Science and Justice 58/2 (2018) 159–165

Contents lists available at ScienceDirect

Science and Justice

journal homepage: www.elsevier.com/locate/scijus

Analysing and exemplifying forensic conclusion criteria in terms of Bayesian decision theory

A. Biedermann[a,*], S. Bozza[b,a], F. Taroni[a]

[a] University of Lausanne, School of Criminal Justice, Lausanne-Dorigny 1015, Switzerland
[b] Ca'Foscari University Venice, Department of Economics, Venice 30121, Italy

ARTICLE INFO

Article history:
Received 16 May 2017
Received in revised form 13 July 2017
Accepted 20 July 2017

Keywords:
Decision-making
Conclusion criteria
Bayesian decision theory
Normative approach

ABSTRACT

There is ongoing discussion in forensic science and the law about the nature of the conclusions reached based on scientific evidence, and on how such conclusions – and conclusion criteria – may be justified by rational argument. Examples, among others, are encountered in fields such as fingermarks (e.g., 'this fingermark comes from Mr. A's left thumb'), handwriting examinations (e.g., 'the questioned signature is that of Mr. A'), kinship analyses (e.g., 'Mr. A is the father of child C') or anthropology (e.g., 'these are human remains'). Considerable developments using formal methods of reasoning based on, for example (Bayesian) decision theory, are available in literature, but currently such reference principles are not explicitly used in operational forensic reporting and ensuing decision-making. Moreover, applied examples, illustrating the principles, are scarce. A potential consequence of this in practical proceedings, and hence a cause of concern, is that underlying ingredients of decision criteria (such as losses quantifying the undesirability of adverse decision consequences), are not properly dealt with. There is merit, thus, in pursuing the study and discussion of practical examples, demonstrating that formal decision-theoretic principles are not merely conceptual considerations. Actually, these principles can be shown to underpin practical decision-making procedures and existing legal decision criteria, though often not explicitly apparent as such. In this paper, we will present such examples and discuss their properties from a Bayesian decision-theoretic perspective. We will argue that these are essential concepts for an informed discourse on decision-making across forensic disciplines and the development of a coherent view on this topic. We will also emphasize that these principles are of normative nature in the sense that they provide standards against which actual judgment and decision-making may be compared. Most importantly, these standards are justified independently of peoples' observable decision behaviour, and of whether or not one endorses these formal methods of reasoning.

1. Introduction

What degree of personal belief should be required before deciding in favour of a particular option? This question is fundamental and arises recurrently. It is inevitable not only in many situations of daily life, but also in virtually any professional area of activity (e.g., economics, engineering, and medicine) [e.g., 19]. In legal contexts, the question of decision takes a highly visible position, mainly because of the direct impact that convictions and acquittals have on all parties involved in the legal process. The ultimate issue is only one decision point, however, among many others in the legal process. Some of those decisions relate to scientific evidence presented by forensic scientists, as will be exemplified later in this paper. Other decision points relate to questions such as whether or not to hear a particular witness, or whether or not to conduct a search.

On a broad view, there are several ways to deal with decision-making. One is dismissively, often justified by reasons such as the need for practically feasible procedures or limitations of resources (e.g., time). Such an approach may be paired with trust in personal experience or a preference for intuitive proceedings. Indeed, there are many day-to-day situations in which a decision must be made and where spending too much time on introspection is neither necessary nor desirable. But there are also other situations in which it is appropriate to formalise intuition – as an integral part of logical

* Corresponding author.
E-mail address: alex.biedermann@unil.ch (A. Biedermann).

http://dx.doi.org/10.1016/j.scijus.2017.07.002

reasoning in the face of uncertainty – and devote time to a serious analysis of how to make a decision, so as to guarantee that throughout decision analysis one is able to measure the quality of decisions [6]. This is typically the case when the stakes involved are high, or adverse decision consequences are severe. One very well known way to look at these different decision perspectives is through Kahneman's notion of fast and slow thinking [17]. In this paper, we posit that professional decision analysis related to the evaluative use of forensic science results in the legal context provides a strong case for the introspective approach.

Once it is agreed that questions of decision ought to be approached through an in-depth perspective, practicing and academic decision analysts commonly distinguish – in law as in other disciplines – between two main accounts, the normative and descriptive [e.g., [5]].[1] The descriptive account takes an interest in people's observable decisional and judgmental behavior. The normative account considers, instead, the rational standards by which judgment and decisions ought to be evaluated. Naturally, the descriptive approach is strongly rooted in empirical considerations. Over the past decades, there has been abundant research on, for example, the elicitation of what various subjects (e.g., judges, citizens and students) consider as 'beyond reasonable doubt' (e.g., [11,26], for a review see also Ref. [14]). Quantitative values obtained in such studies, using various elicitation procedures, vary over a broad range, depending on the experimental conditions. Such general knowledge about the observable properties of human behavior with respect to questions of judgment and decision is valuable, but the more fundamental question is what one's required level of personal degree of belief, before making a particular decision, actually means from a logical point of view. This is a question that pertains to the normative domain [9] and will be a main focus of attention in this paper. We will identify here the normative standpoint in terms of the classic decision-theoretic account, also known as Bayesian decision theory, given by probability and utility theory. Specifically, we will exemplify how this account allows one to capture the essential features of existing forensic decision procedures and conclusion schemes.

This paper is structured as follows. Section 2 introduces notation and one of the forms in which the Bayesian decision-theoretic criterion can be stated. The format we choose is based on the notion of loss for qualifying decision consequences. It is our preferred choice for the purpose of this paper because, compared to other formats, it helps to break down some of the formulaic burden. Further technicalities are confined to the Appendix. Section 3 will exemplify how the criterion allows one to clarify the preferences among decision consequences that are implied by current decision thresholds as used, for example, in kinship analyses in different legal systems. We will also discuss these insights and their relevance for decision practices in other areas of forensic science, such as fingermarks and comparative handwriting examinations. In Section 4, we will emphasize the prescriptive value of the approach. By prescriptive value we mean the potential to provide incentives and means for improving the practical understanding of how to decide based on forensic science results, and how to ensure coherence between decision policies across different forensic disciplines.

2. The Bayesian decision criterion

The basic tenet of the Bayesian decision-theoretic approach is – in one of its formulations – the weighing of losses, quantifying the undesirability of wrong decisions, with one's personal probabilities for such outcomes. In the legal context, typical examples for adverse decision outcomes are the conviction of an innocent person or the acquittal of person who is actually the offender. It is readily seen that such consequences parallel with false identifications and exclusions in the context of forensic identification (or, individualization) [e.g., 7]. More generally, an outcome is defined as what would occur if one makes a decision (e.g., convicting or acquitting, identifying or excluding) given that a particular state of nature holds (e.g., the prosecution's or the defense's case being true). When expressing losses for decision outcomes numerically, and combining them with probabilities for states of nature, one obtains to the concept of expected loss. Decisions can be characterized by their expected loss, and one can use expected loss as a basis to choose among available decisions. In Bayesian decision theory, a common principle says to choose the decision with minimum expected loss.[2] The use of this principle, in a prescriptive sense, as a basis for decision is controversial in the law [e.g., [2]], a topic that is beyond the scope of this paper. Here we will solely concentrate on the analytical use of this principle for the study and review of decision problems that arise in the restricted scope of forensic science.

The concept of expected loss is considered here because, at the time when a decision must be made, the actual state of nature is not known — it is uncertain. If it would be known which state of nature holds, it would be straightforward to select the decision that is optimal under that state of nature, and there would be no necessity to approach the decision problem in a structured way. Clearly, for example, if one would know for sure that a person of interest is *not* the source of a trace found at a crime scene, not identifying that person as the source of the trace would be the optimal decision. If, however, there is uncertainty about the actual state of nature, the presumably sensible way to proceed is to consider the loss that is *expected* for each decision (e.g. considering the loss due to a missed individualization and the loss due to a false identification), and then choose the decision which has the minimum expected loss. Further development of the comparison of the expected losses of two decisions, call them d_1 and d_2, leads to the following standard decision criterion [e.g., [6]] (see Appendix):

$$\text{decide } d_1 \text{ rather than } d_2 \text{ if and only if } \quad \frac{\Pr(\theta_1 \mid I)}{\Pr(\theta_2 \mid I)} > \frac{\mathrm{L}(C_{12})}{\mathrm{L}(C_{21})}, \tag{1}$$

where θ_1 and θ_2 are the two states of nature (e.g., the competing propositions of the prosecution and defense), $Pr(\cdot \mid I)$ denotes probability conditioned on information I, and $\mathrm{L}(\cdot)$ denotes loss associated with a particular consequence C_{ij}, that is the consequence of deciding d_i $(i = 1, 2)$ when the actual state of nature is θ_j $(j = 1, 2)$. Notice that Eq. (1) supposes that correct conclusions, that is deciding d_1 when proposition θ_1 is true, and deciding d_2 when proposition θ_2 is true, have zero losses.[3] In turn, the decisions with adverse consequences C_{12}, that is wrongly deciding d_1 when in fact θ_2 holds, and C_{21}, that is wrongly deciding d_2 when in fact θ_1 holds, have non-zero losses.

A crucial insight of the decision criterion that is exemplified in Eq. (1) is that the question of 'what to decide' does not have an absolute answer, but a relative one. It is relative in the sense that one's degrees of belief, expressed in terms of the odds in favour of θ_1 over θ_2, must be compared against the ratio of the relative losses associated with the two possible ways of deciding wrongly. In particular, the prior (posterior) odds ratio on the left-hand side of Eq. (1) must

[1] From a philosophical perspective, we may add the interpretive account, which concentrates on, for example, the meaning of decision [10].

[2] Note that one can also work with utilities instead of losses. Then the criterion states that one should choose a decision that maximizes the expected utility.

[3] Note that the approach is flexible enough to consider, if required, variations to the assumption that the two correct decision consequences have identical losses. Stated otherwise, the decision-maker may consider that the results of the two ways of deciding correctly are *not* equally desirable. However, in such a case, the reader should observe Eq. (2) in the Appendix instead of the simplified Eq. (1) in Section 2.

exceed the loss ratio on the right-hand side of Eq. (1), given by the loss of the erroneous decision d_1 when θ_2 is true, $L(C_{12})$, and the loss of the erroneous decision d_2 when θ_1 is the case, $L(C_{21})$. While this is an abstract statement of a formulaic expression, with no guidance as to how to quantify the various terms, the key question thus arising is how to substantiate and exemplify Eq. (1) with respect to practical instances of forensic decisions. Addressing this question is the goal of the remainder of this paper.

A second crucial insight conveyed by the decision criterion in Eq. (1) is that one decides not only based on one's beliefs, but also based on personal evaluations of the undesirability of decisions consequences. More precisely, one's beliefs given all the information available at the time when the decision needs to be made have to be related to one's quantifications of inconvenience (or, remorse) associated with adverse decision outcomes. An important corollary of this is that any probabilistic threshold to which a decision-maker adheres when making a decision, reveals a particular expression of preferences among decision outcomes. This is a key property that will be exploited and discussed in Section 3.

Note further that the criterion Eq. (1) involves the term 'Bayesian'. The reason for this is that the state of belief at the time when the decision needs to be made is conditioned on all available information. In a fully formalised perspective, such a belief state corresponds to posterior probabilities obtained through a Bayesian inference process. Strictly speaking, thus, the probabilities $Pr(\theta \mid I)$ in Eq. (1) are actually posterior probabilities $Pr(\theta \mid I, E)$, where E denotes the available scientific evidence. Section 3 will illustrate this through examples.

3. Analysis of examples of forensic decision criteria

3.1. Kinship analysis: paternity

A common example of a legal application involving an explicit numerical decision criterion is kinship analyses, in particular where paternity is alleged. Kinship analyses are widely used in civil cases, but also in criminal cases that may revolve around rape, incest or unidentified victims of homicides. In a wider sense, victim identification after disasters, or war crimes, is also among the relevant areas of application.

Suppose a specific male is alleged to be the father of a specific child. Results are available in terms of measurements taken on genetic markers for the mother, the putative father and the child. In such a case, scientists report a likelihood ratio (also referred to as paternity index [e.g., [13]]) for the proposition that the putative father is the true father, versus the proposition that an unknown person is the father. In addition to a paternity index, scientists may also report the posterior probability for paternity, along with a statement of the assumed prior probability of paternity [13,22]. In some jurisdictions, legal decision-making is based on such a posterior probability. In Switzerland, for example, legal practice follows jurisprudence[4] according to which paternity is considered 'practically proven' if the probability of paternity is at least 0.998 (starting from prior odds of 1). Similar thresholds are applied in many other jurisdictions and kinship scenarios (e.g., in immigration cases). The example below displays such probabilistic thresholds in decision-theoretic terms.

Example 1. What does it mean for a legal decision-maker to say – in terms of quantification of inconvenience associated with adverse decision outcomes – that paternity is established when the probability of paternity $Pr(\theta_1 \mid I, E)$ is, say, at least 0.998? To answer this question, invoke relation Eq. (1) and associate θ_1 and θ_2 with the propositions of, respectively, fatherhood and nonfatherhood, to obtain the ratio $0.998/0.002 = 499$. This ratio must exceed the loss ratio $L(C_{12})/L(C_{21})$, where C_{12} refers to the wrong decision of paternity (i.e., the putative father is not the biological father), and C_{21} is the conclusion of not declaring paternity when the putative father is in fact the father of the child. The Bayesian decision criterion for a probability threshold of 0.998 thus means:

Decide paternity if and only if:

- wrongly deciding paternity is less than 499 times worse than wrongly concluding non-paternity;
- or, equivalently, the loss assigned to an erroneous decision of paternity is smaller than 499 times the loss of an erroneous decision of non-paternity.

As may be seen, what is important in the comparison with the odds for the propositions is the magnitude of the loss ratio, not the absolute loss values assigned to the adverse consequences. Notwithstanding, for the sole purpose of illustrating a numerical example, consider the common 0–1 loss scale. If one considers a wrong conclusion of paternity (to declare paternity when the alleged father is not the biological father), C_{12}, to be the worst consequence, it would receive the maximum loss 1. Then, the loss for a wrong conclusion of non-paternity, consequence C_{21}, must be greater than 1 over 499.

It is often useful to illustrate the critical loss ratio for different threshold probabilities. When the posterior probabilities are, for example, $Pr(\theta_1 \mid I, E) = \{0.95, 0.98, 0.99, 0.995, 0.999\}$ the loss ratio must be inferior to $\{19, 49, 99, 199, 999\}$ in order for d_1 to be preferable to d_2.

An example for the 95% threshold can be found in the New York Civil Practice Law and Rules Rule 4518(d), which states that if "(…) a genetic marker test or DNA test (…) indicates at least a ninety-five percent probability of paternity, the admission of such record or report shall create a rebuttable presumption of paternity, and shall, if unrebutted, establish the paternity of and liability for the support of a child (…)." Considering this decision threshold from a Bayesian decision-theoretic viewpoint thus means that deciding in favour of paternity at a probability of 0.95 is warranted if the loss ratio does not exceed 19. It is worth noting, however, that losses incurred by quite different people are involved here. Further, the loss assessment is the assessment of still a third person who is valuing how "society" is valuing these two different types of losses. This is a complex structure that is important to recognise.

It is important to remind that results of DNA analyses represent, usually, only one item of information, among others, that a decision-maker takes into account. This may lead to situations in which a court may not decide in favour of paternity even though the probability of paternity reaches a value well above, for instance, 0.99. An example for this is the case City & County of San Francisco v. Givens, 85 Cal.App.4th 51, 101 Cal.Rptr.2d 859 (Cal.Ct.App.2000). In this case, a probability of paternity of 0.9992 was found. But despite this figure, the court retained the conclusion non-paternity after taking into account evidence provided by the defendant that was qualified as 'clear and convincing', ascertaining that he had not met the mother in the relevant period of time. This conclusion is not incoherent or in conflict with the formal analysis presented above, nor does it invalidate the particular numerical figure resulting from the DNA analysis. Instead, what the court appears to have been considering at the time of making the decision, was a degree of belief taking into account *all* the information available, including the evidence presented by the defendant. The latter was strongly pointing against the hypothesis of paternity, as is clarified by the following statement: 'The court thereupon entered an order with its finding that there was "no probability" that defendant was the father.'

[4] BGE 101 II 13 *(Arrêt de la IIe Cour civile du 15 mai 1975 dans la cause C. contre F.)*.

3.2. Categorical opinion of identity of source (so-called source attribution)

An interesting insight from the previous section is that common probabilistic standards for determining paternity relate to limiting loss ratios on orders of magnitude that are in the realm of what one may readily apprehend. For example, decision-makers may find it understandable that they ought to take an interest in a question of the kind 'Is wrongly deciding paternity less than x times worse than wrongly concluding non-paternity (where x is a number on the order of several hundred)?'.

This conceptual perspective provides a point of comparison for other situations in which decisions regarding competing propositions need to be made, based on the results of forensic examinations. Areas of interest cover, for example, DNA, fingermarks, toolmarks and handwriting. In these disciplines, target questions may focus on the source of trace material, so-called source attributions: for example, 'is the person of interest the source of the DNA trace (or, fingermark, questioned signature, etc.)?'. But forensic practice in these areas is diverse. As noted by various authors [12,28], fingermark examiners find it acceptable to express categorical opinions regarding the source of fingermarks, whereas DNA analysts would not generally do so for biological traces, despite the fact that statistically more extensive data is available regarding DNA. But even in the area of DNA, intentions and arguments are diverging. In the late 1990's, for example, the FBI followed a policy that allowed scientists to report in terms of categorical conclusions that particular trace material came from a particular person, if the rarity of the DNA profile in the relevant population satisfies certain criteria [15]; recently reiterated in [20]. This can be seen as an instance of the more general – but flawed – argument according to which a low or very low probability for the event of encountering a given finding in another person (or object) from the relevant population entitles one to 'jump' to the conclusion that such an observation is actually impossible, and hence claim that a given state of nature (e.g., common source) is proved.[5] However, on a strict probabilistic account, a sound expression of an opinion about a proposition requires more than the DNA findings alone [3,4]. Moreover, some authors argue that all direct expressions of opinions, by scientists, on questions (i.e., propositions) to which the findings relate are inappropriate [25].

Despite the above observations on the scope of expert reporting, decisions will, at some point, be taken by participants of the legal process. The current analysis and discussion in this paper focuses solely on the logical considerations underlying such decisions, regardless of who will actually make those decisions and whether the formal theory acceptably describes their perception of the decision problem. One way to formulate the generic question in matters concerning source attributions is: 'What is the minimum degree of probability to be required for accepting the proposition of common source as established?' As an aside, note that from a logical point of view this is equivalent to questions asked in relation to the standard of proof in civil and criminal cases, an area in which decades of psychological research has demonstrated a broad variability in what subjects consider as the "preponderance" and "beyond a reasonable doubt" standard [14].

Example 2. To illustrate the Bayesian decision-theoretic justification of categorical source attribution conclusions, consider a case in which the conditional genotype probability (CGP) of the single profile of a crime stain is on the order of 1 in a billion (10^{-9}).[6,7] Imagine further that a person is found to have the same DNA profile as that of the crime stain. On what considerations would it be warranted – from a purely logical point of view – to consider this person to be the source of the crime stain?[8] Following the framework outlined in Section 2, there are two steps to answer this question. The first step focuses on finding the decision-maker's posterior odds. These are obtained by combining the prior odds with the likelihood ratio. In the case here, suppose that the likelihood ratio is given by 1 over the CGP, that is a billion (a $\log_{10}$(LR) of 9). The logarithm to the base ten ($\log_{10}$) is used here to deal more easily with very small and large numbers, and to take advantage of the fact that prior odds and the likelihood ratio combine by addition, rather than multiplication [1]. The prior odds express the degree of belief before considering the information regarding the corresponding DNA profiles. We shall consider the situation variously, under three different suppositions of $\log_{10}$(prior odds) at levels of 0, −3, and −6. Next, assign the likelihood ratio as 1 over the CGP, that is a billion. The combination of the prior odds and the likelihood ratio thus leads to the following $\log_{10}$ posterior odds: $\{9, 6, 3\}$. The second step amounts to comparing the latter result with the loss ratio. According to Eq. (1), deciding in favour of θ_1, that is the person of interest being the source of the crime stain, rather than θ_2, someone else being the source of the crime stain, is warranted if and only if:

$$\begin{aligned}
\log_{10}(\text{prior odds}) + \log_{10}(\text{LR}) &> \log_{10}(\text{loss ratio})\\
\{0, -3, -6\} + 9 &> \log_{10}(\text{loss ratio})\\
\{9, 6, 3\} &> \log_{10}(\text{loss ratio})
\end{aligned}$$

where the loss ratio is given by the loss of a false source attribution divided by the loss of a missed source attribution. It is interesting to note that with prior odds largely in favour of the proposition according to which an unknown person is the source of the crime stain (e.g., $\log_{10}$ prior odds −6), the posterior odds in favour of the proposition of common source is a thousand, which might be seen as insufficient to feel comfortable with concluding common source. However, in the Bayesian decision-theoretic account concluding common source with posterior odds of a thousand is only unwarranted if the loss ratio is larger than this value. Whether this is the case is a judgment in the realm of the individual decision-maker, though practitioners might object that they cannot apprehend the target values and their orders of magnitude. The difficulty here may be thinking about the problem in isolated terms, rather than with respect to a point of comparison. To help decision-makers make up their minds, and give the problem more structure, it may thus be instructive to remind them that in kinship analysis cases, decisions in favour of paternity may be reached for posterior odds that are only approximately half as large (e.g., 499 in Example 1).

Example 3. In forensic handwriting examination it often occurs that the scientist is asked, after comparing questioned writings and known items of writing from a person of interest, whether or not the person of interest is the author of the questioned handwritten text. It is often expected that handwriting examiners are in a position to

[5] See also Ref. [8] for a mathematical argument, based on a proper scoring rule, that rounding off small probabilities to 0 (or high probabilities to 1) is not optimal.

[6] Note that current profiling systems allow one to obtain figures that are smaller than 10^{-9} by several orders of magnitude [16], though reporting should be limited to 1 in a billion for several reasons, including limitations in empirical support, the feasibility of real life comparisons (which need to suppose the presence of relatives in large populations) and difficulties to conceptualize extremely small numbers. The probability of error is a further limiting factor that is usually not taken into account [e.g., 30].

[7] By analogy, the example here can also be given for fingermarks, replacing single profile by friction ridge minutiae configuration and crime stain by fingermark.

[8] For related discussion regarding the uniqueness of DNA profiles, see also Ref. [3].

provide such an opinion because the nature of the examined material, handwriting features, is intricate, and only specialist knowledge that relies exclusively with the scientist would allow one to arrive at reliable conclusions. However, in forensic handwriting examination, too, the issue of identity of source crucially depends on the prior odds, and these lie beyond the area of competence of the handwriting examiner. So, rather than providing a statement about the probability of a proposition, given the findings, handwriting examiners have turned their attention to expressing a probability for the findings, given the competing propositions, to offer guidance in terms of a likelihood ratio [27].[9] In handwriting examinations though, likelihood ratios tend to be much smaller than the figure of a billion discussed in Example 2 for DNA [21]. For the purpose of illustration, suppose that the scientist reports a likelihood ratio of a thousand ($\log_{10}$(LR) of 3). Some forensic scientists illustrate the impact of this result on different prior odds that the recipients of expert information may have, such as $\log_{10}(\text{prior odds}) = \{0, -2, -3\}$. Resulting posterior odds ($\log_{10}$) would thus be $\{3, 1, 0\}$. In this scheme, the scientist is helping the recipient of expert information revising their beliefs in the truth or otherwise of the proposition that the person of interest, rather than an unknown person, is the author of the questioned material. But this is still meeting the initial request only halfway, because the question in the first place was whether or not the person of interest is the author of the questioned handwritten text. This is a question of decision, for which the scientist may offer guidance by explaining that a rational choice requires the recipient of expert information to compare their posterior odds with their loss ratio. For example, for a likelihood ratio of a thousand, the scientist may think of reporting along the following lines:

> 'My findings are on the order of one thousand times more probable if the person of interest is the author of the questioned text than if an unknown person wrote the questioned text. Hence, whatever odds the recipient of expert information assesses that the person of interest is the author, based on other evidence, my findings multiply those odds by one thousand. For example, if the prior odds are even, then the posterior odds are one thousand, but will be less for smaller prior odds. To conclude that the person of interest is the author in the light of these odds of thousand, the ratio of the losses of adverse decision consequences must be smaller than one thousand. Specifically, falsely identifying the person of interest as the author of the questioned handwritten text must not be assessed as being more than one thousand times worse than wrongly missing to recognise the person of interest as the author of the questioned text. If the loss ratio is actually greater than one thousand, then the rational decision in this case is *not* to attribute the questioned text to the person of interest, *despite* the odds being in favour of this proposition. This happens if the prior odds are, for example, 100 or 1000 in favour of the alternative proposition (i.e., an unknown person is the author). In the latter situations, the posterior odds in favour of the proposition that the person of interest is the author of the questioned text are reduced to, respectively, 10 and 1, and hence clearly smaller than the loss ratio of one thousand.'

As noted throughout the previous examples, the importance of the analysis relies in the comparison with the loss ratio. Deciding in favour of the proposition of common source is only warranted as long as the odds in favour of this proposition outweigh the ratio of the losses associated with adverse decision consequences. This does not mean that the odds in favour of the proposition of common source must necessarily be high. To illustrate this, consider the well known particular situation in which it is considered that wrongly identifying the person of interest as the author of the questioned signature is as undesirable as wrongly missing to identify the person of interest as the author. In this case the loss ratio is one and hence deciding in favour of the proposition of common source would be warranted as soon as the probability of this proposition exceeds 0.5. This value is sometimes referred to as a translation of the concept of preponderance standard [e.g., [18]].

4. Discussion and conclusions

A main question that arises with forensic results in the legal process is how to revise one's beliefs in competing propositions of interest. Besides the ultimate issue, propositions can be of various kinds and refer to, for example, the source of forensic traces (e.g., marks or particles) or alleged activities of persons of interest. Legal reasoning, especially the question of what to believe given newly acquired information, is extensively dealt with in both specialised legal [24] and forensic science literature [1,23]. At some point, however, participants of the legal process need to act upon their beliefs, and the question of interest shifts from 'what to believe' to 'what (or, how) to decide'. These questions of decision do not arise in isolation, but are related to questions of belief. Indeed, practitioners often ask 'how sure' they ought to be before making a particular decision. It is of interest, thus, to deal with inference and decision within a coherent whole and to make the connection between these two topics explicit. Bayesian decision theory provides one such framework, and the fact that the probability part of this theory is already used for many forensic inference analyses renders it well suited for extending its use to formal analyses of decision-making. This theoretical framework is analytical and normative, rather than descriptive, in the sense discussed further below [9].

The perspective taken in this paper is that of an individual decision-maker facing a choice among courses of action with uncertainty about decision outcomes. The focus of attention is the question of what constitutes a good decision, to be defined in terms of particular criteria – so-called normative standards – by which decision-making ought to be evaluated [5]. The normative account is valuable in this respect because it allows one to state precisely what it means to make a decision in a particular state of uncertainty. This account provides a reference point against which the observable behavior of real decision-makers, or their thinking before making a decision, may be compared. Such behaviour is, in fact, highly variable, depending on a many situational factors and decision-makers' characteristics. This is evidenced in descriptive research of decision behaviour in legal contexts, such as studies of direct ratings of standards of proof (see [e.g., Ref. [14]] for a review). Such empirical evidence of diversity in opinion is not necessarily problematic, but it is clear that the purely descriptive account leaves deeper and more fundamental issues unaddressed. Examples are, 'what are the fundamental properties of the decisions we make, both in every day and professional situations?', 'how can we better understand current decision practices in terms of their quality (or, 'goodness')?', and 'how can we improve decision literacy both as individuals and as a scientific community?'. Bayesian decision theory, understood in a normative sense, provides a conceptual framework to help address these questions in an analytical sense. In this paper, we have only focused on the normative perspective, though we understand this perspective in a broader sense, alongside other perspectives. These include the mainly empirical and descriptive approaches (i.e., the study of

[9] For a contrary example, see terminology recommended by the Scientific Working Group for Forensic Document Examination (SWGDOC). Paragraph 4.1 of the SWGDOC Standard Terminology for Expressing Conclusions of Forensic Document Examiners (vers. 2013-2) specifies the following: "identification (definite conclusion of identity) — this is the highest degree of confidence expressed by document examiners in handwriting comparisons. The examiner has no reservations whatever (…) the examiner is certain, based on evidence contained in the handwriting, that the writer of the known material actually wrote the writing in question." (Section 4.1).

observable decision behaviour), and prescriptive perspectives (i.e., focusing on how to improve decision behaviour).

To further illustrate the value of the normative approach to decision-making, imagine a forensic science or law student, or a young practitioner who newly enters legal practice. How should they become proficient in sound decision-making? Some commentators hold that proficiency comes through experience, but this offers no solution for the acquiring of knowledge and understanding of decision logic principles — which are independent of empiricism. This pairs with the observation that discussants may refrain from a discourse that focuses on a formal analysis of the kind presented in this paper. As a consequence, decision-making remains obscure and inaccessible to critical review. It is questionable if current practitioners and future generations of members of the judiciary can draw any constructive input from this state-of-the-art. An open-minded approach to the normative perspective offers an alternative to this, in several respects. For one, it is constructive because it comes with modern decision support systems [29] that ease practical applications and that help users get acquainted with the underlying logic of reasoning. Further, it is operational in the sense that it provides the conceptual framework within which real problems may be analysed in a rigorous and coherent way. This provides insight for individual decision-makers which, in turn, is an important preliminary for individual decisions to be explained, justified and conveyed between participants of the legal process. This should contribute to increase decision literacy among forensic and legal practitioners.

A further key feature of the normative approach to decision is its uncompromising transparency. This is relevant in a context in which traditional discussions on standards of proof in terms of single numerical values appear to be without solution, for reasons mentioned above (see also [14]). We have exemplified in this paper that commonly used decision thresholds, despite their arbitrary appearance, have a clearly defined meaning in their Bayesian decision-theoretic (re-)formulation. We have emphasized, for example, that such thresholds are already used in current practice (e.g., in disputed kinship cases), though not formulated explicitly with terminology from Bayesian decision theory. We have also argued that this underlying decision logic provides fundamental insight for other decision problems across forensic disciplines. Practically, this means that discussions can avoid controversial arguments that are prone to lead to impasses. A typical example for a recurrent controversial argument is the minimum degree of personal probability required before making a decision. Instead, discussions could emphasize decision-makers' preferences among decision outcomes that the acceptance of a particular decision threshold inevitably entails. It will then become possible to have an open discourse on how those preferences ought to be framed, depending on the features of the case at hand. The crucial take-home message from this is that any decision in the light of a particular state of belief can be reconstructed in terms of the formal theory emphasized in this paper. This means that, stated otherwise, that our choices reveal our preferences, independently of whether one endorses the theory or not.

Acknowledgments

This research was supported by the Swiss National Science Foundation through Grant No. BSSGI0_155809 and the University of Lausanne. Alex Biedermann gratefully acknowledges the comments received from Professor George F. Sensabaugh and Dr. Charles Brenner during a visiting research stay at University of California, Berkeley.

Appendix A. Bayesian decision criterion

Suppose a case in which the issue is whether a fingermark comes from a given person of interest (POI), or from some unknown person. Let these two competing versions be the relevant states of nature, denoted θ_1 and θ_2. Further, the decision-maker holds probabilities $Pr(\theta_1 \mid I)$ and $Pr(\theta_2 \mid I)$ for these states of nature, conditioned on the information I. In a more general view, it is also common to regard θ_1 and θ_2 as the cases of the prosecution and the defense, respectively. Next, let d_1 and d_2 be the decisions to conclude that the fingermark comes from, respectively, the POI or from some unknown person. When deciding d_i and the state of nature θ_j holds, consequence C_{ij} arises. The loss associated with the consequence C_{ij} is written $\mathrm{L}(C_{ij})$. Combining losses and probabilities leads, for each decision d_i, to the notion of expected loss EL, defined as follows: $\mathrm{EL}(d_i) = \sum_j \mathrm{L}(C_{ij})\,\mathrm{Pr}(\theta_j \mid I)$. One way to choose among the various decisions d_i is to consider their expected loss, and select the decision with the minimum expected loss. This is known as the principle of minimizing expected loss. Formally, this criterion says that to decide d_1 (identifying the POI) rather than d_2 (not identifying the POI) if and only if

$$\mathrm{L}(C_{11})\,\mathrm{Pr}(\theta_1 \mid I) + \mathrm{L}(C_{12})\,\mathrm{Pr}(\theta_2 \mid I) < \mathrm{L}(C_{21})\,\mathrm{Pr}(\theta_1 \mid I) + \mathrm{L}(C_{22})\,\mathrm{Pr}(\theta_2 \mid I), \tag{2}$$

that is $\mathrm{EL}(d_1) < \mathrm{EL}(d_2)$. To simplify this expression, on can invoke a standard $(0, k)$ loss function, where 0 is the loss assigned to the most favourable consequence(s), and k is the loss assigned to the least favourable consequence(s). Here, deciding d_1 when proposition θ_1 is true (consequence C_{11}), and deciding d_2 when proposition θ_2 is true (consequence C_{22}), are correct conclusions and have zero losses. The zero losses $\mathrm{L}(C_{11})$ and $\mathrm{L}(C_{22})$ reduce the expected losses $\mathrm{EL}(d_1)$ and $\mathrm{EL}(d_2)$ in Eq. (2) to, respectively, $\mathrm{L}(C_{12})\mathrm{Pr}(\theta_2 \mid I)$ and $\mathrm{L}(C_{21})\mathrm{Pr}(\theta_1 \mid I)$. Rewriting Eq. (2) leads to (see also, for example, Ref. [6]):

$$\frac{\mathrm{Pr}(\theta_1 \mid I)}{\mathrm{Pr}(\theta_2 \mid I)} > \frac{\mathrm{L}(C_{12})}{\mathrm{L}(C_{21})}.$$

This result shows that, under the chosen type of loss function, the decision criterion involves the comparison between, on the one hand, the odds in favour of θ_1 against θ_2, and on the other hand, the ratio of the losses associated with the two adverse decision consequences.

References

[1] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd ed. ed., John Wiley & Sons, Chichester, 2004.
[2] R.J. Allen, The error of expected loss minimization, Law Prob. Risk 2 (2003) 1–7.
[3] D.J. Balding, When can a DNA profile be regarded as unique? Sci. Justice 39 (1999) 257–260.
[4] D.J. Balding, C.D. Steele, Weight-of-Evidence for Forensic DNA Profiles, 2nd ed., John Wiley & Sons, Chichester, 2015.
[5] J. Baron, Thinking and Deciding, 4th ed. ed., Cambridge University Press, New York, 2008.
[6] J.M. Bernardo, A.F.M. Smith, Bayesian Theory, 2nd ed ed., John Wiley & Sons, Chichester, 2000.
[7] A. Biedermann, S. Bozza, F. Taroni, Decision theoretic properties of forensic identification: underlying logic and argumentative implications, Forensic Sci. Int. 177 (2008) 120–132.
[8] A. Biedermann, P. Garbolino, F. Taroni, The subjectivist interpretation of probability and the problem of individualisation in forensic science, Sci. Justice 53 (2013) 192–200.
[9] A. Biedermann, F. Taroni, C. Aitken, Liberties and constraints of the normative approach to evaluation and decision in forensic science: a discussion towards overcoming some common misconceptions, Law Prob. Risk 13 (2014) 181–191.
[10] L. Buchak, Decision Theory, in: A. Hájek, C. Hitchcock (Eds.), Oxford Handbook of Probability and Philosophy, Oxford University Press, Oxford, 2016.
[11] F.C. Dane, In search of reasonable doubt: a systematic examination of selected quantification approaches, Law Hum. Behav. 9 (1985) 141–158.
[12] I.W. Evett, B.S. Weir, Interpreting DNA Evidence, Sinauer Associates Inc., Sunderland, 1998.

[13] D. Gjertson, C. Brenner, M. Baur, A. Carracedo, F. Guidete, J. Luquef, R. Lessigg, W. Mayrh, V. Pascali, M. Prinz, P. Schneider, N. Morling. ISFG: Recommendations on biostatistics in paternity testing, Forensic Sci. Int. Genet. 1 (2007) 223–231.
[14] R. Hastie, Algebraic Models of Juror Decision Making, in: R. Hastie (Ed.), Inside the Juror. The Psychology of Juror Decision Making, New York, Cambridge University Press, 1993, pp. 84–115.
[15] C. Holden, DNA fingerprinting comes of age, Science 278 (1997) 1407.
[16] A. Hopwood, R. Puch-Solis, V. Tucker, J. Curran, J. Skerrett, S. Pope, G. Tully, Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts, Sci. Justice 52 (2012) 185–190.
[17] D. Kahneman, Thinking, Fast and Slow, Penguin, London, 2011.
[18] D.H. Kaye, Clarifying the burden of persuasion: what Bayesian decision rules do and do not do, Int. J. Evid. Proof 3 (1999) 1–29.
[19] D.V. Lindley, Understanding Uncertainty, revised edition ed., John Wiley & Sons, Hoboken, 2014.
[20] T.R. Moretti, B. Budowle, Letter to the Editor - Reiteration of the statistical basis of DNA source attribution determinations in view of the Attorney General's Directive on "Reasonable Scientific Certainty" statements, J. Forensic Sci. 62 (2017) 1114–1115.
[21] A. Nordgaard, B. Rasmusson, The likelihood ratio as value of evidence — more than a question of numbers, Law Prob. Risk 11 (2012) 303–315.
[22] L. Poulsen, S. Friis, C. Hallenberg, B. Simonsen, N. Morling, A report of the 2009–2011 paternity and relationship testing workshops of the English Speaking Working Group of the International Society For Forensic Genetics, Forensic Sci. Int. Genet. 9 (2014) e1–e2.
[23] P. Roberts, C.G.G. Aitken, The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science (Practitioner Guide No. 3), Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses, Royal Statistical Society's Working Group on Statistics and the Law, 2013.
[24] P. Roberts, A. Zuckerman, Criminal Evidence, 2nd ed. ed., Cambridge University Press, Oxford, 2010.
[25] B. Robertson, G.A. Vignaux, C.E.H. Berger, Interpreting Evidence. Evaluating Forensic Science in the Courtroom, 2nd ed ed., John Wiley & Sons, Chichester, 2016.
[26] R. Simon, L. Mahan, Quantifying burdens of proof: a review from the bench, the jury, and the classroom, Law Soc. Rev. 5 (1971) 319–330.
[27] A. Stockton, S. Day, Bayes, Handwriting and Science, Proceedings of the 59th Annual ASQDE Meeting — Handwriting & Technology: at the Crossroads, Des Moines, Iowa, U.S.A, 2001, pp. 1–10.
[28] D.A. Stoney, What made us ever think we could individualize using statistics? J. Forensic Sci. Soc. 31 (1991) 197–199.
[29] F. Taroni, A. Biedermann, S. Bozza, G. Garbolino, C.G.G. Aitken, Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science. Statistics in Practice, 2nd ed. ed., John Wiley & Sons, Chichester, 2014.
[30] S. Zabell, Book review: statistical DNA forensics: theory, methods and computation, by Wing Kam Fung and Yue-Qing Hu, Law Prob. Risk 11 (2012) 105–110.