specimens) it must be either assailant P_3 or P_4. Given that their DNA profile is unknown, one cannot discriminate between these two scenarios. Hence, P_3 and P_4 appear to be equally likely candidates.

One concludes that the DNA evidence suggests neither Sutton nor Adams as an assailant. Given the circumstantial information considered here, the allele frequencies employed, and the reasoning pattern applied, it is logically inconceivable to arrive at a different conclusion. The probability values of the scenarios can found in Appendix E.5.

5.7 Discussion and Conclusion

The paper presented a BN that aggregates DNA typing results stemming from four distinct specimens on the one hand, and probabilistic analyses on the typing results based on that BN on the other. Meanwhile, each typing result was treated as a mixture profile. The hypotheses of interest were, whether the suspects Sutton and Adams, one of the suspects together with an unknown person, or two unknown persons, committed the sexual assault. The logical organization of the specimens in the light of crime-level hypotheses turned out to be the most challenging task of the present study. Up to date, no publication on how to accomplish a combination of different typing results exists to the best knowledge of the authors. Indeed, the literature on the probabilistic combination of items of evidence is very scarce in general. This meant that the method for the logical organization of the typing results had to be established almost from scratch. The model created produced inferences that corroborate the findings on the typing results made by W.C. Thompson.

5.7.1 Model creation

The authors chose to extend well known concepts for the evaluation of single items of evidence given crime-level propositions. These concepts involve (a) the relevance of the DNA from specimens for the crime in question, and (b) the probability that the DNA is present in the specimen for reasons unrelated to the crime in question [56, 130]. Based on these two concepts, the authors enumerated the different 'worlds', where each world represents a possible setting of (a) and (b) that could have produced the specimens in question. In the present case the initial number of worlds was 256. Generally, the total number of worlds depends sensibly on the number of specimens: the number of worlds increases exponentially relative to the number of specimens. The number of worlds is reduced by different strategies, all of which are essentially based on the exploitation of circumstantial information and forensic expert knowledge. The authors emphasize that without any further information an examination of every single world must be envisaged, which might well be impossible to accomplish from a practical standpoint. However, the rigorous application of circumstantial information and forensic expert knowledge, enables the exclusion of worlds that are impossible to have occurred. A further enquiry on the probability of each worlds, allows us to discriminate between likely and unlikely worlds. The present study retained the three most probable worlds, each possessing a probability of 0.998, 0.001, and 0.001 respectively. Next, the probability of the number of contributors was computed for each specimen. The BN applied for that purpose was proposed in [12]. Again, only the most probable number of contributors were retained for each specimen. Based on the collection of three likeliest worlds, the likeliest number of contributors, and the hypotheses of interest, a logical framework defined upon sets of contributor scenarios can be

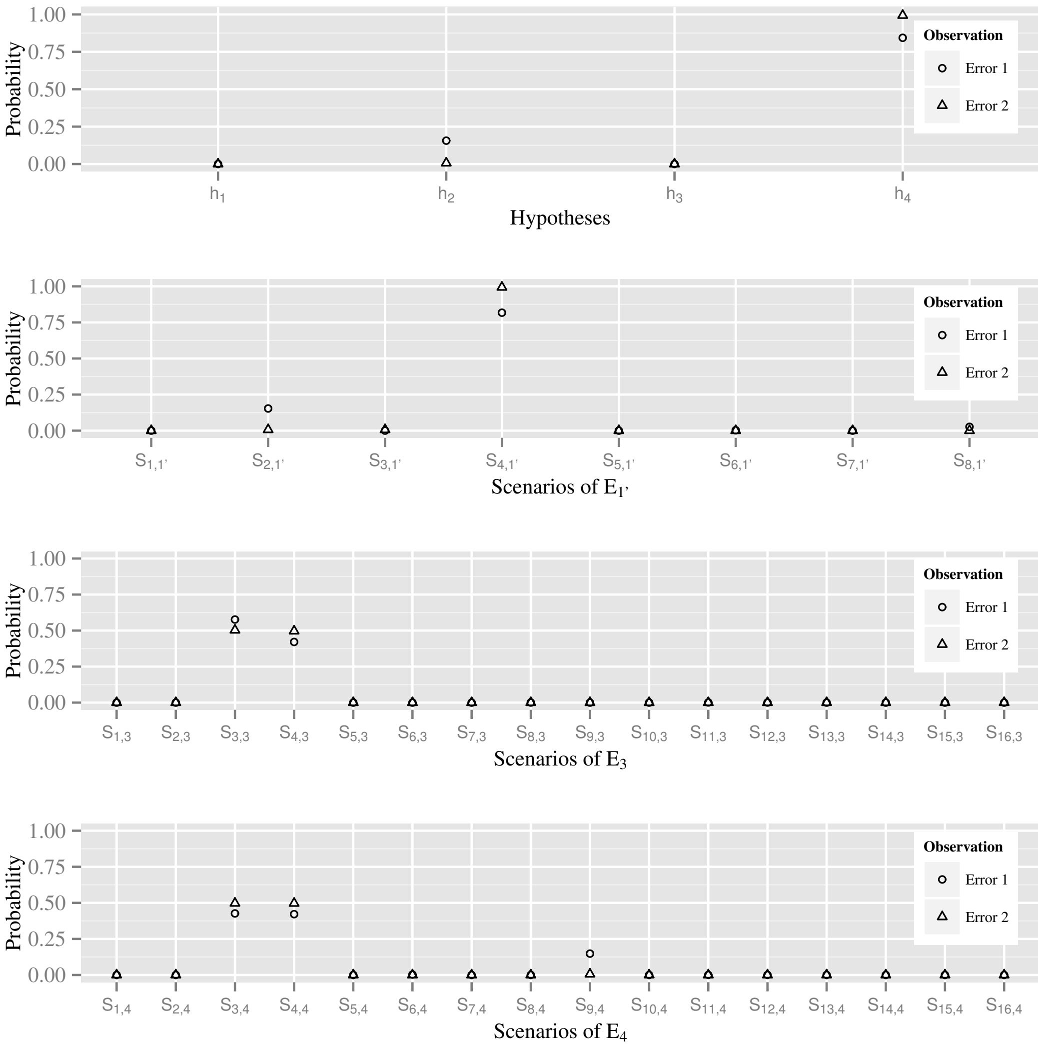


Figure. 5.6 – Posterior probabilities of the hypotheses and the contributor scenarios given the DNA typing results of all the specimens

established for each specimen. A BN reproducing this logical framework was created. This BN served as a gateway to relate the DNA typing results to the hypotheses. The typing results were modeled based on the BNs proposed in [104]. The BN creation process lead to the insight that circumstantial information not only dictates the choice of the relevant population for probability assignments, but also the reasoning pattern itself (i.e., the BN), when items of evidence are combined. That is, such information governs the number of possible worlds and, consequentially, also the number of possible contributor scenarios. BNs require complex reasoning processes to be made transparent and thought through in detail. This enables a clear demonstration on how circumstantial information was employed, what scenarios were considered, and what reasoning structure was applied.

5.7.2 Results produced by the model

The result obtained from computations based on the proposed BN, corroborate the assessments on the typing results made by W.C. Thompson in [139]: the DNA evidence in this case provides no reasonable basis for supporting the guilt of Sutton. More precisely, if the typing results of each specimen are considered in isolation, then they support the hypothesis that Sutton is one of the assailants. However, the support is very weak (i.e., LR ≈ 10 or WoE ≈ 1). In contrast, if the typing results are evaluated jointly, then they clearly support the hypothesis that Sutton is not an assailant of the crime (i.e., LR ≈ 0.185 and WoE ≈ -0.733). The analysis showed further that assigning smaller observational error probabilities to the typing results increases the support in favor of Sutton not being an assailant (i.e., LR ≈ 0.007 and WoE ≈ -2.190). Smaller observational error probabilities also produced smaller relative dissonances among the specimens. The most important feature of the present DNA typing results is, however, the inferential interaction among the different specimens. Namely, the directional change regarding the supported hypothesis, when the evidence is considered in isolation, or in conjunction. This feature was clearly recognized by W.C. Thompson, although, not named as such. David A. Schum was the first person, who gave a clear probabilistic description of the phenomenon he called 'directional change' [125]. The existence of this phenomena is, thus, known for quite some time. However, a discussion on the dangers and prevalence of directional change in forensic case work, or literature is unknown to the authors. Forensic evidence is predominantly evaluated in isolation, and presented to courts from that perspective. This implies that almost nothing is known about the exposure of forensic evidence to such inferential interactions, and about the misinterpretation of evidence due to inferential interactions.

5.7.3 Conclusion

The authors conclude: evidence must be interpreted holistically and not as isolated parts. Such holistic interpretations can only be accomplished based on circumstantial information, and a rigorous application forensic expert knowledge. BNs are invaluable tools for the holistic interpretation of evidence, and can help to render the expert's workflow transparent.

Acknowledgements

This work has beed supported by the Swiss National Science Foundation [grant number 100014_150276].

A Appendix: Example of average uncertainty and information redundance in general patterns of fingerprints

Tables A.1 and A.2 outline the relative frequencies of general patterns of fingerprints in a sample of the Spanish population as observed by Gutièrrez et al. in [67]. Let $X = \{x_1, x_2, x_3, x_4\}$ denote the set of four different general patterns examined in this publication: arch, ulnar loop, radial loop, and whorl. The maximum average uncertainty $H_{\max}(X)$ is obtained if each pattern has the same probability of occurring so that $Pr(x_1) = ... = Pr(x_4) = 1/4$ (i.e., uniform distribution of the patterns). Thus, the maximum average uncertainty is given by

$$H_{\max}(X) = -\sum_{i=1}^{4} Pr(x_i)\log_2 Pr(x_i) = -4 \times \frac{1}{4}\log_2 \frac{1}{4} = 2\text{bits}.$$

The values of the relative frequencies shown in Tables A.1 and A.2 are assigned to the corresponding probabilities $Pr(x_i)$. The average uncertainty of the general pattern $H(X)$ for the thumb of the left hand is given by

$$\begin{aligned} H(X) &= -0.0556 \times \log_2 0.0556 - 0.5707 \times \log_2 0.5707 - 0.0152 \times \log_2 0.0152 - 0.3586 \times \log_2 0.3586 \\ &\approx 1.3160 \text{ bits}, \end{aligned}$$

and its information redundance is

$$R = 1 - \frac{H(X)}{H_{\max}(X)} = 1 - \frac{1.3160}{2} \approx 0.3420,$$

which is a dimensionless quantity. The values for the other fingers in Tables A.1 and A.2 were computed in the same manner. Note, however, that frequency values of 0.0 are omitted when computing the average uncertainty.

The average uncertainty represents the average information transmitted (produced) by an event (here the general pattern X). It can also be interpreted as the average amount of surprise that we associate with an event [90]. Indeed, the larger the average uncertainty the larger the element of surprise, and therefore the more informative the event becomes. Conversely, the smaller the average uncertainty the less surprising, and therefore informative, is the event for us.

Figure A.1 outlines the average uncertainty and the information redundance for each finger on each hand. As can be seen, the little finger has the smallest average uncertainty. The general pattern is the least informative for this finger. This is also reflected in the the relative frequencies: 83.84% of the general patterns of this finger (on both hands) are ulnar loops. Hence, you can expect to observe an ulnar loop almost all the time. This is expressed by a high redundance value of 0.6108 (remember that $0 \leq R < 1$). In contrast, the general pattern of the index has the largest average uncertainty and the smallest redundance value among the fingers. Hence, the general pattern on the index is most informative and the least redundant as compared with the other fingers.

Table. A.1 – Relative frequencies of general patterns, average uncertainty, and information redundance for the fingers of the left hand

		Thumb	Index	Middle finger	Ring finger	Little finger
Frequency:	Arch	0.0556	0.0909	0.0758	0.0354	0.0202
	Ulnar Loop	0.5707	0.4141	0.7424	0.6010	0.8384
	Radial Loop	0.0152	0.1212	0.0253	0.0051	0.0101
	Whorl	0.3586	0.3737	0.1566	0.3586	0.1313
Average uncertainty $H(X)$ in bits		1.3160	1.7409	1.1542	1.1815	0.7785
Information redundance R		0.3420	0.1300	0.4229	0.4093	0.6108

Table. A.2 – Relative frequencies of general patterns, average uncertainty, and information redundance for the fingers of the right hand

		Thumb	Index	Middle finger	Ring finger	Little finger
Frequency:	Arch	0.0202	0.0964	0.0556	0.0556	0.0101
	Ulnar Loop	0.5455	0.3198	0.7626	0.7626	0.8384
	Radial Loop	0.0	0.1574	0.0	0.0	0.0051
	Whorl	0.4343	0.4264	0.1818	0.1818	0.1465
Average uncertainty $H(X)$ in bits		1.1132	1.7955	0.9771	0.9771	0.7250
Information redundance R		0.4434	0.1022	0.5114	0.5114	0.6375

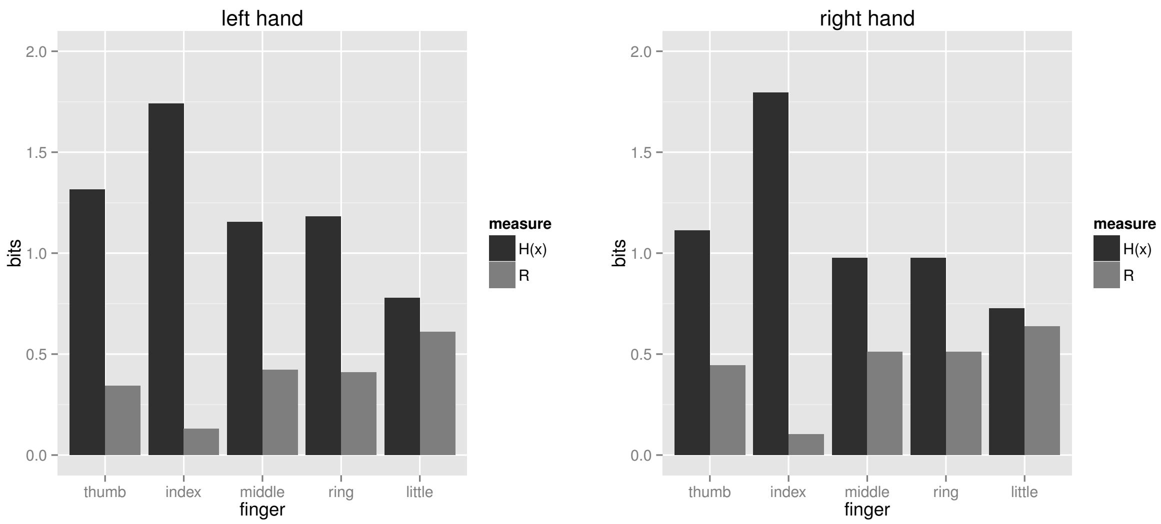


Figure. A.1 – Average uncertainty $H(x)$ and information redundance R of the general pattern for each finger and each hand

B Appendix: Probabilistic ontology of evidence and its combinations

B.1 Reversibility of inconclusive arguments of evidence

By applying condition (ii) we can rewrite $b_1 = 1 - a_1$ and $b_2 = 1 - a_2$. The inferential force of the argument of evidence is then given by

$$V_{r|H} = \frac{a_1 + [\frac{a_2}{1-a_2} - 1]^{-1}}{(1-a_1) + [\frac{a_2}{1-a_2} - 1]^{-1}},$$

which corresponds to the ratio $Pr(h \mid r)/Pr(h \mid \bar{r})$ if the argument is reversible. The posterior probabilities of R can be developed by using the Bayes' theorem so that

$$Pr(h \mid r) = \frac{Pr(r \mid h)Pr(h)}{Pr(r)},$$
$$Pr(h \mid \bar{r}) = \frac{Pr(\bar{r} \mid h)Pr(h)}{Pr(\bar{r})}.$$

Since $Pr(h)$ can be eliminated in the ratio $Pr(r \mid h)Pr(h)/Pr(\bar{r} \mid h)Pr(h)$ we have

$$\frac{Pr(h \mid r)}{Pr(h \mid \bar{r})} = \frac{Pr(r \mid h)}{Pr(\bar{r} \mid h)} \times \frac{Pr(\bar{r})}{Pr(r)} = \frac{Pr(r \mid e)Pr(e \mid h) + Pr(r \mid \bar{e})Pr(\bar{e} \mid h)}{Pr(\bar{r} \mid e)Pr(e \mid h) + Pr(\bar{r} \mid \bar{e})Pr(\bar{e} \mid h)} \times \frac{Pr(\bar{r})}{Pr(r)}.$$

By applying condition (ii) we can rewrite

$$\frac{Pr(h \mid r)}{Pr(h \mid \bar{r})} = \frac{a_2 a_1 + (1-a_2)(1-a_1)}{(1-a_2)a_1 + a_2(1-a_1)} \times \frac{Pr(\bar{r})}{Pr(r)}.$$

The ratio $Pr(\bar{r})/Pr(r)$ can be developed by extending the conversation

$$Pr(r) = \sum_E \sum_H Pr(r \mid E)Pr(E \mid H)Pr(H),$$
$$Pr(\bar{r}) = \sum_E \sum_H Pr(\bar{r} \mid E)Pr(E \mid H)Pr(H).$$

By applying condition (ii) to the ratio $Pr(\bar{r})/Pr(r)$ we obtain

$$\frac{Pr(\bar{r})}{Pr(r)} = \frac{a_2a_1Pr(\bar{h})+(1-a_2)(1-a_1)Pr(\bar{h})+a_2(1-a_1)Pr(h)+(1-a_2)a_1Pr(h)}{a_2a_1Pr(h)+(1-a_2)(1-a_1)Pr(h)+a_2(1-a_1)Pr(\bar{h})+(1-a_2)a_1Pr(\bar{h})}.$$

The impact of the prior probabilities $Pr(h)$ and $Pr(\bar{h})$ is therefore confined in the ratio $Pr(\bar{r})/Pr(r)$. The identity $Pr(h) = Pr(\bar{h})$, that is condition (i), assures that the numerator and the denominator of $Pr(\bar{r})/Pr(r)$ are also identical. As a consequence, $Pr(\bar{r})/Pr(r) = 1$ and

$$\frac{Pr(h \mid r)}{Pr(h \mid \bar{r})} = \frac{Pr(r \mid h)}{Pr(\bar{r} \mid h)},$$

which corresponds to the inferential force of r ($V_{r|H}$):

$$\begin{aligned}\frac{Pr(r \mid h)}{Pr(\bar{r} \mid h)} &= \frac{a_2a_1+(1-a_2)(1-a_1)}{(1-a_2)a_1+a_2(1-a_1)} = \frac{\frac{a_2}{1-a_2}a_1+(1-a_1)}{a_1+\frac{a_2}{1-a_2}(1-a_1)} = \frac{\frac{a_2}{1-a_2}a_1+1-a_1}{a_1+\frac{a_2}{1-a_2}-\frac{a_2}{1-a_2}a_1} = \frac{-a_1[\frac{a_2}{1-a_2}-1]-1}{\frac{a_2}{1-a_2}a_1-a_1-\frac{a_2}{1-a_2}}\\ &= \frac{-a_1[\frac{a_2}{1-a_2}-1]-1}{a_1[\frac{a_2}{1-a_2}-1]-\frac{a_2}{1-a_2}} = \frac{-a_1-[\frac{a_2}{1-a_2}-1]^{-1}}{a_1-\frac{\frac{a_2}{1-a_2}}{[\frac{a_2}{1-a_2}-1]}} = \frac{-a_1-[\frac{a_2}{1-a_2}-1]^{-1}}{a_1-([\frac{a_2}{1-a_2}-1]^{-1}+1)} = \frac{a_1+[\frac{a_2}{1-a_2}-1]^{-1}}{(1-a_1)+[\frac{a_2}{1-a_2}-1]^{-1}}\\ &= V_{r|H}.\end{aligned}$$

B.2 Drag coefficient of chains of reasoning and testimonial evidence

In singly connected chains of reasoning, each variable has one incoming and one outgoing directed arc at most (serial connection). The numerator of the likelihood ratio at the i-th reasoning stage of a total of n reasoning stages is denoted by a_i and the corresponding denominator is denoted by b_i. For n reasoning stages Equation (2.1) becomes, following Schum [125],

$$V_{r|H} = \frac{a_1+D_n}{b_1+D_n}, \tag{B.1}$$

where

$$D_n = D_{n-1} + \frac{b_n}{\prod_{i=1}^{n}(a_i-b_i)} = D_{n-1} + \frac{1}{\prod_{i=2}^{n-1}(a_i-b_i)}[\frac{a_n}{b_n}-1]^{-1}. \tag{B.2}$$

As can be seen from Equation (B.1), the inferential drag is recursively accumulated with every additional reasoning stage. Thus, in general, the longer the argument based on a SCCR, the larger the inferential drag, and the weaker the inferential force.

The BN for the item of testimonial evidence is a chain of reasoning with three stages (i.e. perceptional sensitivity, objectivity, and veracity). The inferential force is therefore given by

$$V_{r|H} = \frac{a_1+D_4}{b_1+D_4} = \frac{a_1+\frac{b_2}{(a_2-b_2)}+\frac{b_3}{(a_2-b_2)(a_3-b_3)}+\frac{b_4}{(a_4-b_4)(a_4-b_4)}}{b_1+\frac{b_2}{(a_2-b_2)}+\frac{b_3}{(a_2-b_2)(a_3-b_3)}+\frac{b_4}{(a_4-b_4)(a_4-b_4)}}.$$

B.3 Bypassing an intermediary variable

The operation of bypassing an intermediary variable requires two conditions that must cumulatively apply [28]: (i) the variable to be bypassed is neither a query variable (such as H) nor an evidence variable (such as R); (ii) the variable to be bypassed cannot have more than one child variable.
Let X be the variable to bypass and $x \in X$ a variable state. Let Y be the child of X and $y \in Y$ a variable state. Then the NPT of Y has to be updated in order to 'inherit' the properties of X. Such an update is realized according to the following rule

$$Pr'(y \mid pa(X), pa(Y)) = \sum_{x \in X} Pr(y \mid x, pa(Y)) Pr(x \mid pa(X)), \tag{B.3}$$

where, $pa(X)$ denotes the parents of X and $pa(Y)$ the parents of Y apart from X.

B.4 Bypassing intermediary variables in the argument of credibility of testimonial evidence

In the context of evaluating the report provided by a human source, let a_2^{hum} and b_2^{hum} be probabilities required to subsume all three attributes constituting the argument of credibility (observational sensitivity, objectivity, and veracity) into a single reasoning stage. The inferential force that a report R exerts upon E is obtained by using Equations (B.1) and (B.2)

$$V_{r|e} = \frac{a_1 + \frac{b_2}{(a_2-b_2)} + \frac{b_3}{(a_2-b_2)(a_3-b_3)}}{b_2 + \frac{b_2}{(a_2-b_2)} + \frac{b_3}{(a_2-b_2)(a_3-b_3)}} = \frac{a_1^{\text{hum}}}{b_2^{\text{hum}}}.$$

The conditional probabilities a_2^{hum} and b_2^{hum} can be simplified to

$$\begin{aligned} a_2^{\text{hum}} &= a_1(a_2-b_2)(a_3-b_3) + b_2(a_3-b_3) + b_3, \\ b_2^{\text{hum}} &= b_1(a_2-b_2)(a_3-b_3) + b_2(a_3-b_3) + b_3. \end{aligned} \tag{B.4}$$

This reduction of chain-length by bypassing variables S (observational sensitivity) and O (objectivity) allows us to compare testimonial evidence to tangible evidence (e.g., physical sensors) by letting them perform the same task [125]. This makes it possible to classify an item of human testimony according to Bernoulli's cases and sub-cases.

B.5 Bypassing the intermediary variables G and F

Start by bypassing variable F by writing

$$\begin{aligned} Pr(e \mid h) &= Pr(e \mid f)Pr(f \mid h) + Pr(e \mid \overline{f})Pr(\overline{f} \mid h) = a_1, \\ Pr(e \mid \overline{h}) &= Pr(e \mid f)Pr(f \mid h) + Pr(e \mid \overline{f})Pr(\overline{f} \mid h) = b_1. \end{aligned}$$

Assume that $Pr(e \mid f) = 1$ and $Pr(e \mid \overline{f}) = \gamma$. Thus, we can write

$$\begin{aligned} Pr(e \mid h) &= Pr(f \mid h) + \gamma Pr(\overline{f} \mid h), \\ Pr(e \mid \overline{h}) &= Pr(f \mid \overline{h}) + \gamma Pr(\overline{f} \mid \overline{h}). \end{aligned}$$

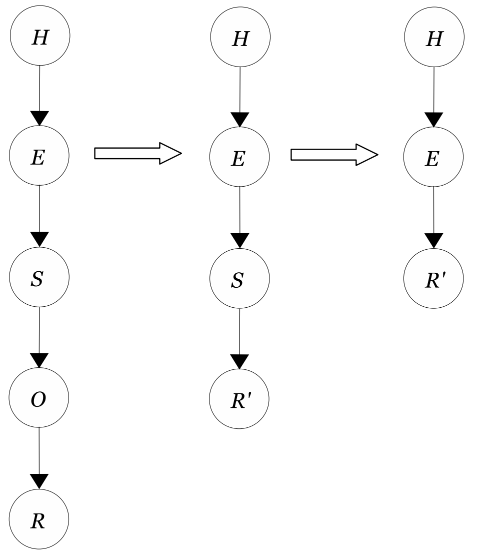


Figure. B.1 – Bypassing the intermediary variables S and O in the argument of credibility of testimonial evidence.

Next, extend the conversation to variable G to obtain

$$\begin{aligned}
Pr(e \mid h) =& Pr(f \mid g,h)Pr(g) + Pr(f \mid \overline{g},h)Pr(\overline{g}) + \gamma[Pr(\overline{f} \mid g,h)Pr(g) + Pr(\overline{f} \mid \overline{g},h)Pr(\overline{g})],\\
Pr(e \mid \overline{h}) =& Pr(f \mid g,\overline{h})Pr(g) + Pr(f \mid \overline{g},\overline{h})Pr(\overline{g}) + \gamma[Pr(\overline{f} \mid g,\overline{h})Pr(g) + Pr(\overline{f} \mid \overline{g},\overline{h})Pr(\overline{g})].
\end{aligned}$$

It is common in forensic literature to denote $Pr(g)$ more shortly as r, which is also called the 'relevance term' [133]. $Pr(\overline{g})$ is, therefore, $(1-r)$. Logical assignments of 1 and 0 apply to, respectively, $Pr(f \mid g,h)$ and $Pr(f \mid \overline{g},h)$ for the case where only one offender is involved [56]. The complementary events thus have probabilities $Pr(\overline{f} \mid g,h) = 0$ and $Pr(\overline{f} \mid \overline{g},h) = 1$. For instance, $Pr(f \mid g,h)$ is the probability of the suspect being the source of the trace (f) given that the trace was left by the offender during the commission of the crime (g) and given that the suspect is the offender (h). The above assignment thus means that given g and h , the event f is taken to be certain. As a consequence, the expression for a_1 can be rewritten as

$$a_1 = r + (1-r)\gamma.$$

Further logical assignments are $Pr(f \mid g,\overline{h}) = 0$ and $Pr(\overline{f} \mid g,\overline{h}) = 1$ [56]. The term $Pr(f \mid \overline{g},\overline{h})$ is the probability that the suspect's trace is at the scene for reasons unrelated to the crime. This probability is commonly written p in forensic literature. Thus, $Pr(\overline{f} \mid \overline{g},\overline{h}) = 1-p$. For $Pr(e \mid \overline{h})$ one can then obtain

$$b_1 = p(1-r) + \gamma[r + (1-p)(1-r)].$$

The inferential force of the argument of relevance is therefore given by

$$V_{e'|h} = \frac{a_1}{b_1} = \frac{r + (1-r)\gamma}{p(1-r) + \gamma[r + (1-p)(1-r)]}.$$

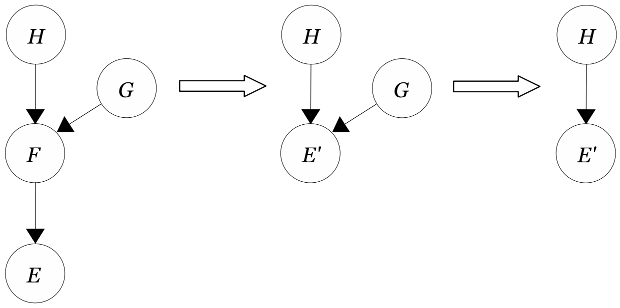


Figure. B.2 – Bypassing the intermediary variables of the BN for the evaluation of tangible evidence given crime-level propositions.

This corresponds therefore to type (a) argument of relevance. If, however, $r = 1$ then the argument of relevance is of type (c).

C Appendix: Graphical probabilistic analysis of the combination of items of evidence

C.1 Development of Equation (3.2)

In order to develop Equation (3.2) from Equation (3.1), an 'extension of the conversation' to the intermediate, unobserved variable E, is necessary:

$$LR_{E^*,E^{c*}} = \frac{Pr(E^*,E^{c*}|H)}{Pr(E^*,E^{c*}|H^c)} = \frac{Pr(E*,E^{c*}|E)Pr(E|H)+Pr(E*,E^{c*}|E^c)Pr(E^c|H)}{Pr(E*,E^{c*}|E)Pr(E|H^c)+Pr(E*,E^{c*}|E^c)Pr(E^c|H^c)} \; .$$

Then one can consider that the joint probability of the two reports $\{E^*,E^{c*}\}$, given E, is given by the product of the conditional probabilities of the individual reports, given E, because of their conditional independence as implied by Figure 3.1 (a). For $h_1 = Pr(E^* \mid E)$, $m_2 = Pr(E^{c*} \mid E)$, $f_1 = Pr(E^* \mid E^c)$ and $c_2 = Pr(E^{c*} \mid E^c)$, one thus obtains:

$$\begin{aligned} LR_{E^*,E^{c*}} &= \frac{\overbrace{Pr(E^*,E^{c*}|E)}^{h_1m_2}Pr(E|H)+\overbrace{Pr(E^*,E^{c*}|E^c)}^{f_1c_1}\overbrace{Pr(E^c|H)}^{1-Pr(E|H)}}{\underbrace{Pr(E^*,E^{c*}|E)}_{h_1m_2}Pr(E|H^c)+\underbrace{Pr(E*,E^{c*}|E^c)}_{f_1c_1}\underbrace{Pr(E^c|H^c)}_{1-Pr(E|H)}} \\ &= \frac{Pr(E|H)h_1m_2 - Pr(E|H)f_1c_2 + f_1c_2}{Pr(E|H^c)h_1m_2 - Pr(E|H^c)f_1c_2 + f_1c_2} = \frac{Pr(E|H)[\frac{h_1m_2}{f_1c_2}-1]+1}{Pr(E|H^c)[\frac{h_1m_2}{f_1c_2}-1]+1} = \frac{Pr(E|H)+[\frac{h_1m_2}{f_1c_2}-1]^{-1}}{Pr(E|H^c)+[\frac{h_1m_2}{f_1c_2}-1]^{-1}} \; . \end{aligned}$$

C.2 Development of $LR_{E_2=e_2|E_1=e_1}$

For a case in which the fingermark and the footwear mark are asymmetrically independent (i.e., whenever $a_2 \neq a'_2$; Table 3.5), the likelihood ratio for the footwear mark evidence (E_2) is conditioned on the fingermark evidence E_1:

$$LR_{E_2=e_2|E_1=e_1} = \frac{Pr(E_2=e_2|E_1=e_1,S=s,H_p)}{Pr(E_2=e_2|E_1=e_1,S=s,H_d)} \; . \tag{C.1}$$

When assessing uncertainty about E_2, one needs to account for uncertainty about F_2. Using notation introduced so far, the likelihood ratio thus develops as follows:

$$LR_{E_2=e_2|E_1=e_1} = \frac{\left(\begin{array}{c}\overbrace{Pr(E_2=e_2|F_2=f_2)}^{h_1}Pr(F_2=f_2|E_1=e_1,S=s,H_p)\\ +\underbrace{Pr(E_2=e_2|F_2=\bar{f}_2)}_{f_2}\underbrace{Pr(F_2=\bar{f}_2|E_1=e_1,S=s,H_p)}_{1-Pr(F_2=f_2|E_1=e_1,S=s,H_p)}\end{array}\right)}{\left(\begin{array}{c}\overbrace{Pr(E_2=e_2|F_2=f_2)}^{h_2}Pr(F_2=f_2|E_1=e_1,S=s,H_d)\\ +\underbrace{Pr(E_2=e_2|F_2=\bar{f}_2)}_{f_2}\underbrace{Pr(F_2=\bar{f}_2|E_1=e_1,S=s,H_d)}_{1-Pr(F_2=f_2|E_1=e_1,S=s,H_d)}\end{array}\right)}$$

$$= \frac{Pr(F_2=f_2|E_1=e_1,S=s,H_p)h_2-Pr(F_2=f_2|E_1=e_1,S=s,H_p)f_2+f_2}{Pr(F_2=f_2|E_1=e_1,S=s,H_d)h_2-Pr(F_2=f_2|E_1=e_1,S=s,H_d)f_2+f_2}$$

$$= \frac{Pr(F_2=f_2|E_1=e_1,S=s,H_p)+[\frac{h_2}{f_2}-1]^{-1}}{Pr(F_2=f_2|E_1=e_1,S=s,H_d)+[\frac{h_2}{f_2}-1]^{-1}}$$

Accounting for uncertainty about G_2, that is the relevance of the footwear mark, the numerator extends to:

$$\begin{aligned} Pr(F_2=f_2|E_1=e_1,S=s,H_p) = {} & Pr(F_2=f_2|E_1=e_1,S=s,G_2=g_2,H_p)\underbrace{Pr(G_2=g_2)}_{r_2} \\ & +Pr(F_2=f_2|E_1=e_1,S=s,G_2=\bar{g}_2,H_p)\underbrace{Pr(G_2=\bar{g}_2)}_{1-r_2}, \end{aligned}$$

where

$$\begin{aligned} Pr(F_2=f_2|E_1=e_1,S=s,G_2=g_2,H_p) = {} & \underbrace{Pr(F_2=f_2|F_1=f_1,G_2=g_2,H_p)}_{w}Pr(F_1=f_1|E_1=e_1,S=s,H_p) \\ & +\underbrace{Pr(F_2=f_2|F_1=\bar{f}_1,G_2=g_2,H_p)}_{w}\underbrace{Pr(F_1=\bar{f}_1|E_1=e_1,S=s,H_p)}_{1-Pr(F_1=f_1|E_1=e_1,S=s,H_p)} \\ = {} & w, \end{aligned}$$

and

$$\begin{aligned} Pr(F_2=f_2|E_1=e_1,S=s,G_2=\bar{g}_2,H_p) = {} & \underbrace{Pr(F_2=f_2|F_1=f_1,G_2=\bar{g}_2,H_p)}_{0}Pr(F_1=f_1|E_1=e_1,S=s,H_p) \\ & +\underbrace{Pr(F_2=f_2|F_1=\bar{f}_1,G_2=\bar{g}_2,H_p)}_{0}Pr(F_1=\bar{f}_1|E_1=e_1,S=s,H_p) \\ = {} & 0. \end{aligned}$$

Thus, $Pr(F_2=f_2|E_1=e_1,S=s,H_p)=wr_2$.

In the denominator one has:

$$\begin{aligned} Pr(F_2=f_2|E_1=e_1,S=s,H_d) = {} & Pr(F_2=f_2|E_1=e_1,S=s,G_2=g_2,H_d)\underbrace{Pr(G_2=g_2)}_{r_2} \\ & +Pr(F_2=f_2|E_1=e_1,S=s,G_2=\bar{g}_2,H_d)\underbrace{Pr(G_2=\bar{g}_2)}_{1-r_2} \end{aligned}$$

where

$$\begin{aligned}
Pr(F_2=f_2|E_1=e_1,S=s,G_2=g_2,H_d) &= \underbrace{Pr(F_2=f_2|F_1=f_1,G_2=g_2,H_d)}_{0}Pr(F_1=f_1|E_1=e_1,S=s,H_d) \\
&\quad +\underbrace{Pr(F_2=f_2|F_1=\bar{f}_1,G_2=g_2,H_d)}_{0}Pr(F_1=\bar{f}_1|E_1=e_1,S=s,H_d) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
Pr(F_2=f_2|E_1=e_1,S=s,G_2=\bar{g}_2,H_d) &= \underbrace{Pr(F_2=f_2|F_1=f_1,G_2=\bar{g}_2,H_d)}_{a_2}Pr(F_1=f_1|E_1=e_1,S=s,H_d) \\
&\quad +\underbrace{Pr(F_2=f_2|F_1=\bar{f}_1,G_2=\bar{g}_2,H_d)}_{a_2'}\underbrace{Pr(F_1=\bar{f}_1|E_1=e_1,S=s,H_d)}_{1-Pr(F_1=f_1|E_1=e_1,S=s,H_p)} \\
&= Pr(F_1=f_1|E_1=e_1,S=s,H_d)(a_2-a_2')+a_2'.
\end{aligned}$$

From here, a further development of $Pr(F_1=f_1|E_1=e_1,S=s,H_d)$ is needed. Consider thus Bayes' theorem:

$$Pr(F_1=f_1|E_1=e_1,S=s,H_d) = \frac{Pr(E_1=e_1|S=s,F_1=f_1)Pr(F_1=f_1|H_d)}{\begin{pmatrix} Pr(E_1=e_1|S=s,F_1=f_1)Pr(F_1=f_1|H_d) \\ +Pr(E_1=e_1|S=s,F_1=\bar{f}_1)Pr(F_1=\bar{f}_1|H_d) \end{pmatrix}}$$

where

$$\begin{aligned}
Pr(E_1=e_1|S=s,F_1=f_1)Pr(F_1=f_1|H_d) &= \underbrace{Pr(E_1=e_1|S=s,F_1=f_1)}_{1} \\
&\quad \times [\underbrace{Pr(F_1=f_1|G_1=g_1,H_d)}_{0}Pr(G_1=g_1) \\
&\quad +\underbrace{Pr(F_1=f_1|G_1=\bar{g}_1,H_d)}_{a_1}\underbrace{Pr(G_1=\bar{g}_1)}_{1-r_1}] \\
&= a_1(1-r_1)
\end{aligned}$$

and

$$\begin{aligned}
Pr(E_1=e_1|S=s,F_1=\bar{f}_1)Pr(F_1=\bar{f}_1|H_d) &= [\underbrace{Pr(E_1=e_1|U=u,F_1=\bar{f}_1)}_{1}\underbrace{Pr(U=u)}_{\gamma} \\
&\quad +\underbrace{Pr(E_1=e_1|U=\bar{u},F_1=\bar{f}_1)}_{0}Pr(U=\bar{u})] \\
&\quad \times [\underbrace{Pr(F_1=\bar{f}_1|G_1=g_1,H_d)}_{1}\underbrace{Pr(G_1=g_1)}_{r_1} \\
&\quad +\underbrace{Pr(F_1=\bar{f}_1|G_1=\bar{g}_1,H_d)}_{1-a_1}\underbrace{Pr(G_1=\bar{g}_1)}_{1-r_1}] \\
&= \gamma\{r_1+(1-a_1)(1-r_1)\}.
\end{aligned}$$

One thus has:

$$Pr(F_1 = f_1|E_1 = e_1, S = s, H_d) = \frac{a_1(1-r_1)}{a_1(1-r_1)+\gamma[r_1+(1-a_1)(1-r_1)]} \ .$$

The likelihood ratio $LR_{E_2=e_2|E_1=e_1}$ initially defined in Equation (C.1) thus becomes:

$$LR_{E_2=e_2|E_1=e_1} = \frac{wr+[h_2/f_2-1]^{-1}}{\left\{\frac{a_1(1-r_1)}{a_1(1-r_1)+\gamma\{r_1+(1-a_1)(1-r_1)\}}(a_2-a_2')+a_2'\right\}(1-r_2)+[h_2/f_2-1]^{-1}} \ .$$

Note that when $a_2 = a_2'$, one obtains

$$LR_{E_2=e_2|E_1=e_1,S=s} \quad = \quad \frac{wr+[h_2/f_2-1]^{-1}}{a_2'(1-r_2)+[h_2/f_2-1]^{-1}}$$

which is equivalent to a result of an algebraic approach previously presented in [46].

D Appendix: Investigating evidential phenomena in combined evidence

D.1 Derivation of argument structures of (a) and (b') from (b)

Let $R_1 = \{r_1, \bar{r}_1\}$ and $R_2 = \{r_2, \bar{r}_2\}$ represent the two reports, $E_1 = \{e_1, \bar{e}_1\}$ and $E_2 = \{e_2, \bar{e}_2\}$ the two events corresponding to each report, and $H = \{h, \bar{h}\}$ the hypotheses of interest. Assume that we received two reports $r_1 \in R_1$ and $r_2 \in R_2$. Then the inferential force provided by both reports upon H, following the argument structure shown in Figure 4.1 (b), is given by

$$\frac{Pr(r_1, r_2 \mid h)}{Pr(r_1, r_2 \mid \bar{h})} = \frac{Pr(r_1 \mid h)}{Pr(r_1 \mid \bar{h})} \frac{Pr(r_2 \mid r_1, h)}{Pr(r_2 \mid r_1, \bar{h})},$$

where

$$\frac{Pr(r_1 \mid h)}{Pr(r_1 \mid \bar{h})} = \frac{Pr(r_1 \mid e_1)Pr(e_1 \mid h) + Pr(r_1 \mid \bar{e}_1)\overbrace{Pr(\bar{e}_1 \mid h)}^{1-Pr(e_1|h)}}{Pr(r_1 \mid e_1)Pr(e_1 \mid \bar{h}) + Pr(r_1 \mid \bar{e}_1)\underbrace{Pr(\bar{e}_1 \mid \bar{h})}_{1-Pr(e_1|\bar{h})}} = \frac{Pr(e_1 \mid h)[Pr(r_1 \mid e_1) - Pr(r_1 \mid \bar{e}_1)] + Pr(r_1 \mid \bar{e}_1)}{Pr(e_1 \mid \bar{h})[Pr(r_1 \mid e_1) - Pr(r_1 \mid \bar{e}_1)] + Pr(r_1 \mid \bar{e}_1)}$$

$$= \frac{Pr(e_1 \mid h) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}}{Pr(e_1 \mid \bar{h}) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}}$$

and similarly

$$\frac{Pr(r_2 \mid r_1, h)}{Pr(r_2 \mid r_1, \bar{h})} = \frac{Pr(r_2 \mid e_2)Pr(e_2 \mid r_1, h) + Pr(r_2 \mid \bar{e}_2)\overbrace{Pr(\bar{e}_2 \mid r_1, h)}^{1-Pr(e_2|r_1,h)}}{Pr(r_2 \mid e_2)Pr(e_2 \mid r_1, \bar{h}) + Pr(r_2 \mid \bar{e}_2)\underbrace{Pr(\bar{e}_2 \mid r_1, \bar{h})}_{Pr(e_2|r_1,\bar{h})}}$$

$$= \frac{Pr(e_2 \mid r_1, h)[Pr(r_2 \mid e_2) - Pr(r_2 \mid \bar{e}_2)] + Pr(r_2 \mid \bar{e}_2)}{Pr(e_2 \mid r_1, \bar{h})[Pr(r_2 \mid e_2) - Pr(r_2 \mid \bar{e}_2)] + Pr(r_2 \mid \bar{e}_2)}$$

$$= \frac{Pr(e_2 \mid r_1, h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_2 \mid r_1, \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}.$$

The probabilities $Pr(e_2 \mid r_1, h)$ and $Pr(e_2 \mid r_1, \bar{h})$ can further be developed so that

$$\frac{Pr(r_2 \mid r_1, h)}{Pr(r_2 \mid r_1, \bar{h})} = \frac{Pr(e_2 \mid e_1, h)Pr(e_1 \mid r_1, h) + Pr(e_2 \mid \bar{e}_1, h)\overbrace{Pr(\bar{e}_1 \mid r_1, h)}^{1-Pr(e_1 \mid r_1, h)} + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_2 \mid e_1, \bar{h})Pr(e_1 \mid r_1, \bar{h}) + Pr(e_2 \mid \bar{e}_1, \bar{h})\underbrace{Pr(\bar{e}_1 \mid r_1, \bar{h})}_{1-Pr(e_1 \mid r_1, \bar{h})} + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}.$$

Finally, one obtains for the inferential force of r_2

$$\frac{Pr(r_2 \mid r_1, h)}{Pr(r_2 \mid r_1, \bar{h})} = \frac{Pr(e_1 \mid r_1, h)[Pr(e_2 \mid e_1, h) - Pr(e_2 \mid \bar{e}_1, h)] + Pr(e_2 \mid \bar{e}_1, h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_1 \mid r_1, \bar{h})[Pr(e_2 \mid e_1, \bar{h}) - Pr(e_2 \mid \bar{e}_1, \bar{h})] + Pr(e_2 \mid \bar{e}_1, \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}. \tag{D.1}$$

Derivation of (a) from (b) To say that the events E_1 and E_2 are naturally redundant is equal to saying that $E_1 = E_2$. Therefore, the inferential force of r_2 (D.1) becomes

$$\frac{Pr(e_1 \mid r_1, h)[Pr(e_2 \mid e_1, h) - Pr(e_2 \mid \bar{e}_1, h)] + Pr(e_2 \mid \bar{e}_1, h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_1 \mid r_1, \bar{h})[Pr(e_2 \mid e_1, \bar{h}) - Pr(e_2 \mid \bar{e}_1, \bar{h})] + Pr(e_2 \mid \bar{e}_1, \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}} = \frac{Pr(e_1 \mid r_1, h) + [\frac{Pr(r_2|e_1)}{Pr(r_2|\bar{e}_1)} - 1]^{-1}}{Pr(e_1 \mid r_1, \bar{h}) + [\frac{Pr(r_2|e_1)}{Pr(r_2|\bar{e}_1)} - 1]^{-1}}.$$

As a result, the inferential force that the two reports exert on H is given by

$$\frac{Pr(r_1, r_2 \mid h)}{Pr(r_1, r_2 \mid \bar{h})} = \frac{Pr(e_1 \mid h) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}}{Pr(e_1 \mid \bar{h}) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}} \frac{Pr(e_1 \mid r_1, h) + [\frac{Pr(r_2|e_1)}{Pr(r_2|\bar{e}_1)} - 1]^{-1}}{Pr(e_1 \mid r_1, \bar{h}) + [\frac{Pr(r_2|e_1)}{Pr(r_2|\bar{e}_1)} - 1]^{-1}},$$

which is identical to the inferential force following the argument structure involving two reports relating to a single event as depicted in Figure 4.1 (a).

Derivation of (b') from (b) If E_1 and E_2 are conditionally independent given H, then the inferential force of r_2 (D.1) becomes

$$\frac{Pr(e_1 \mid r_1, h)[Pr(e_2 \mid e_1, h) - Pr(e_2 \mid \bar{e}_1, h)] + Pr(e_2 \mid \bar{e}_1, h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_1 \mid r_1, \bar{h})[Pr(e_2 \mid e_1, \bar{h}) - Pr(e_2 \mid \bar{e}_1, \bar{h})] + Pr(e_2 \mid \bar{e}_1, \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}} = \frac{Pr(e_2 \mid h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_2 \mid \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}.$$

The joint inferential force of both reports is therefore given by

$$\frac{Pr(r_1, r_2 \mid h)}{Pr(r_1, r_2 \mid \bar{h})} = \frac{Pr(e_1 \mid h) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}}{Pr(e_1 \mid \bar{h}) + [\frac{Pr(r_1|e_1)}{Pr(r_1|\bar{e}_1)} - 1]^{-1}} \frac{Pr(e_2 \mid h) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}}{Pr(e_2 \mid \bar{h}) + [\frac{Pr(r_2|e_2)}{Pr(r_2|\bar{e}_2)} - 1]^{-1}},$$

which corresponds to the inferential force of the argument structure depicted in Figure 4.1 (b').

D.2 Interpretation of $R_{e_2|e_1}$

Table. D.1 – Relationship between the measure $R_{e_2|e_1}$ and the inferential force of e_2

$R_{e_2\|e_1}$	e_2 supports h over $\bar{h}$ ($W(h:e_2)>1$)	e_2 supports $\bar{h}$ over h ($W(h:e_2)<1$)	Type of interaction
$R_{e_2\|e_1}<0$	$W(h:e_2\mid e_1)>W(h:e_2)$	$W(h:e_2\mid e_1)<W(h:e_2)$	Synergy
$R_{e_2\|e_1}=0$	$W(h:e_2\mid e_1)=W(h:e_2)$	$W(h:e_2\mid e_1)=W(h:e_2)$	Cond. independence
$1>R_{e_2\|e_1}>0$	$W(h:e_2)>W(h:e_2\mid e_1)>1$	$W(h:e_2)<W(h:e_2\mid e_1)<1$	Redundance
$R_{e_2\|e_1}=1$	$W(h:e_2\mid e_1)=1$	$W(h:e_2\mid e_1)=1$	
$R_{e_2\|e_1}>1$	$W(h:e_2\mid e_1)<1$	$W(h:e_2\mid e_1)>1$	Directional change

D.3 Derivations for types of inferential interaction in terms of weight of evidence

In the case where both items of evidence provide weight in favor of h ($W(h:e_1)>0$ and $W(h:e_2)>0$), the following derivations apply:

Synergy The events e_1 and e_2 are said to be inferentially synergistic given the hypotheses H if the inequality $W(h:e_2\mid e_1)>W(h:e_2)$ holds. A reformulation of this inequality into $0>W(h:e_2)-W(h:e_2\mid e_1)$, and by applying Equation 4.8 leads to the conclusion that the joint weight of evidence of e_1 an e_2 is larger than the sum of each weight, that is, $W(h:e_1,e_2)>W(h:e_1)+W(h:e_2)$.

Conditional independence The events e_1 and e_2 are said to be conditionally independent given the hypotheses H if $Pr(e_2\mid e_1,H)=Pr(e_2\mid H)$. In other words, one has $W(h:e_2\mid e_1)=W(h:e_2)$. Hence, $W(h:e_2)-W(h:e_2\mid e_1)=0$, and from Equation 4.8 it follows that $W(h:e_1,e_2)=W(h:e_1)+W(h:e_2)$.

Redundance An event e_2 is said to be redundant given e_1, if $0<W(h:e_2\mid e_1)<W(h:e_2)$ holds. Hence, Equation 4.8 states that the lower boundary of the joint weight of evidence in favor of h provided by e_2 and e_1 is $W(h:e_1)<W(h:e_1,e_2)$, and that the upper boundary is $W(h:e_1,e_2)<W(h:e_1)+W(h:e_2)$ (i.e., $W(h:e_1)<W(h:e_1,e_2)<W(h:e_1)+W(h:e_2)$). In other words, the domain of values associated with redundance is bounded by the values of complete redundance ($W(h:e_1,e_2)=W(h:e_1)$) and conditional dependence ($W(h:e_1,e_2)=W(h:e_1)+W(h:e_2)$).

Complete redundance Event e_2 is said to be completely redundant given e_1 and the hypotheses H, if $W(h:e_2\mid e_1)=0$ holds. In such cases Equation 4.8 reduces to $W(h:e_1,e_2)=W(h:e_1)$.

Directional change The occurrence of event e_1 is said to induce a directional change in e_2 given hypotheses H (i.e., event e_2 supports $\bar{h}$ once e_1 is known to have occurred), if $W(h:e_2\mid e_1)<0$ holds. From Equation 4.8 it follows that $W(h:e_1,e_2)<W(h:e_1)$. The derivations for cases where $W(h:e_1)<0$ and $W(h:e_2)<0$ can be obtained analogously.

D.4 Values of $ia(a_0 : \mathbf{D})$ and their implication on the presence of a directional change or synergy

D.4.1 Proving that $ia(a_0 : \mathbf{D}) > 1$ implies $W(a_0 : \mathbf{D}^{dc}) \neq 0$.

The statement that in all cases, where $W(a_0 : d_{i,j}) > 0, d_{i,j} \in \mathbf{D}^m$, an interaction value larger than one implies the presence of an effect of directional change ($W(a_0 : \mathbf{D}^{dc}) \neq 0$) is to be proven here. The proof is established by exploiting the fact that a value of $ia[W(a_0 : D)] > 1$ being equivalent to $W_\perp(a_0 : \mathbf{D}^m) < -W(a_0 : \mathbf{D}^m)$, implies that $W(a_0 : \mathbf{D}^{dc}) \neq 0$. More formally we have,

$$ia(a_0 : \mathbf{D}) > 1 \Longleftrightarrow W_\perp(a_0 : \mathbf{D}^m) < -W(a_0 : \mathbf{D}^m), \text{ and} \tag{D.2}$$

$$W_\perp(a_0 : \mathbf{D}^m) < -W(a_0 : \mathbf{D}^m) \Longrightarrow W(a_0 : \mathbf{D}^{dc}) \neq 0. \tag{D.3}$$

The equivalence stated in D.2 follows directly from Equation 4.10. This equation is used again as a starting point for the demonstration of statement D.3, notably

$$ia(a_0 : \mathbf{D}) = \frac{W_\perp(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^m)}{W_\perp(a_0 : \mathbf{D}^m)} > 1,$$

from where it follows that

$$-W(a_0 : \mathbf{D}^m) > 0$$
$$-W(a_0 : \mathbf{D}^s) - W(a_0 : \mathbf{D}^{pr}) - W(a_0 : \mathbf{D}^{dc}) > 0.$$

Finally, one obtains

$$-W(a_0 : \mathbf{D}^{dc}) > W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}).$$

For this inequality to hold, it is necessary that $W(a_0 : \mathbf{D}^{dc}) \neq 0$ since $W(a_0 : \mathbf{D}^s) \leq 0$, $W(a_0 : \mathbf{D}^{pr}) \leq 0$, and $W(a_0 : \mathbf{D}^{dc}) \geq 0$ by definition (see D.3). This result allows Statement D.2 and D.3 to be rewritten in a more informative manner

$$ia(a_0 : \mathbf{D}) > 1 \Longleftrightarrow -W(a_0 : \mathbf{D}^{dc}) > W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}) \Longrightarrow W(a_0 : \mathbf{D}^{dc}) \neq 0.$$

By applying the same reasoning to situations, in which $W(a_0 : d_{i,j}) < 0, d_{i,j} \in \mathbf{D}^m$ one obtains

$$ia(a_0 : \mathbf{D}) < 1 \Longleftrightarrow -W(a_0 : \mathbf{D}^{dc}) < W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}) \Longrightarrow W(a_0 : \mathbf{D}^{dc}) \neq 0.$$

D.4.2 Proving that $ia(a_0 : \mathbf{D}) < 0$ implies $W(a_0 : \mathbf{D}^s) \neq 0$'.

In situations, in which $W(a_0 : d_{i,j}) > 0$, where $d_{i,j} \in \mathbf{D}^m$ a value of $ia(a_0 : \mathbf{D}) < 0$ indicates that $\mathbf{D}^s \neq 0$. More precisely, one has the relationships

$$ia(a_0 : \mathbf{D}) < 0 \Longleftrightarrow W_\perp(a_0 : \mathbf{D}^m) > W(a_0 : \mathbf{D}^m), \text{ and} \tag{D.4}$$

$$W_\perp(a_0 : \mathbf{D}^m) > W(a_0 : \mathbf{D}^m) \Longrightarrow W(a_0 : \mathbf{D}^s) \neq 0. \tag{D.5}$$

The equivalence in D.4 follows directly from Equation 4.10, which is used – just as in D.4.1 – to prove the implication in D.5. Notably,

$$ia(a_0 : \mathbf{D}) = \frac{W_\perp(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^m)}{W_\perp(a_0 : \mathbf{D}^m)} > 0,$$

where $W_{\perp}(a_0 : \mathbf{D}^m) \neq 0$. Thus,

$$\begin{aligned} W(a_0 : \mathbf{D}^m) &> W_{\perp}(a_0 : \mathbf{D}^m) \\ W(a_0 : \mathbf{D}^s) + W(a_0 : \mathbf{D}^{pr}) + W(a_0 : \mathbf{D}^{dc}) &> W_{\perp}(a_0 : \mathbf{D}^m), \end{aligned}$$

and finally one has

$$W(a_0 : \mathbf{D}^s) > W_{\perp}(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^{pr}) - W(a_0 : \mathbf{D}^{dc}).$$

For this inequality to hold, it is necessary that an effect of synergy is present ($W(a_0 : \mathbf{D}^s) \neq 0$) since $W_{\perp}(a_0 : \mathbf{D}^m)$, and $W(a_0 : \mathbf{D}^{pr})$ take values larger than zero each and $W(a_0 : \mathbf{D}^{dc})$ a value smaller than zero by definition (see D.3). Thus one can write

$$ia(a_0 : \mathbf{D}) < 0 \Longleftrightarrow W(a_0 : \mathbf{D}^s) > W_{\perp}(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^{pr}) - W(a_0 : \mathbf{D}^{dc}) \Longrightarrow W(a_0 : \mathbf{D}^s) \neq 0.$$

Applying an analogous reasoning to situations, in which $W(a_0 : d_{i,j}) < 0$, where $d_{i,j} \in \mathbf{D}^m$ leads to

$$ia(a_0 : \mathbf{D}) > 0 \Longleftrightarrow W(a_0 : \mathbf{D}^s) < W_{\perp}(a_0 : \mathbf{D}^m) - W(a_0 : \mathbf{D}^{pr}) - W(a_0 : \mathbf{D}^{dc}) \Longrightarrow W(a_0 : \mathbf{D}^s) \neq 0.$$

D.5 Development of Equation 4.15

D.5.1 Development of $Pr(a_0 \mid \mathbf{D}^{i-1}, a_1)$.

By using the Bayes' theorem $Pr(a_0 \mid \mathbf{D}^{i-1}, a_1)$ becomes

$$Pr(a_0 \mid \mathbf{D}^{i-1}, a_1) = \frac{Pr(\mathbf{D}^{i-1} \mid a_0)Pr(a_0 \mid a_1)}{Pr(\mathbf{D}^{i-1} \mid a_0)Pr(a_0 \mid a_1) + Pr(\mathbf{D}^{i-1} \mid \overline{a_0})Pr(\overline{a_0} \mid a_1)} = \frac{F(a_0 : \mathbf{D}^{i-1})Pr(a_0 \mid a_1)}{F(a_0 : \mathbf{D}^{i-1})Pr(a_0 \mid a_1) + Pr(\overline{a_0} \mid a_1)}$$

The probability $Pr(a_0 \mid \mathbf{D}^{i-1}, \overline{a_1})$ can be developed analogously.

D.6 Deriving values of $W_{\text{diss}}(a_k : \mathbf{D})$

Let $\mathbf{D}^m$ denote the subset of items providing positive weight and the subset $\mathbf{D}^{n-m}$ the subset of items providing negative weight. Equation 4.20 can be written as

$$\begin{aligned} W_{\text{diss}}(a_k : \mathbf{D}) &= \overbrace{|W(a_k : \mathbf{D}^m)| + |W(a_k : \mathbf{D}^{n-m})|}^{W_{\text{pot}}(a_k:\mathbf{D})} - \overbrace{|W(a_k : \mathbf{D}^m) + W(a_k : \mathbf{D}^{n-m})|}^{W_{\text{ex}}(a_k:\mathbf{D})} \\ &= W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m}) - |W(a_k : \mathbf{D}^m) + W(a_k : \mathbf{D}^{n-m})|. \end{aligned} \tag{D.6}$$

The expressed weight can take three possible values, that is, it can be (i) larger than zero, (ii) exactly zero, or (iii) smaller than zero. A case-by-case analysis of Equation D.6 produces Equation 4.21:
Case (i). A value of $W_{\text{ex}}(a_k : \mathbf{D}) > 0$ is obtained if $W(a_k : \mathbf{D}^m) < -W(a_k : \mathbf{D}^{n-m})$ and Equation D.6 becomes

$$W_{\text{diss}}(a_k : \mathbf{D}) = W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m}) - \big(-W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m})\big) = 2W(a_k : \mathbf{D}^m).$$

Case (ii). A value of $W_{\text{ex}}(a_k : \mathbf{D}) = 0$ is obtained if $W(a_k : \mathbf{D}^m) = -W(a_k : \mathbf{D}^{n-m})$ and Equation D.6 becomes

$$W_{\text{diss}}(a_k : \mathbf{D}) = W_{\text{pot}}(a_k : \mathbf{D}) = W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m}),$$

which is equivalent to $2W(a_k : \mathbf{D}^m)$ and $-2W(a_k : \mathbf{D}^{n-m})$ respectively.
Case (iii). A value of $W_{\text{ex}}(a_k : \mathbf{D}) > 0$ is obtained if $W(a_k : \mathbf{D}^m) > -W(a_k : \mathbf{D}^{n-m})$ and Equation D.6 becomes

$$W_{\text{diss}}(a_k : \mathbf{D}) = W(a_k : \mathbf{D}^m) - W(a_k : \mathbf{D}^{n-m}) - \big(W(a_k : \mathbf{D}^m) + W(a_k : \mathbf{D}^{n-m})\big) = -2W(a_k : \mathbf{D}^{n-m}).$$

D.6.1 Proving that $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$ in configuration (i)

Section 4.4.3 and D.6 showed that $W_{\text{diss}}(a_k : \mathbf{D})$ takes three different values depending on whether the expressed weight is larger, equal, or smaller than zero (i.e. case (i) to (iii)). The same applies for $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ and $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ each. Hence, there are 3^2 possible cases when considering the amount of $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ relative to that of $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$. The proof given here consists of showing that the relationship $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) > W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ is false in each of the 3^2 possible cases. The 3^2 cases are denoted by two roman numerals such as, for instance, case (i.ii), where the first numeral corresponds to the case of $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ and the second to the case of $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$.
Note also that in a line of reasoning the weight for any evidence $d_{\cdot,i}$ is bounded by the weight provided by the preceding event $d_{\cdot,i-1}$ in that line, that is, $W(a_0 : \overline{d_{\cdot,i-1}}) \leq W(a_0 : d_{\cdot,i}) \leq W(a_0 : \overline{d_{\cdot,i-1}})$ (see [125, 78]). Thus, applied to the present case, where the two lines of reasoning are independent given A_0, one has $W(a_0 : d_{1,1}) \leq W(a_0 : d_{1,0})$ and $-W(a_0 : d_{2,1}) \leq -W(a_0 : d_{2,0})$ (with equality only if $Pr(d_{1,1} \mid \overline{d_{1,0}}) = 0$ and $Pr(d_{2,1} \mid \overline{d_{2,0}}) = 0$ respectively).
In case (i) for $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ one has

$$W(a_0 : d_{1,1}) < -W(a_0 : d_{2,1}) \Rightarrow W_{diss}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,1}).$$

- *Case (i.i).* One has $W(a_0 : d_{1,0}) < -W(a_0 : d_{2,0}) \Rightarrow W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,0})$, and $W(a_0 : d_{1,1}) \leq W(a_0 : d_{1,0})$. Since $W_{diss}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,1}) \leq W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,0})$ one concludes that $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.

- *Case (i.ii).* One has $W(a_0 : d_{1,0}) = -W(a_0 : d_{2,0}) \Rightarrow W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = W(a_0 : d_{1,0}) - W(a_0 : d_{2,0}) = 2W(a_0 : d_{1,0})$. Thus, one concludes that $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$ for the same reasons as in case (i.i).

- *Case (i.iii).* One has $W(a_0 : d_{1,0}) > -W(a_0 : d_{2,0}) \Rightarrow W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = -2W(a_0 : d_{2,0})$, $-W(a_0 : d_{2,1}) \leq -W(a_0 : d_{2,0})$, and $W(a_0 : d_{1,1}) < -W(a_0 : d_{2,1})$. Moreover, $W(a_0 : d_{1,1}) < -W(a_0 : d_{2,0})$ applies. As a consequence, one has $W_{diss}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,1}) < W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = -2W(a_0 : d_{2,0})$ and concludes that $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) < 0$.

In case (ii) for $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ one has

$$W(a_0 : d_{1,1}) = -W(a_0 : d_{2,1}) \Rightarrow W_{diss}(a_0 : d_{1,1}, d_{2,1}) = W(a_0 : d_{1,1}) - W(a_0 : d_{2,1}) = 2W(a_0 : d_{1,1}) = -2W(a_0 : d_{2,1}).$$

- *Case (ii.i).* One has $W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,0})$ (see case (i.i)), which implies $W_{diss}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,1}) \leq W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,0})$. Hence, one has $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.

- *Case (ii.ii).* One has $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,0})$ (see case (i.ii)), which implies $W_{diss}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,1}) \leq W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = 2W(a_0 : d_{1,0})$. Thus, one has $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.

- *Case (ii.iii).* One has $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = -2W(a_0 : d_{2,0})$ (see case (i.iii)), which implies $W_{diss}(a_0 : d_{1,1}, d_{2,1}) = -2W(a_0 : d_{2,1}) \leq W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = -2W(a_0 : d_{2,0})$. Thus, one has $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.

In case (iii) for $W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1})$ one has

$$W(a_0 : d_{1,1}) > -W(a_0 : d_{2,1}) \Rightarrow W_{diss}(a_0 : d_{1,1}, d_{2,1}) = -2W(a_0 : d_{2,1}).$$

- *Case (iii.i).* One has $W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = 2W(a_0 : d_{1,0})$ (see case (i.i)). Given that $W(a_0 : d_{1,1}) \leq W(a_0, d_{1,0})$ and $W(a_0 : d_{1,1}) > -W(a_0 : d_{2,1})$, it follows that $-W(a_0 : d_{2,1}) < W(a_0, d_{1,0})$. Hence, $W_{diss}(a_0 : d_{1,1}, d_{2,1}) < W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1})$ one has $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) < 0$.

- *Case (iii.ii).* For $W(a_0 : d_{1,0}) = -W(a_0 : d_{2,0})$ one has $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = -2W(a_0 : d_{2,0})$. Since, $-W(a_0 : d_{2,1}) \leq -W(a_0 : d_{2,0})$ it follows that $W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) = -2W(a_0 : d_{2,1}) \leq W_{\text{diss}}(a_0 : d_{1,0}, d_{2,0}) = -2W(a_0 : d_{2,0})$ and one concludes $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.
- *Case (iii.iii).* One has $W_{\text{diss}_0}(a_0 : \mathbf{D}_1) = -2W(a_0 : d_{2,0})$ (see case (i.iii)). For the same reasons as in case (iii.ii) one concludes that $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$.

Note that in cases (i.iii) and (iii.i) the relationship $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) < 0$ was obtained and $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) \leq 0$ the in all the remaining cases. Thus, in situations of configuration (i) the assertion $W_{\text{diss}_1}(a_0 : d_{1,1}, d_{2,1}) = W_{\text{diss}}(a_0 : d_{1,1}, d_{2,1}) - W_{\text{diss}_0}(a_0 : d_{1,1}, d_{2,1}) \not> 0$ holds.

E Appendix: *State of Texas vs Josiah Sutton*

E.1 Computationally optimized BN for contributor scenarios

The BN for the scenarios of specimen $E_{1'}$ is outlined first, followed by the BN for the specimens E_3 and E_4.

BN for scenarios of $E_{1'}$. The nodes `H` and `N1` remain unchanged. Table E.1 outlines the node definitions from top to bottom and from left to right in BN shown in Figure E.1. Nodes that connect to other BNs are colored in light grey and their names are written in black (i.e., `P1inE1` to `P5inE1`, and `H1E1` to `H4E1`, and `P1E1` to `P3E1`). Node `xorS1toS8` is an evidence node and `H` a query node. Two things happen by instantiating `xorS1toS8`. First, mutual exclusiveness is established among $S_{1\leq s\leq 8,1'} \in \mathbf{S}_{1'}$ so that $Pr(\mathbf{S}_{1'} \mid \mathbf{I}) = 1$. Second, the nodes for the possible contributors are `P1inE1` to `P5inE1` d-connected to `H` through the mutually exclusive scenarios of $\mathbf{S}_1$.

BN for scenarios of E_3 and E_4. The nodes `H`, `W`, `N3` and `N4` are the same as those discussed in Section 5.4.5. Table E.2 outlines the node definitions from top to bottom and from left to right in BN shown in Figure E.2. The BNs for the scenarios of specimens E_3 and E_4 are essentially the same. Therefore, only the node definitions and the BN for E_3 are shown. The difference is refined to the nodes `S1to43Wnot3` to `S9to123Wnot3` and `S13to163Wnot3`, and `S1to44Wnot2` to `S9to124Wnot2` and `S13to164Wnot2` as shown in the table. Again by instantiating `true` in `xorS1to163`, the mutual exclusiveness among the scenarios is established and the hypothesis node is d-connected to the nodes representing the possible contributors.

In order to assemble the BNs for the contributor scenarios of all the specimens one has to merge nodes that are identical. These are `H`, `HP1` to `HP4`, `W`, `S11S51` to `S41S81`, and `P1Ss1` to `P3Ss1`.

Table. E.1 – Node definitions for the BN of contributor scenarios of $E_{1'}$

Node	Definition	Purpose
P1inE1 to P5inE1	Boolean, uniform distribution	Setting up possible contributors $\mathbf{P}_{1'}$
P34notinE1 to P12notinE1	e.g., $\neg$P3inE1 $\wedge$ $\neg$P4inE1	Exclude possible contributors
P12inE1 to P34inE1	e.g., P1inE1 $\wedge$ P2inE1	Include possible contributors
P12E1 to P34E1	e.g., P12inE1 $\wedge$ P34notinE1	Actual contributors among $P_1,..,P_4$
P12isE1 to P34isE1	e.g., P12E1 $\wedge$ $\neg$P5inE1	Exclude possible contributor P_5
P125isE1 to P345isE1	e.g., P12E1 $\wedge$ P5inE1	Include possible contributor P_5
S11	P12isE1 $\wedge$ H=1	
S21	P13isE1 $\wedge$ H=2	
S31	P23isE1 $\wedge$ H=3	Account for H
S41	P34isE1 $\wedge$ H=4	
S51 to S81	cf., S11 to S41	
xorS11S21	S11 $\veebar$ S21	Setting up m.e.* between $S_{1,1'}$ and $S_{2,1'}$
xorS31S41	S31 $\veebar$ S41	Setting up m.e.* between $S_{3,1'}$ and $S_{3,1'}$
xorS51S61	S51 $\veebar$ S61	Setting up m.e.* between $S_{5,1'}$ and $S_{6,1'}$
xorS71S81	S71 $\veebar$ S81	Setting up m.e.* between $S_{7,1'}$ and $S_{8,1'}$
xorS11toS41	xorS11S21 $\veebar$ xorS31S41	Setting up m.e.* among $S_{1\leq s\leq 4,1'}$
xorS51toS81	xorS51S61 $\veebar$ xorS71S81	Setting up m.e.* among $S_{5\leq s\leq 8,1'}$
N1is2	xorS1toS41 $\wedge$ N1 = 2	Establish $N_{1'} = 2$ for $S_{1\leq s\leq 4,1'}$
N1is3	xorS5toS81 $\wedge$ N1 = 2	Establish $N_{1'} = 3$ for $S_{5\leq s\leq 8,1'}$
xorS1toS8	N1is2 $\veebar$ N1is3	Setting up m.e.* among $S_{1\leq s\leq 8,1'}$
S11S51	S11 $\vee$ S51	
S21S61	S21 $\vee$ S61	
S31S71	S31 $\vee$ S71	
S41S81	S41 $\vee$ S81	Used to condition $S_{s,i}, i \in \{3,4\}$ on $S_{s,1'}$
P1Ss1	S11S51$\vee$ S21S61	
P2Ss1	S11S51$\vee$ S31S71	
P3Ss1	S21S61$\vee$ S31S71 $\vee$ S41S81	

*mutual exclusiveness

Table. E.2 – Node definitions for the BN of contributor scenarios of E_3 and E_4

Node	Definition	Purpose
`P1inE3` to `P6inE3`	Boolean, uniform distribution	Setting up possible contributors $\mathbf{P}_3$
`P12inE3` to `P34inE3`	e.g., `P1inE3` ∧ `P2inE3`	Include possible contributors
`P34notinE3` to `P12notinE3`	e.g., ¬`P3inE3` ∧ ¬`P4inE3`	Exclude possible contributors
`P12E3` to `P34E3`	e.g., `P12inE3` ∧ `P34notinE3`	Establish actual contributors among P_1 to P_4
`P234notinE3`	`P34notinE3` ∧ ¬`P2inE3`	Exclude possible contributors
`P134notinE3`	`P34notinE3` ∧ ¬`P1inE3`	
`P124notinE3`	`P12notinE3` ∧ ¬`P4inE3`	
`P124notinE3`	`P12notinE3` ∧ ¬`P3inE3`	
`P1E3` to `P4E3`	e.g. `P234notinE3` ∧ `P1inE3`	Establish actual contributor P_1 to P_4
`P56notinE3`	¬`P5inE3` ∧¬ `P6inE3`	Exclude possible contributors P_5 and P_6
`P5not6inE3`	`P5inE3` ∧¬ `P6inE3`	Include P_5 and exclude P_6
`P6not5inE3`	¬`P5inE3` ∧ `P6inE3`	Include P_6 and exclude P_5
`P12isE3` to `P46isE3`	e.g., `P12E3` ∧ `P56notinE3`	Establish actual contributors among $\mathbf{P}_3$
`P12isE3E1` to `P34isE3E1`	e.g., `P12isE3` ∧ `S11S51`	Account for scenarios $\mathbf{S}_{1'}$
`P1isE3E1` to `P46isE3E1`*	e.g., `P1isE3` ∧ `P5Ss1`	
`HP1`	`H=1` ∨ `H=2`	Isolate assailant P_1
`HP2`	`H=1` ∨ `H=3`	Isolate assailant P_2
`HP3`	`H=2` ∨ `H=3` ∨ `H=4`	Isolate assailant P_3
`HP4`	`H=4`	Isolate assailant P_4
`S13` to `S43`	e.g., `P12isE3E1` ∧ `H = 1`	Account for H
`S53` to `S163`	e.g., `P1isE3E1` ∧ `HP1`	
`xorS1to23` to `xorS15to163`	e.g., `S13` ⊻ `S23`	Setting up m.e.
`xorS1to43` to `xorS13to163`	e.g., `xorS1to23` ⊻ `xorS3to43`	
`S1to43Wnot3` to `S9to123Wnot3`	e.g., `xorS1to43` ∧¬ `W = 3`	Establish $W \neq 3$ for $S_{1\leq s\leq 12,3}$
`S13to163Wnot3`	e.g., `xorS13to163` ∧ `W = 3`	Establish $W = 3$ for $S_{13\leq s\leq 16,3}$
For $\mathbf{S}_4$ one has		
`S1to44Wnot2` to `S9to124Wnot2`	e.g., `xorS1to44` ∧¬ `W = 2`	Establish $W \neq 2$ for $S_{1\leq s\leq 12,4}$
`S13to164Wnot2`	e.g., `xorS13to164` ∧ `W = 2`	Establish $W = 2$ for $S_{13\leq s\leq 16,4}$
`S1to43Nis2`, `S9to123Nis2`, and `S13to163Nis2`	e.g., `S1to43Wnot3` ∧ `N3 = 2`	Establish $N_3 = 2$ for $S_{1\leq s\leq 4,3}$ and $S_{9\leq s\leq 16,3}$
`S5to83Nis1`	`S5to83Wnot3` ∧ `N3 = 1`	Establish $N_3 = 2$ for $S_{5\leq s\leq 8,3}$
`xorS1to83` and `xorS9to163`	e.g., `S1to43Nis2` ⊻ `S9to123Nis2`	Setting up m.e.
`xorS1to163`	`xorS1to83` ⊻ `xorS9to163`	

*Note, that P_4 can only be a contributor if `S41S81` applies. Hence, there is no need to isolate P_4 in a single node as was done for P_1 with `P1Ss1`.

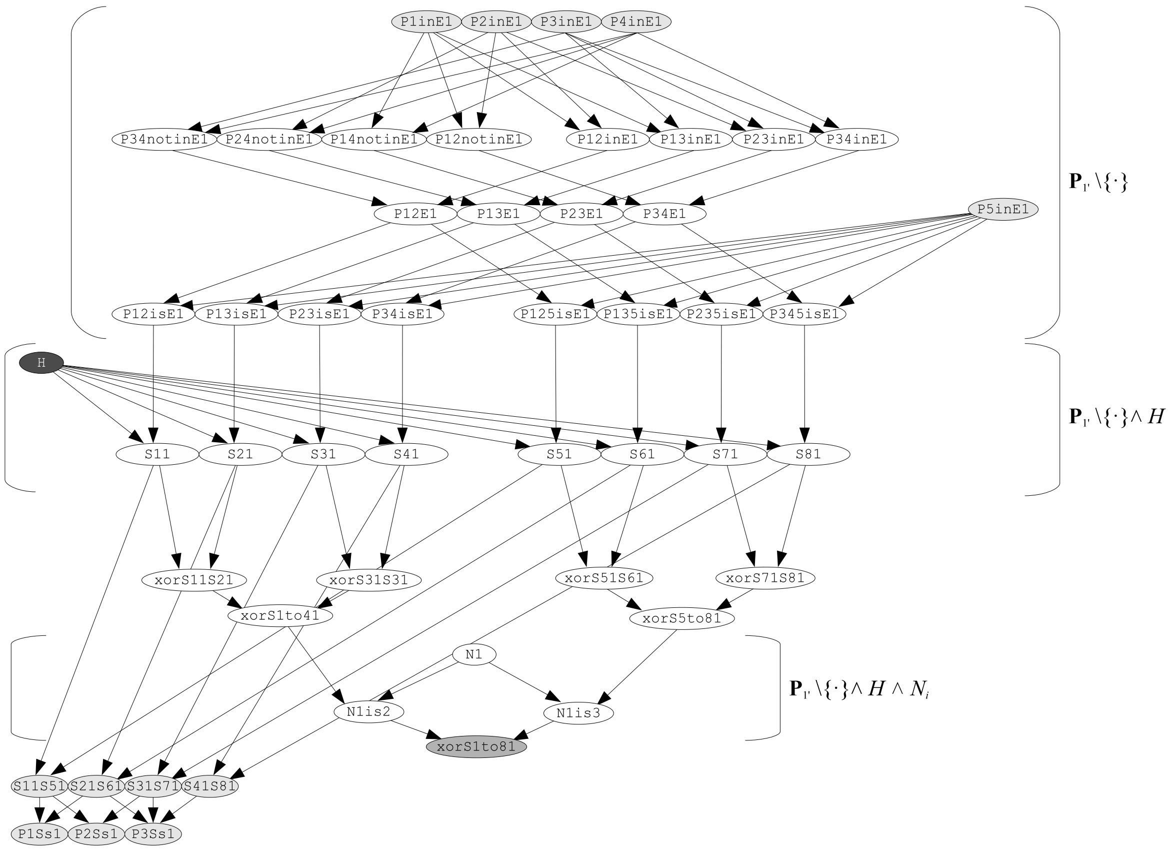


Figure. E.1 – BN for handling the scenarios of $\mathbf{S}_{1'}$

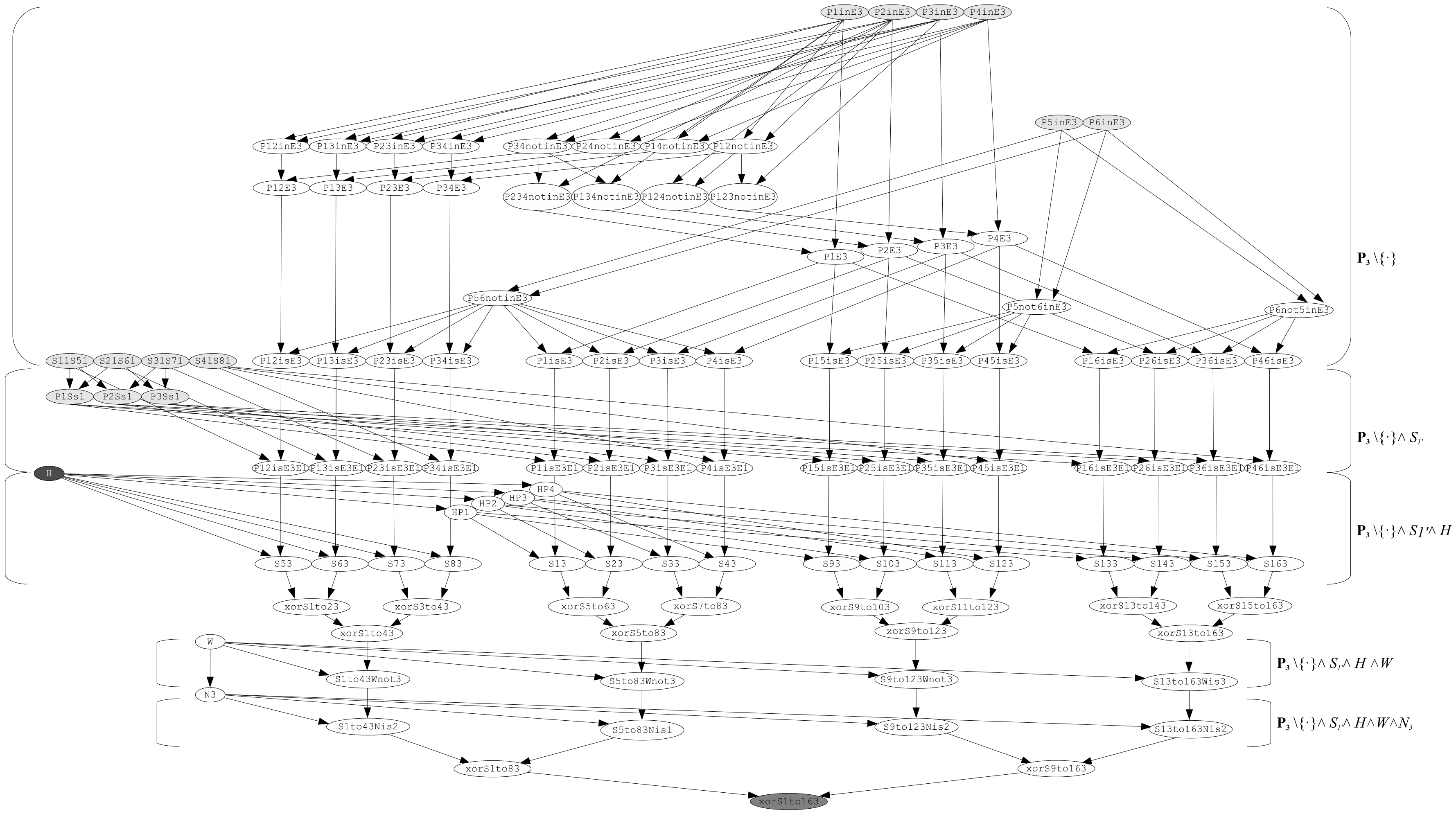


Figure. E.2 – BN for handling the scenarios of $\mathbf{S}_3$. Structure of the BN for the scenarios $\mathbf{S}_4$ is the same.

E.2 Locus modeling for LDLR

Figure E.3 shows the BN used for evaluating the typing results for all three specimens at locus L_2 (LDLR). The outer brackets annotated by specimen represent the subgraphs for the evaluation of the marker L_2 for the corresponding specimen. The nodes in that subgraph associated with the specimen possess suffixes indicating their affiliation to the specimen (i.e., `-E1`, `-E3`, and `-E4`). Note also that the genotype of each person is only created once, that is for P_1 to P_5 along with the subgraph associated with $E_{1'}$, and P_6 along with the subgraph associated with E_3. However, their corresponding nodes possess no suffixes indicating the specimen in their name. This highlights the fact that the genotype of a person is affiliated to the person himself but not to the the specimen. The node definitions for these nodes are as described in Section 5.5. The nodes `P1inE1` to `P5inE1`, `P1inE3` to `P6inE3`, and `P1inE4` to `P6inE4` remain exactly the same as before and were created once for each specimen. These are the nodes that connect to the BN for the contributor scenarios. The remaining brackets are annotated with the conjunction or disjunction symbols. The nodes embraced by these brackets are defined as logical conjunctions and disjunctions of their parent nodes. The other loci were constructed analogously and vary only regarding the number of alleles they possess.

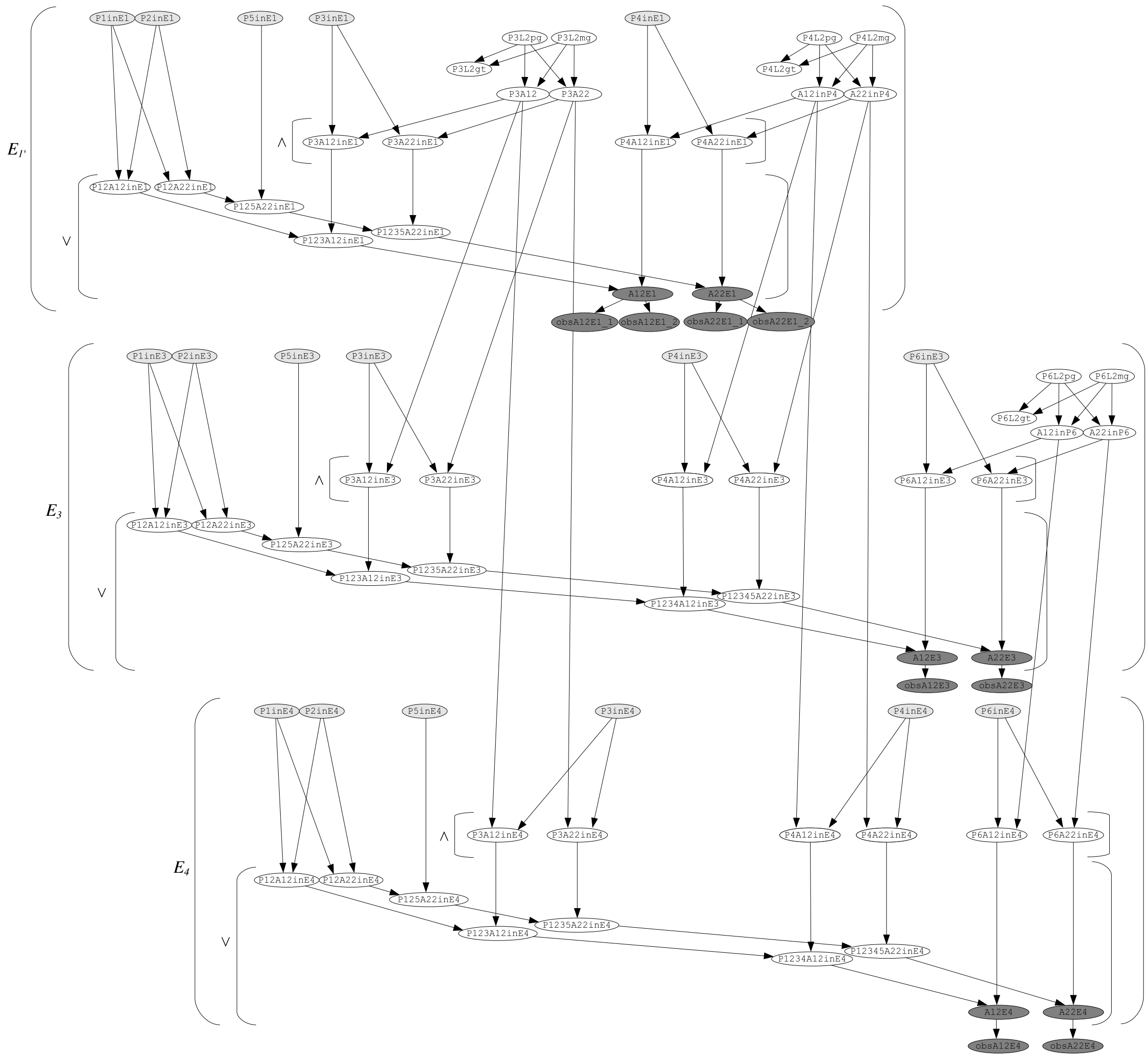


Figure. E.3 – BN for evaluating the typing results of marker L_2 for all three specimens

E.3 Results for the computation of WoE and LR regarding the question of weather Sutton is an assailant or not

Table. E.3 – WoE of each allele, where the specimens are considered jointly.

Allele	$W(h_p : obs\mathbf{A}_{1'} \mid \mathbf{I})$		$W(h_p : obs\mathbf{A}_3 \mid obs\mathbf{A}_{1'}, \mathbf{I})$		$W(h_p : obs\mathbf{A}_4 \mid obs\mathbf{A}_{1'}, obs\mathbf{A}_3, \mathbf{I})$	
	Error 1	Error 2	Error 1	Error 2	Error 1	Error 2
1.1 DQA	4.896953e-01	4.896953e-01	-9.040374e-03	-2.993216e-05	-6.784527e-06	-1.236032e-04
1.2 DQA	1.080890e-01	1.083075e-01	-2.269334e-06	-1.939803e-11	3.227787e-09	-1.053216e-10
2 DQA	2.923574e-01	2.919095e-01	4.395348e-01	4.753759e-01	-1.493273e+00	-2.514227e+00
3 DQA	-4.237306e-02	-4.225744e-02	2.565181e-02	8.414982e-04	-6.917515e-01	7.904631e-03
4.1 DQA	-7.393097e-02	-7.380779e-02	-4.586104e-01	-4.762677e-01	-5.225200e-01	-1.493236e+00
4.2/4.3 DQA	-6.294434e-02	-6.301437e-02	4.117635e-04	6.398949e-07	-9.907414e-01	-9.961593e-01
other DQA	4.146608e-15	4.821637e-16	-2.783044e-07	-2.946503e-07	-3.031043e-04	-3.659946e-04
A LDLR	2.380921e-01	2.380921e-01	-2.368358e-01	-2.368731e-01	7.515001e-05	-3.236420e-04
B LDLR	1.065125e-03	1.065651e-03	-1.608808e-03	-1.692454e-03	2.259615e-04	1.944489e-05
A GYPA	3.402475e-02	3.404226e-02	-7.488293e-02	-7.788652e-02	4.240599e-04	1.289654e-05
B GYPA	-8.670682e-02	-8.677364e-02	1.265243e-01	1.305224e-01	-1.586006e-01	-1.541305e-01
A HBGG	4.878020e-02	4.880945e-02	-5.496909e-02	-5.667072e-02	1.538954e-04	-3.204213e-05
B HBGG	2.032571e-01	2.034157e-01	7.699710e-05	-3.001932e-06	-9.896642e-04	-1.084754e-03
C HBGG	-1.175926e-01	-1.177323e-01	1.546352e-01	1.580841e-01	7.125000e-05	-1.929296e-04
A D7S8	5.634194e-03	5.637917e-03	-3.909567e-02	-3.940541e-02	1.088239e-04	-1.126366e-05
B D7S8	-1.568720e-01	-1.568711e-01	1.900820e-01	1.904063e-01	1.140496e+00	9.759675e-01
A Gc	7.901676e-02	7.905834e-02	9.932158e-06	-3.466362e-07	-1.346097e-04	-2.310922e-03
B Gc	4.418692e-04	4.414413e-04	3.550712e-06	8.922776e-09	3.645673e-09	2.124992e-12
C Gc	1.702418e-01	1.703772e-01	3.524921e-05	-5.161000e-07	-2.192079e-04	-3.790766e-03
20 D1S80	-2.831001e-01	-2.831597e-01	-	-	-	-
21 D1S80	-4.067892e-01	-4.095781e-01	-	-	-	-
24 D1S80	-1.591810e+00	-1.679658e+00	-	-	-	-
25 D1S80	1.142455e+00	1.143455e+00	-	-	-	-
28 D1S80	1.861009e+00	2.021184e+00	-	-	-	-
34 D1S80	3.499234e-02	1.410091e-03	-	-	-	-
other D1S80	3.553468e-02	1.331101e-03	-	-	-	-

Table. E.4 – WoE of each allele, where each specimen is considered in isolation.

Allele	$W(h_p : obs\mathbf{A}_{1'} \mid \mathbf{I})$		$W(h_p : obs\mathbf{A}_3 \mid \mathbf{I})$		$W(h_p : obs\mathbf{A}_4 \mid \mathbf{I})$	
	Error 1	Error 2	Error 1	Error 2	Error 1	Error 2
1.1 DQA	1.127565e+00	4.896953e-01	-6.268951e-01	-2.813092e-01	-6.154154e-01	-2.760765e-01
1.2 DQA	2.488842e-01	1.083075e-01	-3.216665e-01	-1.945143e-01	-3.302723e-01	-1.984645e-01
2 DQA	6.731777e-01	2.919095e-01	-8.400561e-02	-1.163595e-03	1.168671e-01	2.312748e-03
3 DQA	-9.756758e-02	-4.225744e-02	-6.112695e-02	-2.124136e-03	-2.325595e-01	-8.218724e-03
4.1 DQA	-1.702323e-01	-7.380779e-02	1.426850e-02	1.846253e-03	8.466582e-03	2.849330e-03
4.2/4.3 DQA	-1.449347e-01	-6.301437e-02	-2.062532e-02	-1.325255e-04	-4.371378e-02	-2.735682e-04
other DQA	-1.211253e-13	8.804310e-14	-1.032992e-06	-2.566668e-08	-2.585248e-06	-6.193033e-08
A LDLR	5.482272e-01	2.380921e-01	-9.001128e-04	-2.050075e-04	6.464492e-03	-3.994464e-04
B LDLR	2.452541e-03	1.065651e-03	-2.705628e-04	-5.653528e-05	1.011683e-03	-1.338895e-04
A GYPA	7.834489e-02	3.404226e-02	-6.943639e-04	-1.544682e-04	3.966223e-03	-3.273143e-04
B GYPA	-1.996498e-01	-8.677364e-02	-3.289287e-03	-3.094306e-04	-6.669965e-03	-8.574484e-04
A HBGG	1.123206e-01	4.880945e-02	1.247287e-03	-2.626899e-04	4.850867e-03	-7.622002e-04
B HBGG	4.680167e-01	2.034157e-01	3.143007e-03	6.977744e-04	7.720935e-03	1.646427e-03
C HBGG	-2.707670e-01	-1.177323e-01	-1.073731e-03	-2.670972e-04	7.674363e-03	-8.219244e-04
A D7S8	1.297321e-02	5.637917e-03	-3.395431e-04	-9.437831e-05	3.450265e-03	-2.676100e-04
B D7S8	-3.612110e-01	-1.568711e-01	-4.700629e-03	-5.505020e-04	2.825292e-02	1.800781e-03
A Gc	1.819428e-01	7.905834e-02	1.817361e-03	2.460510e-04	1.016238e-02	2.087866e-04
B Gc	1.017441e-03	4.414413e-04	-2.050459e-04	-3.918724e-05	2.212039e-03	-2.077229e-05
C Gc	3.919963e-01	1.703772e-01	2.450779e-03	4.392031e-04	2.559932e-02	3.755731e-04
20 D1S80	-6.518620e-01	-2.831597e-01	-	-	-	-
21 D1S80	-9.366668e-01	-4.095781e-01	-	-	-	-
24 D1S80	-3.665278e+00	-1.679658e+00	-	-	-	-
25 D1S80	2.630601e+00	1.143455e+00	-	-	-	-
28 D1S80	4.285131e+00	2.021184e+00	-	-	-	-
34 D1S80	8.057283e-02	1.410091e-03	-	-	-	-
other D1S80	8.182163e-02	1.331101e-03	-	-	-	-

E.4 Posterior probabilities of H

Table. E.5 – Posterior probabilities of H

	Posterior probabilities			
	h_1	h_2	h_3	h_4
Error 1	3.139338e-32	1.562185e-01	1.465165e-24	8.437815e-01
Error 2	4.472688e-47	6.410677e-03	1.690472e-34	9.935893e-01

E.5 Probabilities of the scenarios

Table. E.6 – Posterior probabilities of each scenario $S_{s,i}$

Scenarios	$Pr(S_{s,1'} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$		$Pr(S_{s,3} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$		$Pr(S_{s,4} \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$	
	Error 1	Error 2	Error 1	Error 2	Error 1	Error 2
$S_{1,i}$	3.519133e-43	1.821919e-55	3.347282e-14	5.201095e-17	2.540382e-07	1.897841e-07
$S_{2,i}$	1.534932e-01	6.410393e-03	1.568539e-37	1.781598e-52	4.499045e-37	5.912073e-52
$S_{3,i}$	2.922752e-25	6.410393e-03	5.764686e-01	5.029782e-01	4.265627e-01	4.972019e-01
$S_{4,i}$	8.175203e-01	9.935095e-01	4.209310e-01	4.965689e-01	4.215299e-01	4.967753e-01
$S_{5,i}$	3.139338e-32	4.472688e-47	7.607106e-42	3.342425e-60	1.309285e-41	4.483963e-59
$S_{6,i}$	2.725353e-03	2.840879e-07	3.347098e-05	1.182586e-08	2.966946e-03	9.131375e-06
$S_{7,i}$	1.172890e-24	1.725839e-36	7.535579e-31	1.086494e-44	3.669249e-31	5.265075e-45
$S_{8,i}$	2.626115e-02	7.980058e-05	5.998413e-04	2.275267e-05	3.071794e-05	8.741191e-08
$S_{9,i}$	-	-	4.025857e-08	3.313212e-12	1.480442e-01	5.857877e-03
$S_{10,i}$	-	-	1.886520e-31	1.134917e-47	2.621879e-31	1.824821e-47
$S_{11,i}$	-	-	1.289026e-03	3.696254e-05	3.474798e-04	1.851456e-05
$S_{12,i}$	-	-	6.445755e-04	3.578080e-05	3.443375e-04	1.850134e-05
$S_{13,i}$	-	-	2.668948e-08	4.107609e-11	1.611565e-04	1.163449e-04
$S_{14,i}$	-	-	3.685653e-31	4.384680e-46	1.648614e-31	2.271331e-46
$S_{15,i}$	-	-	1.816764e-05	1.787385e-04	1.120041e-05	1.339438e-06
$S_{16,i}$	-	-	1.523849e-05	1.786409e-04	1.163655e-06	8.599962e-07
$Pr(\mathbf{S}_i \mid \mathbf{A}^*_{1'}, \mathbf{A}^*_3, \mathbf{A}^*_4, \mathbf{I})$	1	1	1	1	1	1

Bibliography

[1] P. Achinstein. Concepts of evidence. *Mind, New Series*, 87:22–45, 1978.

[2] C. G. G. Aitken and A. Gammerman. Probabilistic reasoning in evidential assessment. *Journal of the Forensic Science Society*, 29:303–316, 1989.

[3] C. G. G. Aitken and D. A. Stoney. *The Use of Statistics in Forensic Science*. Ellis Horwood, New York, 1991.

[4] C. G. G. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 2nd edition, 2004.

[5] 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.

[6] T. J. Anderson, D. A. Schum, and W. L. Twining. *Analysis of Evidence*. Cambridge University Press, Cambridge, 2nd edition, 2005.

[7] Aristotle. *De Rethorica Ad Alexandrum*, volume XI. Claredon Press, Oxford, 1924.

[8] A. Bell, J. Swenson-Wright, and K. Tybjerg. *Evidence*. The Darwin College Collection. Cambridge University Press, Cambridge, 2008.

[9] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, Chichester, 2000.

[10] J. Bernoulli. *Ars Conjectandi*, volume 3 and 4. Verlag Von Wilhelm Engelmann, Leipzig, 1899.

[11] A. Biedermann. *Bayesian Networks and the Evaluation of Scientific Evidence in Forensic Science*. PhD thesis, Université de Lausanne, Lausanne, 2007.

[12] 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, 2012.

[13] A. Biedermann and T. Hicks. The importance of critically examining the level or propositions when evaluating forensic DNA results. *Frontiers in Genetics*, 7:1–4, 2016.

[14] A. Biedermann and F. Taroni. A probabilistic approach to the joint evaluation of firearm evidence and gunshot residues. *Forensic Science International*, 162:18–33, 2006.

[15] A. Biedermann, F. Taroni, O. Delemont, C. Samenadi, 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, 2005.

[16] A. Biedermann, F. Taroni, O. Delemont, C. Samenadi, 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, 2005.

[17] C. H. Brenner, R. Fimmers, and M. P. Baur. Likelihood ratios for mixed stains when the number of donors cannot be agreed. *International Journal of Legal Medicine*, 109:218–219, 1996.

[18] J. S. Buckleton, C. M. Triggs, and C. Champod. An extended likelihood ratio framework for interpreting evidence. *Science & Justice*, 46:69–78, 2006.

[19] C. Bundesen and T. Habekost. *The Handbook of Cognition*, chapter 4 Attention, pages 105–129. Sage Publications, London, 2006.

[20] W. L. Buntine. Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, 2:159–225, 1994.

[21] C. Champod. Edmond Locard - Numerical standards & "probable" identifications. *Journal of Forensic Identification*, 45:136–163, 1995.

[22] C. Champod, C. J. Lennard, P. Margot, and M. Stoilovic. *Fingerprints and Other Ridge Skin Impressions*. CRC Press, Boca Raton, 2004.

[23] 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, 1998.

[24] R. Coquoz, J. Comte, D. Hall, T. Hicks, and F. Taroni. *Preuve pas l'ADN: La génétique au service de la justice*. Collection Sciences Forensiques. Presses polytechniques et universitaires romandes, Lausanne, 3ème édition edition, 2013.

[25] R Core Team. R: A Language and Environment for Statistical Computing, version 3.1.0, 2014.

[26] F. Crispino, O. Ribaux, M. Houck, and P. Margot. Forensic science - A true science? *Australian Journal of Forensic Science*, 43:157–176, 2011.

[27] J. M. Curran. An introduction to Bayesian credible intervals for sampling error in DNA profiles. *Law, Probability and Risk*, 4:115–126, 2005.

[28] A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, New York, 2009.

[29] A. P. Dawid. Introduction. In P. Dawid, L. Twining, and M. Vasilaki, editors, *Evidence, Inference, and Enquiry*, number 271, pages 1–9, Oxford, 2011. The British Academy, Oxford University Press.

[30] 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.

[31] 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.

[32] 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.

[33] B. de Finetti. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'I.H.P*, 7:1–68, 1937.

[34] B. de Finetti. Logical foundations and measurement of subjective probability. *Acta Psychologica*, 34(129-145), 1970.

[35] B. de Finetti. *Probability, Induction and Statistics*. John Wiley & Sons, New York, 1972.

[36] B. de Finetti. Probabilism: A critical essay on the theory of probability an on the value of science. *Erkenntnis*, 31:169–223, 1989.

[37] B. de Finetti. *Philosophical Lecture on Probability - collected, edited, and annotated by Alberto Mura*. Synthese Library. Springer Science+Business Media B.V., Sassari, 2010.

[38] W. Edwards. Influence diagrams, Bayesian imperialism, and the Collins case: An appeal to reason. *Cardozo Law Review*, 13:1025–1074, 1991.

[39] W. Edwards, H. Lindman, and L. J. Savage. Bayesian statistical inference for psychological research. *Psychological Review*, 70:193–242, 1963.

[40] I. W. Evett. Evaluating DNA profiles in a case where the defence is "It was my brother". *Journal of the Forensic Science Society*, 32:5–14, 1992.

[41] 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.

[42] I. W. Evett. Expert evidence and forensic misconceptions of the nature of exact science. *Science & Justice*, 36:118–122, 1996.

[43] 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.

[44] I. W. Evett, G. Jackson, J. A. Lambert, and S. McCrossan. The impact of the principles of evidence interpretation and the structrue and content of statements. *Science & Justice*, 40:233–239, 2002.

[45] I. W. Evett, G. J. Jackson, J. A. Lambert, and S. McCrossan. The impact of the principles of evidence interpretation on the structure and and content of the statements. *Science & Justice*, 40:233–239, 2000.

[46] 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.

[47] 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, 2000.

[48] W. B. Fairley and W. Mosteller. A conversation about Collins. *University of Chicago Law Review*, 41:242–253, 1974.

[49] P. Feyerabend. *Against Method*. Verso, London, 1975.

[50] J. G. Fichte. *Fichtes Werke. Band I, Zur theoretischen Philosophie I, hrsg. von Immanuel Hermann Fichte, 1845/1846*. Walter de Gruyter & Co., Berlin, 1971.

[51] S. E. Fienberg. When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, 1:1–41, 2003.

[52] J. Franklin. The *Science of Conjecture: Evidence and Probability before Pascal*. John Hopkins University Press, Baltimore, 2002.

[53] R. D. Friedman. Assessing evidence. *Michigan Law Review*, 94:1810–1838, 1996.

[54] H. Gaifman and M. Snir. Probabilities over rich languages, testing and randomness. *The Journal of Symbolic Logic*, 47:495–548, 1982.

[55] D. Garber and S. Zabell. On the emergence of probability. *Archive for History of Exact Sciences*, 21:33–53, 1979.

[56] P. Garbolino and F. Taroni. Evaluation of scientific evidence using Bayesian networks. *Forensic Science International*, 125:149–155, 2002.

[57] A. J. M. Garrett. The history of probability theory. In G. J. Erickson, J. T. Rychert, and C. R. Smith, editors, *Maximum Entropy and Bayesian Methods*, volume 98 of *Fundamental Theories of Physics*, pages 223–238. Springer Netherlands, 1998.

[58] I. J. Good. *Probability and the Weighing of Evidence*. Butler and Tanner, London, 1950.

[59] I. J. Good. Effective sampling rates for signal detection; Or can the Gaussian model be salvaged? *Information and Control*, 3(116-140), 1960.

[60] I. J. Good. Weight of evidence, corroboration, explanatory power, information and the utiliy of experiments. *Journal of the Royal Statistical Society. Series B*, 22:319–331, 1960.

[61] I. J. Good. A derivation of the probabilistic explanation of information. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28:578–581, 1966.

[62] I. J. Good. On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17:319–321, 1967.

[63] I. J. Good. Studies in the histroy of probability and statistics. XXXVII A. M. Turing's statistical work in World War II. *Biometrika*, 66:393–396, 1979.

[64] I. J. Good. *Good Thinking - The Foundations of Probability and Its Applications*. Dover Publications, Mineola, 1983.

[65] I. J. Good. Weight of Evidence: A Brief Survey. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 2*, pages 249–270, North Holland, 1985. Proceedings of the 2nd Valencia International Meeting 1983, Elsevier Science Publishers.

[66] I. J. Good and W. I. Card. The diagnostic process with special reference to errors. *Methods of Information in Medicine*, 10:176–188, 1971.

[67] E. Gutièrrez-Redomero, C. Alonso-Rodriguez, J. Hernandez-Hurtado, and L. Rodriguez-Villalba. Distribution of the *minutiae* in the fingerprints of a sample of the Spanish population. *Forensic Science International*, 208:79–90, 2011.

[68] I. Hacking. Combined evidence. In Sören Stenlund, Ann-Mari Henschen-Dahlquist, Lars Lindahl, Lennart Nordenfelt, and Jan Odelstad, editors, *Logical Theory and Semantic Analysis*, volume 63, pages 113–123. Springer Netherlands, 1974.

[69] I. Hacking. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference*. Cambridge University Press, Cambridge, 1975.

[70] A. Hall and J. Ballantyne. Novel Y-STR typing strategies reveal the genetic profile of the semen donor in extended interval post-coital cervicovaginal samples. *Forensic Science International*, 136:58–72, 2003.

[71] A. Hepler, A. P. Dawid, and V. Leucari. Object-Oriented graphical representation of complex patterns of evidence. *Law, Probability and Risk*, 6:275–293, 2007.

[72] HPD. Houston police department offense report, incident no. 137009698, supplement no.0022, June 16 1999.

[73] K. Inman and N. Rudin. The origin of evidence. *Forensic Science International*, 126:11–16, 2002.

[74] H. Jeffreys. *Theory of Probability*. Claredon Press, Oxford, 3rd edition, 1961.

[75] F. V. Jensen. *An Introduction to Bayesian Networks*. UCL Press, London, 1996.

[76] F. V. Jensen, B. Chamberlain, T. Nordahl, and F. Jensen. Analysis in HUGIN of data conflict. In *Uncertainty in Artificial Intelligence: Proceedings of the Sixth Conference*, pages 546–554, Mountain View, 1990. Association of Uncertainty in Artificial Intelligence, Elsevier Science Publishers.

[77] F. V. Jensen and T. D. Nielsen. *Bayesian Networks and Decision Graphs*. Springer, New York, 2nd edition, 2007.

[78] 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.

[79] J. B. Kadane and D. A. Schum. *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. John Wiley & Sons, New York, 1996.

[80] I. Kant. *Kritik der Reinen Vernunft*. Warthogsbooks, http://kickme.to/Warthogsbooks, 1787.

[81] D. H. Kaye and J. J. Koehler. The misquantification of probative value. *Law and Human Behavior*, 27:645–658, 2003.

[82] S. S. Kind. *The Scientific Investigation of Crime*. Forensic Science Services, Harrogate, 1987.

[83] U. B. Kjaerulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer, New York, 2nd edition, 2013.

[84] N. Köller, K. Niessen, M. Riess, and E. Sadorf. *Probability Conclusions in Expert Opinions on Handwriting. Substantiation and Standardization of Probability Statements in Expert Opinions*. Luchterhand, München, 2004.

[85] K. Konis. RHugin, R package version 8.0, 2010.

[86] K. B. Korb and A. E. Nicholson. *Bayesian Artificial Intelligence*. Chapman & Hall/CRC, Boca Raton, 2nd edition, 2011.

[87] K. B. Laskey. Conflict and surprise: Heuristics for model revision. In *Uncertainty in Artificial Intelligence: Proceedings of the Seventh Conference*, pages 197–204, San Mateo, 1991. Morgan Kaufmann Publishers.

[88] K. B. Laskey and S. M. Mahoney. Network fragments: Representing knowledge for constructing probabilistic models. In D. Geiger and P. Shenoy, editors, *Uncertainty in Artificial Intelligence: Proceedings of the Thirteenth Conference*. Morgand Kaufmann, San Fransisco, 1997.

[89] K. B. Laskey, D. A. Schum, P. C. G. Costa, and T. Janssen. Ontology of evidence. In *Proceedings of the Third International Ontology for the Intelligence Community Conference (OIC 2008)*, Fairfax, 2008.

[90] B. P. Lathi. *Modern Digital and Analog Communication Systems*. Oxford University Press, New York, 3rd edition, 1998.

[91] S. L. Lauritzen and J. Mortera. Bounding the number of contributors to mixed DNA stains. *Forensic Science International*, 130:125–126, 2002.

[92] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Society. Series B (Methodological)*, 50:157–224, 1988.

[93] R. O. Lempert. Modeling relevance. *Michigan Law Review*, 75:1021–1057, April-May 1977.

[94] T. S. Levitt and K. B. Laskey. Computational inference for evidential reasoning in support of judicial proof. *Cardozo Law Review*, 22:1691–1731, 2001.

[95] D. V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27:986–1005, 1956.

[96] D. V. Lindley. The philosophy of statistics. *The Statistician*, 49:293–337, 2000.

[97] D. V. Lindley. *Understanding Uncertainty*. John Wiley & Sons, Hoboken, 2006.

[98] P. Lipton. *Evidence and Explanation*, chapter 1. The Darwin College Collection. Cambridge University Press, Cambridge, 2008.

[99] E. Locard. *L'enquête Criminelle et les Méthodes Scientifiques*. Flammarion, Paris, 1920.

[100] T. D. Lyon and J. J. Koehler. Relevance ratio: Evaluating the probative value of expert testimony in child sexual abuse cases. *Cornell Law Review*, 82:43–78, 1996.

[101] S. M. Mahoney and K. B. Laskey. Constructing situation specific networks. In G. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence: Proceedings of the Fourteenth Conference on Artificial Intelligence*. Morgand Kaufmann, San Fransisco, 1998.

[102] K. A. Mayntz-Press, L. M. Sims, A. Hall, and J. Ballantyne. Y-STR profiling in extended interval ($\leq$ 3 days) postcoital cervicovaginal samples. *Journal of Forensic Sciences*, 53:342–348, 2008.

[103] J. Mortera. Analysis of DNA mixtures using probabilistic expert systems. In P. L. Green, N. L. Hjort, and S. Richardson, editors, *Highly Structured Stochastic Systems*. Oxford University Press, Oxford, 2003.

[104] J. Mortera, A. P. Dawid, and S. L. Lauritzen. Probabilistic expert systems for DNA mixture profiling. *Theoretical Population Biology*, 63:191–205, 2003.

[105] National Academy of Sciences (NAS). *Strengthening Forensic Science in the United States: A Path Forward*. National Academic Press, Washington, 2009.

[106] Court of Appeal. R. v. T [2010] EWCA Crim 2439, October 2010.

[107] E. N. O'Neal, S. H. Decker, C. Spohn, and K. Tellis. Condom use during sexual assault. *Journal of Forensic and Legal Medicine*, 20:605–609, 2013.

[108] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, 1988.

[109] B. L. Peterson, B. Su, R. Chakraborty, B. Budowle, and R. E. Gaensslen. World population data for the HLA-DQA1, PM and D1S80 loci with least and most common profile frequencies for combinations of loci estimated following NRC II guidlines. *Journal of Forensic Sciences*, 45:118–145, 2000.

[110] J. H. Poincaré. *The Value of Science*. The Science Press, New York, 1907.

[111] J. H. Poincaré. *La Valeur de la Science*. Ernest Flammarion, Paris, 1908.

[112] O. Pourret, P. Naim, and B. Marcot, editors. *Bayesian Networks: A Practical Guide to Applications*. John Wiley & Sons, Chichester, 2008.

[113] C. Quigley. *The Evolution of Civilization*. MacMillan and Co., New York, 1961.

[114] C. Quigley. Assumption and inference on human origins. *Current Anthropology*, 12:519–540, 1971.

[115] Quintilian. *Institutio Oratoria*, volume IV-VI. The Loeb Classical Library, Great Britain, 1977.

[116] B. Robertson and G. A. Vignaux. Probability - The logic of the law. *Oxford Journal of Legal Studies*, 13:457–478, 1993.

[117] B. Robertson and G. A. Vignaux. Taking fact analysis seriously. *Michigan Law Review*, 91:1442–1464, 1993.

[118] B. Robertson and G. A. Vignaux. *Interpreting Evidence, Evaluating Forensic Science in the Courtroom*. John Wiley & Sons, Chichester, 1995.

[119] C. Roux, B. Talbot-Wright, J. Robertson, F. Crispino, and O. Ribaux. The end of the (forensic science) world as we know it? The example of trace evidence. *Philosophical Transactions of the Royal Society B*, 370:1–8, 2015.

[120] S. Russel and P. Norvig. *Artificial Intelligence - A Modern Approach*. Prentice Hall, Upper Saddle River, 2010.

[121] 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.

[122] L. J. Savage. *The Foundations of Statistics*. Dover Publications, New York, 1972.

[123] D. A. Schum. Sorting out the effects of witness sensitivity an response-criterion placement upon the inferential value of testimonial evidence. *Organizational Behavior and Human Performance*, 27:153–196, 1981.

[124] D. A. Schum. Inference networks and the evaluation of evidence: Alternative analyses. In K. B. Laskey and H. Prade, editors, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 575–584, San Francisco, 1999. Morgan Kaufmann Publishers.

[125] D. A. Schum. *The Evidential Foundations of Probabilistic Reasoning*. Northwestern University Press, Evanston, 2001.

[126] D. A. Schum. A science of evidence: Contributions from law and probability. *Law, Probability and Risk*, 8:197–231, 2009.

[127] D. A. Schum and A. W. Martin. Formal and empirical research on cascaded inference in jurisprudence. *Law & Society Review*, 1:105–152, 1982.

[128] D. A. Schum and J. R. Morris. Assessing the competence and credibility of human sources of intelligence evidence: Contributions from law and probability. *Law, Probability and Risk*, 6:247–274, 2007.

[129] D. A. Schum and P. E. Pfeiffer. Observer reliability and human inference. *IEEE Transactions on Reliability*, 22:170–176, 1973.

[130] 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.

[131] F. Taroni, C. G. G. Aitken, G. Garbolino, and A. Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, Chichester, 2006.

[132] F. Taroni and A. Biedermann. Inadequacies of posterior probabilities for the assessment of scientific evidence. *Law, Probability and Risk*, 4:89–114, 2005.

[133] F. Taroni, A. Biedermann, S. Bozza, and P. Garbolino. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. John Wiley & Sons, Chichester, 2nd edition, 2014.

[134] 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.

[135] F. Taroni, S. Bozza, A. Biedermann, G. Garbolino, and C. G. G. Aitken. *Data analysis in forensic science: A Bayesian decision perspective*. Statistics in Practice. John Wiley & Sons, Chichester, 2010.

[136] F. Taroni, R. Marquis, M. Schmittbuhl, A. Biedermann, A. Thiéry, and S. Bozza. Bayes factor for investigative assessment of selected handwriting features. *Forensic Science International*, 242:266–273, 2014.

[137] P. Thagart. Why wasn't O.J. convicted? Emotional coherence and legal inference. *Cognition and Emotion*, 17:361–383, 2003.

[138] W. C. Thompson. Review fo DNA evidence in State of Texas v. Josiah Sutton, February 2003.

[139] W. C. Thompson. Beyond bad apples: Analyzing the role of forensic science in wrongful convictions. *Southwestern University Law Review*, 37:112–119, 2009.

[140] 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.

[141] P. Tillers. Webs of things in the mind: A new science of evidence. *Michigan Law Review*, 87:1225–1258, 1989.

[142] L C van der Gaag and S Renooij. Analysing sensitivity data form probabilistic networks. In *UAI '01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, pages 530–537, San Francisco, 2001. Morgan Kaufmann Publishers.

[143] J. W. von Goethe. *Faust*. The Riverside Press, Cambridge, 1870.

[144] J. W. von Goethe. *Faust*. The Project Gutenberg, http://gutenberg.aol.de, 2010.

[145] J. H. Wigmore. The problem of proof. *Illinois Law Review*, 8:77–103, 1913.

[146] J. H. Wigmore. *The Science of Judicial Proof: As Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials*. Brown, Boston, 3rd edition, 1937.