

A formal approach to qualifying and quantifying the ‘goodness’ of forensic identification decisions

A. Biedermann^{a,*}, S. Bozza^{b,a}, F. Taroni^a, P. Garbolino^c

^aUniversity of Lausanne, School of Criminal Justice, 1015 Lausanne-Dorigny, Switzerland

^bCa' Foscari University of Venice, Department of Economics, 30121 Venice, Italy

^cUniversità Iuav di Venezia, Department of Architecture and Arts, 30135 Venice, Italy

Abstract

In this paper we review and analyse common understandings of the degree to which forensic inference of source – also called identification or individualisation – can be approached with statistics and is referred to, increasingly often, as a decision. We also consider this topic from the strongly empirical perspective of PCAST [32] in its recent review of forensic science practice. We will point out why and how these views of forensic identification as a decision, and empirical approaches to it (namely experiments by multiple experts under controlled conditions), provide only descriptive measures of expert performance and of general scientific validity regarding particular forensic branches (e.g., fingerprint examination). Although relevant to help assess whether the identification practice of a given forensic field can be trusted, these empirical accounts do not address the separate question of what ought to be a sensible, or ‘good’ in some sense, (identification-)decision to make in a particular case. The latter question, as we will argue, requires additional considerations, such as decision-making goals. We will point out that a formal approach to qualifying and quantifying the relative merit of competing forensic decisions can be considered within an extended view of statistics in which data analysis and inference are a necessary but not sufficient preliminary.

Keywords: Forensic individualisation, Decision analysis, Decision theory, Minimax, Decision-theoretic criterion

It is not uncommon to hear of nonsense about conclusions being “scientifically proved.” [10, at p. 10]

1. Introduction: is forensic identification (only) a case of statistics?

The mindset of many forensic scientists associates forensic identification of source with statistics. According to this view, forensic identification of source – also sometimes referred to more shortly as identification or individualisation¹ [6] – can benefit from statistics. A more stringent view would consider that forensic identification *is*, by its nature, a statistical procedure. This, however, might be criticised as being a restrictive view. It is interesting to note in this context that at the first ICFIS conference, held in Edinburgh (April 2nd–4th, 1990), then running under the shorter title *International Conference on Forensic Statistics*, Dennis Lindley noted: “Our topic might better be described as ‘forensic inference’, rather than ‘forensic statistics’, since not all evidence is based on statistics (or data)” [24, p. 83]. According to Lindley’s account, the questions that arise in forensic science invoke, but do not reduce to statistics. This

*Corresponding author. – Submission for the conference theme ‘Statistical Foundations of Forensic Identification of Source’ of the 10th International Conference on Forensic Inference and Statistics, September 5th–8th 2017, South Dakota State University, Minneapolis, Minnesota (USA).

Email address: alex.biedermann@unil.ch (A. Biedermann)

¹Historically, individualization has been used as a term for the forensic source attribution process [21], though practice developed a preference for the term identification. Strictly speaking, however, an identification means the assignment of an item to a particular class or set [19]. We will use the terms identification and individualisation interchangeably throughout this paper, but with a preference for the term individualisation.

connects to the view expressed by Evett according to whom “people call this [the interpretation of evidence] statistics, but it is not actually statistics, it is inference, it is reasonable reasoning in the face of uncertainty.”²

In the above quote, Lindley uses the term ‘statistics’ in the focused sense of analysing data and using them to help people revise their beliefs about propositions about which they are uncertain. On other occasions, Lindley favoured a broader definition that involves data analysis only as a first step: “(...) the first task of a statistician is to develop a (probability) model to embrace the client’s interests and uncertainties. It will include the data and any parameters that are judged necessary. Once accomplished, the mechanics of the calculus take over and the required inference is made” [25, p. 318]. Inference, however, is not the end of the matter; Lindley also noted that “[i]f decisions are involved, the model needs to be extended to include utilities, followed by another mechanical operation of maximizing expected utility. One attractive feature is that the whole procedure is well defined and there is little need for ad hoc assumptions” [25, p. 318]. Lindley refers here to the logical extension of probability to decision theory, which is a point that is systematically found in his writings [e.g., 23, 26]. In legal literature, decision theory has already been raised some 50 years ago [17]. At the time, decision theory was in the course of becoming a core area in statistics, as noted in [10]:

“Years ago a statistician might have claimed that statistics deals with the processing of data. As a result of relatively recent formulations of statistical theory, today’s statistician will be more likely to say that statistics is concerned with decision making in the face of uncertainty. Its applicability ranges from almost all inductive sciences to many situations that people face in everyday life when it is not perfectly obvious what they should do.” [at p. 1]

In this paper, we will discuss forensic individualisation from the point of view of this extended, decision-oriented definition of statistics. The reasons for this are twofold. The first reason is that forensic individualisation cannot be resolved by inference alone, as argued previously by [33].³ The second reason is that the consideration of forensic individualisation as a decision – using elements of decision theory – is a novel, contemporary development and raises new points of discussion [11], mainly regarding applicability. We will consider several of these discussion points in this paper, but leave more technical developments [e.g., 3, 4] aside, except for short summaries in Appendices A and B. Specifically, we will analyse and discuss the conceptual insights that derive from the proposed decision-theoretic perspective with respect to conclusions and recommendations of [31] in its recent review of current forensic science practice, especially PCAST’s call for an empirical approach to scientific validity – a call reiterated recently in an AAAS report [37].

This paper is organised as follows. Section 2 will address the following points:

- the distinction between formal properties of decision theory and accounts that focus on the empirical study of how people actually make decisions (Section 2.1);
- formal approaches to decision analysis in individual cases with and without taking into account uncertainty through probability (Sections 2.2 and 2.3).

This outline will allow us to state and comment on the formal properties of decision theory – the theory focusing on the logic of decision – in a broad perspective, clarifying conceptual demarcations with respect to other decision approaches. In particular, we will point out in what sense these properties allow one to define *precisely*, qualitatively and quantitatively, what may generically be called the ‘goodness’ of decisions. The latter term is, otherwise, often used in a more colloquial sense, and hence left undefined. Discussion and conclusions are presented in Section 3, with an emphasis on the questions that arise when relating the formal concepts of decision theory to the applied context of forensic individualisation. We will argue that our decision-theoretic account of forensic individualisation is not intended to constrain forensic practice, but to provide all discussants with a powerful analytic instrument to favour insight, bettering the understanding and help articulate what is at stake in the forensic individualisation process.

²Acceptance speech of I. W. Evett for his doctorate honoris causa, received on the occasion of the 100th anniversary of the School of Criminal Justice, University of Lausanne, June 24th 2009. Text in brackets added by the authors.

³On the same position, see [8] for a recent review, and Recommendation 3.7 of the report of the [15]: “Because empirical evidence and statistical reasoning do not support a source attribution to the exclusion of all other individuals in the world, latent print examiners should not report or testify, directly or by implication, to a source attribution to the exclusion of all others in the world.” (at p. 72)

2. Formal properties of forensic decisions and forensic decision analysis

2.1. Distinguishing decision empiricism from the logic of decision

The notion of decision has become a popular word in forensic science, not least because some branches have given it the prominent role of labelling its conclusions. For example, experts in forensic identification or individualisation (e.g., of fingerprints) now refer to their conclusions as individualisation *decisions*.⁴ While this provides evidence of the field's continuous endeavour to rethink its principles and practice, critical commentators [e.g., 11] argue that the renaming of friction ridge mark examiners' conclusions as decisions was a mere change of label, leaving aside fundamental developments in the understanding of expert conclusions in formal decision theoretic terms [e.g., 3]. Although this is a trenchant criticism, it should not be interpreted as a discouragement of the promotion of research in this area.

We can distinguish two aspects related to forensic uses of the word decision, referred here to as decision empiricism and the logic of decision, respectively. Below, we provide a brief sketch of them in turn (see also Table 1 for a summary). Decision empiricism, as it is understood here, places a strong focus on, first, recognising that what forensic expert reporting ultimately amounts to is decision-making. Second, decision empiricism focuses on studying – by observation – how human experts actually make decisions. The ‘how’ of observable decision making is explored, however, in variable depths and often, the study reduces to merely recording factual output without actually trying to understand how the scientists arrived at given outputs (i.e., their reported conclusions). For example, in the most basic form, empirical study involves the ‘testing’ of a procedure (under controlled conditions), as applied by a scientist, and record the scientist's performance during this task. This is typically done when developing a new method or technique. On a more elaborate level, decision empiricism extends to a behavioural perspective that has led to novel avenues of interdisciplinary research with areas such as cognitive (neuro-)science [12], generating new and valuable knowledge about the conditions and factors that influence human decision-making. Broadly speaking, such research aims to improve decision-making in the sense of helping people making the ‘right’ decisions, and hopefully more often. An idea underlying this perspective is that the expert decision tasks refer to determinations⁵ (e.g., individualising or not a person of interest) for which there is a ground truth against which particular expert conclusions can be compared [39]. The crucial point here is that such research strives to characterise the performance of experts⁶ across multiple instances – under controlled conditions – of a particular forensic expert task (e.g., fingerprint examination).⁷ Clearly, such information is valuable to assess questions such as whether experts, including the methods and techniques they use, are able to do what they claim to do, and whether experts are better in their area of expertise than lay persons.⁸ The answer to such questions will also help assess whether an expert ought to be trusted and be assigned to a particular scientific task in a given case of interest. However, and this is an essential point, the overall performance of an expert – as monitored across repeated exercises (tests) under controlled conditions – is not a direct measure of the ‘goodness’ of a particular decision (to be) made in a given case at hand. Stated otherwise, general performance indicators of an expert, including any technical system or device that is used during the analysis process, does not answer the question of how ‘good’ – or accurate (when an underlying truth-state can be considered) – an individual decision is. Expert performance measures also do not answer the question, or instruct, how to actually make a ‘good’ decision in any individual case that an examiner may face. These questions fall into the domain of the logic of decision, as outlined below.

⁴See, for example, the SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology) ‘Guideline for the Articulation of the Decision-Making Process for the individualisation in Friction Ridge Examination (Latent/Tenprint)’, https://www.nist.gov/sites/default/files/documents/2016/10/26/swgfast_articulation_1.0_130427_1.pdf (page last visited April 15th 2017): “individualisation is the decision by an examiner that there are sufficient features in agreement to conclude that two areas of friction ridge impressions originated from the same source.” (Paragraph 10.2.2.) The recent AAAS report [37] also systematically refers to examiners' conclusions as decisions.

⁵There are many problems associated with the traditional practice of reporting individualisation conclusions, especially where they amount to categorical conclusions. A discussion of these problems is beyond the scope of this paper, but see, for example, [7] for a recent review.

⁶This may include the whole examination process, extending also to aspects such as the chain of custody.

⁷This leads to data that can be summarised in terms of standard diagnostic performance measures, such as rates of false positives and false negatives, and related notions such as sensitivity and specificity – widely used, for example, in the medical literature.

⁸Note also that studies “(...) in which many examiners render decisions about many independent tests (typically, involving “questioned” samples and one or more “known” samples) and the error rates are determined” [31, at p. 5–6] are considered to be part of the requirements to establish scientific validity.

Table 1: Summary of main differences between the notions of decision empiricism and the logic of decision.

Decision empiricism	Logic of decision
<ul style="list-style-type: none"> • Empirical study of human decision behaviour • Comparison with ground truth (controlled conditions) • Focus on performance across multiple exercises (decisions) • External descriptive viewpoint on experts (or a method) 	<ul style="list-style-type: none"> • Focus on the logical structure of the decision problem • Unknown state of nature (ground truth) • Focus on the relative ‘merit’ of the individual decision • Analytical viewpoint of the individual decision-maker

The logic of decision is the second aspect related to the forensic use of the word decision that we consider here (Table 1). Instead of taking an external point of view as in the empirical approach mentioned above (i.e., looking at how people behave in the face of particular decision tasks), the logic of decision reverses the viewpoint and focuses on a personalized perspective: it asks you (the reader, us, or any other person in their own way) to imagine yourself in the position of an individual decision-maker facing the problem of choosing one of a list of feasible decisions, also sometimes called alternative courses of action. Stated otherwise, instead of looking at multiple decisions made by one or several experts, and comparing each of these decisions (e.g., expert conclusions in fingerprint examination) with the respective ground truth, the logic of decision relates to decision-making under uncertainty⁹ in the individual case. It asks the question: given the basic elements of your (ours, anybody’s) decision problem, what is the best possible course of action to choose? A defining feature of the kind of situations dealt with here is that the true state of affairs is unknown, and not observable, which introduces uncertainty about the decision outcome to be obtained. For example, in real cases, it cannot be known with certainty whether a given person of interest, or an unknown person, is the source of an examined fingerprint. By making an individualisation decision in such a situation we must accept that, potentially, our conclusion amounts to a false identification. The question of interest thus becomes: how can a decision-maker – in any given case – qualify¹⁰ and compare the relative ‘goodness’, or merit, of rival courses of action when their outcomes cannot be known with certainty? Later sections of this paper will further elaborate on this question. We will present and compare formal ways to approach it.

The above two accounts of the notion of decision are neither original nor new. They are elementary and largely uncontested, yet – interestingly – ongoing discussions in forensic science and the law, in many instances, do not properly distinguish between the two perspectives. Much of current controversies over the thoroughness (e.g., validity) of forensic science, as evidenced for example by the reports of PCAST [31], and previously the NAS [28], focus on aspects of decision empiricism – notably performance in multiple experiments under controlled conditions. Posing this challenge to forensic science theory and practice is laudable and necessary, but it is not the end of the matter, because it deals only with a general measure of how trustworthy an expert and/or an employed methods are/is. As noted above, it is a completely different matter to devise a coherent approach to decision-making in an individual case, and to qualify the relative merit of rival decisions available in a particular instance. Forensic literature on this topic is rather scarce.¹¹

Throughout this paper, we will argue that decision empiricism and the logic of decision are related, and that the former is an essential,¹² but not a sufficient preliminary for the latter. Sections 2.2 and 2.3 outline, stepwise, features of standard approaches for decision analysis in the individual case.

⁹On the notion of uncertainty as it is understood in this paper, see also footnote 13.

¹⁰Term ‘qualifying’ is used here to express a *qualitative* preference, that is an assertion that is less strong than one based on quantitative terms. For example, one decision may be preferred to another decision *qualitatively*, i.e. without quantifying how much one decision is preferable compared to another. The term ‘qualifying’ can also be understood as the attribution of a quality to a decision, e.g. specifying the conditions under which one decision can be *qualified* as being preferable to (or, in more general language, ‘qualified as better than’) another decision.

¹¹We leave aside here categorical conclusion schemes of the kind ‘if A is observed, then conclude B’ because they often violate logical considerations or are unbalanced (i.e., they do not consider at least one alternative proposition). As an example, see applications in forensic toxicology, such as the analysis of alcohol markers in hair [20], where results of analyses are compared against cut-off values in order to reach a predefined conclusion.

¹²On the importance of general performance measures, see also Lander’s reply to the ANZFSS: “The reliability of evidence presented in courts

Table 2: Decision and loss matrices with d_1 and d_2 denoting, respectively, the decisions to individualise and not to individualise a suspect. The states of nature θ_1 and θ_2 denote, respectively, the proposition that the person of interest is the source of the trace and that an unknown person is the source of the trace. The notation $L(C_{ij})$ with $i, j = \{1, 2\}$ represents the loss associated with the consequence C_{ij} , that is when taking decision d_i and the state of nature θ_j holds.

		States of nature: The trace comes from	
		... the person of interest (θ_1)	... an unknown person (θ_2)
Decision matrix:	individualise (d_1)	C_{11}	C_{12}
	do not individualise (d_2)	C_{21}	C_{22}
Loss matrix:	individualise (d_1)	$L(C_{11})$	$L(C_{12})$
	do not individualise (d_2)	$L(C_{21})$	$L(C_{22})$

2.2. Decision-making in the individual case (I): deciding *without* considering uncertainty about states of nature

Recognising that the study of how to coherently approach decision-making in the individual case is a topic worthy of investigation, we now turn our attention to different decision methods. As a first approach, consider one that invokes a limited number of elements, namely decisions, states of nature and a measure for the desirability or undesirability of decision consequences. Note, however, that for the time being, we leave aside an important feature of our decision problem, that is uncertainty – expressed by probability¹³ – about the states of natures. The method to be considered in this section is known as the minimax method [e.g., 10]. This decision rule has wide developments in game theory but they will not be considered here, mainly because the assumption of an active adversarial player – who will react upon the decision maker’s choice – does not correspond to a feature of our decision problem. Here we consider only the non-probabilistic minimax decision rule for individual decisions. While we do not endorse this method for the problem of individualization, for reasons explained in due course, the method is useful to provide a formal description of some but not all features of a decision problem as seen from the viewpoint of the individual decision-maker. It will help us make a series of prima facie reasonable formalisations that we can use to blend the reality of forensic decision-making.

We use the term ‘state of nature’ here as a synonym for proposition or, more generally, one version of an event of interest. Suppose two propositions, denoted θ_1 and θ_2 , short for the propositions that a given suspect is the source of a trace found on a crime scene, and that an unknown person is the source, respectively. Let the decisions (i.e., available courses of action) be d_1 , individualise the person of interest, and d_2 , do not individualise the person of interest.¹⁴ This leads to the basic representation in terms of a decision matrix (Table 2), with C_{ij} denoting the consequence of deciding d_i when the state of nature θ_j holds.

As a last ingredient of the analysis, qualify decision consequences C_{ij} in terms of their undesirabilities, by using the concept of loss, written $L(C_{ij})$ for short. In particular, assume what is known as a 0 – 1 loss function, and assign 0 losses to a correct individualisation (i.e., consequence C_{11}) and a correct non-individualisation (i.e., consequence C_{22}). This translates the idea that nothing is ‘lost’ by making accurate decisions (e.g., deciding d_1 , individualising the person of interest when in fact this person is the source of the trace, θ_1). Further, assign the loss 1 to the worst consequence¹⁵ consequence, C_{12} , that is deciding d_1 , identifying the person of interest, when in fact θ_2 is true (i.e., an unknown person is the source of the trace). This leaves us with the consequence C_{21} , a missed individualisation. Assume that this consequence is less adverse than a false individualisation (C_{12}), so that we can assign a loss $L(C_{12}) < 1$. A qualitative assignment of this kind will suffice to pursue our discussion.¹⁶ Table 3 summarises all loss

primarily depends on whether the underlying forensic feature-comparison method has been empirically shown to be scientifically reliable – and only then on the details of the case.” [22, p. 2]

¹³Note that in some branches, such as operational research, decision analysis with quantification of uncertainty about states of nature, using probability, is known as decision analysis ‘under risk’, and decision analysis without taking into account probability is referred to as decision ‘under uncertainty’. In this paper, we use the expression ‘decision analysis under uncertainty’ for situations in which probability is taken into account [e.g., 26]. The reason for this is that uncertainty is quantified by probability and, in its personalist interpretation, is – by definition – always available to the decision-maker.

¹⁴In a more general development, decision d_2 could be broken down to decisions ‘inconclusive’ and ‘exclusion’ [e.g., 3].

¹⁵Note that a different scale may be chosen for the loss function.

¹⁶See, for example, [3, 4] and [35] for further discussion on the elicitation of loss values for a missed individualisation.

assignments.

To select a course of action, the minimax method proceeds as follows. Start by highlighting, for each decision, d_1 and d_2 , the worst consequence. In Table 3 this is shown with circles around the highest loss values that may incur for each decision. In case of decision d_1 , individualising the person of interest, this is the loss $L(C_{12}) = 1$ for a false identification. For decision d_2 , not individualising the person of interest, this is the loss $L(C_{21}) < 1$ for a missed individualisation. Next, the minimax method says to choose the decision that minimises the maximum loss over the different states of nature. Stated otherwise, the method says to look at the worst consequence(s) associated with each decision, and then choose the decision(s) that implies the minimum loss if the worst case occurs.

Table 3: Loss matrix with d_1 and d_2 denoting, respectively, the decisions to individualise and not to individualise a suspect. The states of nature θ_1 and θ_2 denote, respectively, the proposition that the person of interest is the source of the trace and that an unknown person is the source of the trace. Maximum losses for each decision are highlighted with a circle.

Decisions		States of nature: The trace comes from	
		... the person of interest (θ_1)	... an unknown person (θ_2)
Loss matrix:	individualise (d_1)	0	1
	do not individualise (d_2)	< 1	0

Applying this method to our running example of forensic individualisation would mean to always decide d_2 , not individualising the person of interest. Proceeding in this way would ensure that, in case one's determination is erroneous, the loss incurred would be minimal. It is minimal with respect to the loss associated with the worst case (consequence C_{12}) of the alternative decision, d_1 , set to 1.

Is this method a feasible approach for forensic individualisation? Intuitively, it may seem attractive to say that one strives to ensure that decision-making avoids excessive losses, but for forensic purposes, following this principle would be the end of individualisation. Completely avoiding the possibility of incurring the worst consequence can only be achieved by never selecting the individualisation decision d_1 . But, this does not map to current decision policies that consider individualisations among the admissible conclusions.

Indeed, forensic practitioners *do* reach individualisations, and they do so – presumably – when there is a high probability that the person of interest is the source θ_1 , given the information I available. The standard argument being that if $\Pr(\theta_1 | I)$ is high, there is a high 'prospect' of the consequence being C_{11} , a correct individualisation. Since this consequence has a minimum loss, the decision to individualise (d_1) does not appear to be an unreasonable course of action. Conversely, the argument also says that even though the decision to individualise d_1 may lead to a false identification – that is the worst consequence (and hence incurring the highest loss) – the decision d_1 may be considered acceptable because the probability for the person of interest not being the source of the trace, $\Pr(\theta_2 | I)$, is (sufficiently) low.

It is readily seen that in the above paragraph, we have extended our considerations about decision-making to a further element that is not part of the decision method considered so far in this section: the probabilities for the states of nature. In essence, probabilities translate the extent of available information about propositions of interest. We can thus see that the minimax method would lead us to ignore information and even make the quest for and coherent use of information – which is the heart of forensic science¹⁷ and the legal process – meaningless: whatever the amount of information that we could gather in support of the person of interest being the source of the trace (proposition θ_1), compared to the alternative proposition of an unknown person being the source of the trace (proposition θ_2), the minimax method would instruct us not to individualise.

As an intermediate conclusion we can thus retain that one's preferences among decision consequences (i.e., what one prefers) is not the only essential feature of decision-making. It is equally essential to consider one's beliefs about the states of nature (i.e., what one thinks is the case), conditioned by all the information I available at the time when a

¹⁷Especially forensic inference and statistics.

decision needs to be made. A relevant question thus is how we can make this intuition formally precise. The next section will deal with this topic in more detail (see also Appendix A for technical developments). This intermediate conclusion is not meant, however, to dismiss the minimax method in principle. It may find consideration in contexts where forensic scientists either are not able, for practical reasons, to get acquainted with the relevant information for assessing their personal probabilities, or they are not allowed, in principle, to have access to such information, so that they must take a decision under a ‘veil of ignorance’, so to speak (see Appendix B for an example).

2.3. *Decision-making in the individual case (II): decision taking into account uncertainty about states of nature*

In the previous Section we have considered a decision criterion that is based on the highest loss that may incur for each decision. However, we have concluded that this may be unsatisfactory to the extent that there is no clear acknowledgment of the fact that there is more than one potential outcome associated with each decision, and that we have case specific probabilities and losses for those outcomes. Taking these elements a step further leads to the idea of weighing, for each decision, the loss of each consequence (not only a decision’s worst consequence) by the probability of its occurrence, and summing these weighted losses. The result is known as the mathematical concept of expected loss and provides a formal criterion for the ‘goodness’ of a decision: it is a criterion that allows us to quantify the loss we ‘expect’ from a particular decision. In our case, we work with expected losses, rather than utilities (for which one can also compute the expectation), because we chose to characterise decision consequences in terms of losses. Further, since we can consider the expected loss for each decision, we have a criterion that allows us to compare different decisions, and then make a choice based on the result of this comparison. The decision rule here is to select the decision that offers the minimum expected loss. Note that this is a difference with respect to the minimax rule discussed in the previous Section which considers the decision that ensures minimal loss in the worst case.

While the minimax method can be applied easily in the contexts considered here, because it only requires one to look at the loss matrix and sort out the appropriate decision, the expected loss method is slightly more technical, but only at first sight. The reason for this is that it involves sums of probability-weighted losses. These may not be readily obtained without computational effort. However, as shown in the development in Appendix A, by choosing a suitable loss function and keeping the decision space limited (here a 2×2 decision table) it is possible to confine the considerations to a comparison between the ratio of probabilities for states of nature (i.e., odds) and the ratio of losses associated with adverse consequences. Specifically, applied to our example of forensic individualisation, the criterion given in Appendix A (Equation 2) says to choose decision d_1 (individualise) if and only if the odds in favour of the person of interest being the source of the trace are greater than the ratio of the loss of a false identification, $L(C_{12})$, and the loss of a missed individualisation, $L(C_{21})$. In numerical terms, this criterion thus says that when the odds in favour of the person of interest, rather than an unknown person, being the source of the trace is – for example – a thousand, then the decision to individualise (d_1) is optimal if and only if the loss associated with a false individualisation is *less* than a thousand times greater than the loss associated with a missed individualisation. This way of stating the decision criterion has the advantage of avoiding technicalities deriving from determining expected values. Instead, thinking is directed towards comparing orders of magnitudes of, on the one hand, probabilities, and on the other hand, of losses associated with decision consequences.¹⁸

The above decision-theoretic account – also called the Bayesian decision theoretic criterion (see Appendix A) – can be interpreted in the wider sense of capturing the essence of the common saying ‘the more is at stake the more you ought to be convinced (or, sure)’, identifying ‘stakes’ with the relative losses of adverse decision consequences and ‘being convinced’ with the odds in favour of the proposition of interest (here, the proposition that the person of interest is the source of the trace). By invoking the Bayesian decision-theoretic criterion, we merely formalize quantities and their relationships that we are already using. It provides us with a framework to clearly articulate, with rigorous and precise terms, the inevitable conceptual issues that underlie forensic individualisation. Specifically, this account shows that probabilities regarding the propositions of interest (i.e., the person of interest versus an unknown person being the

¹⁸Note, however, that if the decision problem is more complex, in particular when the number of available courses of action is greater than two (e.g., one may consider a third decision d_3 , such as reporting ‘inconclusive’) and/or a loss function not necessarily bounded between 0 and 1, the decision-theoretic criterion does not simplify to the above-mentioned comparison in terms of odds and loss ratio (see Also Appendix A), and some additional computational effort is required.

source of the trace) are a relevant consideration, though not in isolation, but in an interaction with the decision-maker's preferences among adverse decision consequences.

3. Discussion and conclusions

It has become increasingly popular to refer to forensic conclusions reported by scientists as decisions. For example, the report of the [31] uses expressions such as “examiners’ decision” [p. 9], “identification decisions” [p. 100] and “experts’ decisions” [p. 101]. In scholarly work, the term “forensic decision making” [13] can be found. Once it is admitted that expert conclusions are decisions, the natural consequence seems to be that forensic scientists – including the methods and techniques they are using – should be subjected to empirical testing, in order to see how ‘good’ scientists are in their decision-making. Specifically, the [31] report concludes:

“(…) appropriately designed (….) studies are required, in which many examiners render decisions about many independent tests (typically, involving “questioned” samples and one or more “known” samples) and the error rates are determined.” [p. 143]

While we unreservedly agree with the view that such empirical information is valuable for characterising *general* scientific validity of a given forensic branch, and of individual examiners, we also emphasize that there is room for a moment of reflection and think about the following question: is the above understanding of decision and its empirical investigation really the end of the matter? As exemplified below, the forensic community appears to be rather unclear about what ought to be done with data of empirical validation studies.

Suppose we had data to determine performance characteristics¹⁹ of, for example, a fingerprint examiner, to an extent that we would consider sufficient for that expert to be allowed to give testimony. Can we expect then that people will know what *exactly* to do with such data? It appears that the answer is negative, as even some academic writers offer only vague replies. For example, [14] rightly ask “(…) what could the juror conclude on the basis of the experiment (…)?” [p. 10], but then only note that “[t]he juror can reason with this information to infer something about the particular case – to deduce the particular from the general” [p. 10]. Strangely, however, these authors leave the term ‘something’ essentially undefined, while there is standard theory – particularly well established in medical literature – on how to use diagnostic performance measures (e.g., sensitivity and specificity) for inference about hypotheses of interest. An example is given by [18], and [36] provide a discussion in the context of DNA profiles and on how to take into account the potential of error.

Yet other authors use empirical data to inform judgement about whether forensic scientists can be regarded as experts [38]. As mentioned above, this is relevant for general considerations but of quite limited help for decisional questions in the individual case.

Turning now to the question of decision, we can ask again: do diagnostic performance measures tell recipients of expert information anything about how to make a ‘good’ decision in a particular case at hand? And similarly, a forensic expert may ask: how can empirically derived performance measures help me make a sound decision (e.g., an identification decision) in the particular case I face? The truth is that general diagnostic performance measures are *no* direct guides to decisions in the individual case. For example, even though the proportion of false positive answers by an expert – or a group of experts – in a series of experiments under controlled conditions was low, there is no way of asserting then that an identification decision in a particular case at hand is ‘probably’ correct (i.e., the proposition that the potential source of the trace is in fact the source being true), or was the optimal choice to be made. Claiming the contrary would mean to confuse decision empiricism with the logic of decision (see Section 2.1). The reason for this is that

- standard diagnostic performance measures are defined in terms of the proportion of correct determinations in a series of experiments under controlled conditions (i.e., comparisons with respect to ground truths), whereas

¹⁹See, for example, [34] on data regarding the proportion of correct and erroneous conclusions of fingerprint examiners in a series of experiments conducted under controlled conditions.

- the qualification and quantification of the ‘goodness’ of a decision in an individual case – as understood in this paper (Section 2.3 and Appendix A) – is measured on the basis of the decision-maker’s *case-specific* preferences among decision consequences and the assessment of the probability of obtaining the various consequences.

We have argued throughout this paper that the former, decision empiricism, is an important and necessary, but not sufficient preliminary of the latter – decision logic. The two are connected, though, through inference and statistics, which take on the task of quantifying the probative strength of an expert’s report (e.g., a reported correspondence between a trace and a reference print) with respect to competing propositions of interest (e.g., the person of interest is the source of the trace versus some unknown person is the source), and taking into account relevant expert performance descriptors (e.g., from controlled empirical studies). This seeks to assist decision-makers – which may be scientists themselves – in revising their beliefs about the main propositions in view of the expert’s report. But then, answering the question of how to decide, in the case at hand, is screened off from the previously considered empirical elements: instead, decision is a consideration of both one’s current belief state, given all the information available at the time when the decision needs to be made, and one’s goals, expressed in terms of preferences among the possible decision consequences.

The decision-theoretic framework that we have sketched here (Section 2.3 and Section A) makes this view formally precise, but remains entirely general. For example, the theory only emphasizes (i) that measuring uncertainties by probabilities, and undesirability of adverse consequences by means of losses is crucial, and (ii) how these elements ought to be combined to ensure coherence. But the theory does neither say whose probabilities and losses (utilities) are meant, nor how they ought to be assigned. It is clear, thus, that individualisation in this account is not confined to only scientific or statistical considerations, but requires a broader perspective, calling for an active role of other participants in the legal process [11].

Appendix A: Decision theoretic criterion

Denote by $\Pr(\theta_1 | I)$ and $\Pr(\theta_2 | I)$ the decision-maker’s probability for two states of nature θ_1 and θ_2 , called here propositions, or hypotheses, given the information I . Note that the development can also be given with respect to more than two states of nature without much additional effort. As an example in a forensic science context, θ_1 and θ_2 could refer to the propositions that a person of interest (POI) is, or is not, respectively, the source of a mark or trace (e.g., a fingerprint, blood stain) found on a crime scene, or the person whose handwriting is present on a questioned document, and so on.

Denote by d_1 and d_2 two decisions, one of which must be taken in the light of uncertainty about whether θ_1 or θ_2 is true. There may also be more than two decisions, as is the case for the states of nature θ .²⁰ In a situation involving forensic individualisation, d_1 and d_2 may refer to, for example, the decisions of considering or not the POI as the source of a particular trace. Let C_{ij} be the consequence of deciding d_i when the state of nature θ_j holds, and define the undesirability of a decision consequence C_{ij} in terms of the loss $L(C_{ij})$. See Table 2 for a summary.

Combining probabilities for states of nature and losses for decision consequences allows one to obtain the *expected* value, associated with a particular course of action (i.e., decision). Specifically, when qualifying decision consequences in terms of losses, this leads to the notion of expected loss EL of decisions d_i , defined as follows:

$$\text{EL}(d_i) = \sum_j L(C_{ij}) \Pr(\theta_j | I).$$

The principle of minimizing expected loss then says to choose the decision that has the minimum expected loss. The decision criterion for deciding d_1 (individualisation) over d_2 (not individualise), can thus be written as:

$$\text{if } \underbrace{L(C_{11}) \Pr(\theta_1 | I) + L(C_{12}) \Pr(\theta_2 | I)}_{\text{EL}(d_1)} < \underbrace{L(C_{21}) \Pr(\theta_1 | I) + L(C_{22}) \Pr(\theta_2 | I)}_{\text{EL}(d_2)}, \text{ decide } d_1. \quad (1)$$

²⁰See [3] for a development with an additional decision called ‘inconclusive’, alongside with decisions ‘individualisation’ and ‘exclusion’.

Note that the quantification of undesirability of real-world decision outcomes are not necessarily monetary. What matters is that the decision-maker can provide a ranking of the decision consequences according to their desirability (or, undesirability). For the case at hand, one may agree about the following ranking

$$(C_{11}, C_{22}), C_{21}, C_{12},$$

where equally preferred consequences are in brackets, and consequences with decreasing desirability are written from the left to the right. The ranking above thus means that deciding wrongly in favour of the POI (i.e., a missed individualisation, C_{21}) is considered a less adverse consequence than wrongly deciding against the POI (i.e., a false individualisation, C_{12}). The best consequences are correct conclusions, that is the individualising the POI if that person is truly the source of the trace (C_{11}), and not individualising the POI if an unknown person is the source of the trace (C_{22}).

Suppose a $(0, k)$ loss function is chosen, assigning a 0 loss to the most favorable consequence(s), and a loss equal to k to the least favorable consequence(s). According to this scale, the pair (C_{11}, C_{22}) , that is correct individualisations and non-individualisations, represent the most desirable consequences and their associated losses are set equal to 0. Then, the expected loss of deciding ‘individualisation’ reduces to $EL(d_1) = L(C_{12}) \Pr(\theta_2 | I)$, because the term $L(C_{11})$ is zero. Analogously, the expected loss of not individualising the POI reduces to $EL(d_2) = L(C_{21}) \Pr(\theta_1 | I)$, and the decision theoretic criterion in (1) becomes

$$L(C_{12}) \Pr(\theta_2 | I) < L(C_{21}) \Pr(\theta_1 | I).$$

By rearranging the terms, one can obtain (see also, for example, [2]):

$$\frac{\Pr(\theta_1 | I)}{\Pr(\theta_2 | I)} > \frac{L(C_{12})}{L(C_{21})}. \quad (2)$$

In words, the decision-theoretic criterion amounts to a comparison between the odds in favor of θ_1 against θ_2 , and the relative losses associated with the two possible ways of deciding wrongly, that is the ratio of the loss of a false individualisation to the loss of a missed individualisation. This is valid for any loss function of this type. Note that the above criterion is also known as *Bayesian* decision-theoretic criterion in situations where the probabilities for the states of nature are conditioned by, for example, additional information (i.e., evidence) E , written $\Pr(\theta | I, E)$, obtained through Bayesian updating of $\Pr(\theta | I)$. Equation (2) is also discussed in legal literature [e.g., 9]. A more elaborate version of the formula for general loss/utility functions is discussed in [16] and [27]. Presentations in statistical literature can be found, for example, in [1], [2], [29] and [30].

Suppose further that a $(0, 1)$ scale is chosen, that is $k = 1$. The advantage of this is that the assignment of numerical values to intermediate consequences can be made through a coherent comparison with a standard that is given by two reference consequences [23]. The expected loss of decision d_1 then reduces further to $EL(d_1) = \Pr(\theta_2 | I)$, since a false individualisation, C_{12} , is the worst consequence, with the maximum loss assigned, that is $L(C_{12}) = 1$. The (Bayesian) decision criterion then further simplifies to:

$$\text{if } \underbrace{\Pr(\theta_2 | I)}_{EL(d_1)} < \underbrace{L(C_{21}) \Pr(\theta_1 | I)}_{EL(d_2)}, \text{ decide } d_1,$$

or, by rearranging terms

$$\frac{\Pr(\theta_2 | I)}{\Pr(\theta_1 | I)} < L(C_{21}).$$

Under the assumption of a $(0, 1)$ loss function, the decision criterion thus amounts to a comparison between the odds against θ_1 and the loss that is associated to the intermediate consequence C_{21} .

Appendix B: Example for the minimax method for individual decisions

The use of the minimax method seems to be appropriate in decision situations where the decision maker has to take a decision under a ‘veil of ignorance’ because they have not access to the relevant information, and they cannot have

Table 4: Loss matrix with d_1 and d_2 denoting, respectively, the decisions to report and not to report a candidate profile found after a search in a DNA database. The states of nature θ_1 and θ_2 denote, respectively, the person of interest is the source of the trace and an unknown person is the source of the trace. Maximum losses for each decision are highlighted with a circle.

Decisions		States of nature: The trace comes from	
		... the person of interest (θ_1)	... an unknown person (θ_2)
Loss matrix:	report candidate profile (d_1)	0	$\textcircled{< 1}$
	do not report profile (d_2)	$\textcircled{1}$	0

access, by factual or normative reasons. Imagine a investigative situation in which an incomplete DNA profile (i.e., a DNA profile with no detected alleles for some loci) is searched against a database of DNA profiles. A candidate profile is found. It shows the same alleles as those observed in the profile of the trace, but also additional alleles at some loci for which the trace apparently is homozygous (i.e., shows only one corresponding allele). There is established theory for evaluating the strength of such a finding for inference of source [e.g., 5], but this is not the question of interest here. Rather, the question of interest is which of the two decisions d_1 , report the candidate profile to the laboratory (and hence to the recipient of expert information), and d_2 , do not report the candidate profile, should be selected. There are two states of nature, θ_1 , the person of interest is the source of the questioned DNA trace, and θ_2 , an unknown person is the source of the trace. Table 4 gives a possible assignment of losses to decision consequences (i.e., a $0 - k$ loss function with $k = 1$): reporting a DNA profile when the person of interest is the source of the trace, and not reporting a DNA profile when an unknown person is the source of the trace, are correct determinations and hence are assigned zero loss. Incorrect determinations, namely reporting a profile when the person of interest is not the source (i.e., a ‘false alarm’, consequence C_{12}) and failing to report a profile when the person of interest is the source (consequence C_{21}), have loss values different from zero. Specifically, suppose the following assignments:

- $L(C_{21}) = 1$: Failing to report²¹ a profile when in fact the person of interest is the source of the trace (consequence C_{21}) is considered the worst consequence because it amounts to deprive investigative authorities from a relevant candidate.
- $L(C_{12}) < 1$: This assignment reflects the investigative priority according to which falsely reporting a profile of a person who is not the source of the trace (consequence C_{12}) is less adverse than missing a person who truly is the source of the trace (consequence C_{21}), hence $L(C_{12}) < L(C_{21})$. Stated otherwise, the inconvenience (or, loss associated with) pursuing a false lead is considered smaller than the inconvenience resulting from missing a relevant lead.²²

Note that the above assignments are made for the sole purpose of this example and readers are invited to choose their own assignments.

Return now to our scientist who conducted the database search and who is in charge of deciding whether the candidate profile, found as a result of the database search, is to be reported (d_1) or not (d_2). The reporting scientist has no knowledge at all of the case to which the trace relates and the circumstances of that case. Beliefs about whether or not the person of interest (θ_1) or an unknown person (θ_2) is the source of the trace are above and beyond the area of competence of the reporting scientist (i.e., probabilities for states of nature cannot be reasonably conceived), yet the scientist needs to make a decision under a condition that we would call of ‘ignorance’, rather than of ‘uncertainty’, to stress the point that is not the task of the reporting scientist to assess the probabilities of the case (see fn. 21). If the losses are as specified above, and the decision-maker’s goal is to minimize the maximum potential loss when making a decision, then the minimax method would advise to always select d_1 (i.e., report the candidate profile).

It is interesting to note that the above conclusion, always deciding d_1 (reporting the candidate profile), does *not* necessarily result when deciding based on the principle of minimum expected loss (Appendix A). As a brief illustration

²¹Note that reporting in this example only means to inform relevant authorities about the observed correspondence between the DNA profile of the trace and the profile of the person of interest. Such a report does *not* individualise the person of interest or otherwise convey a conclusion regarding inference of source.

²²This assignment of losses is different from the one considered in Appendix A (and Table 3).

of this, suppose that for $L(C_{12})$, the loss for reporting a profile when the person of interest is not the source, a value close to 1 is chosen. This would mean that the loss ratio on the righthand side of Equation 2 would be approximately 1, too. With such a loss structure, reporting the candidate profile (decision d_1) would only be optimal – in the sense of the criterion of minimising expected loss – if and only if the odds in favour of the person of interest being the source of the trace, $\Pr(\theta_1 | I) / \Pr(\theta_2 | I)$, are rather high, that is at least 1, approximately. This, however, may be hardly the case, or justifiable, if the person of interest is found as a result of a database search and there is no or little other prior information pointing towards this person.

Acknowledgments

This research was supported by the Swiss National Science Foundation through grant No. BSSGI0.155809 and the University of Lausanne. Some of the writing of this paper was carried out during a visiting research stay of Alex Biedermann at New York University School of Law.

References

- [1] J O Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1985.
- [2] J M Bernardo and A F M Smith. *Bayesian Theory*. John Wiley & Sons, Chichester, second edition, 2000.
- [3] A Biedermann, S Bozza, and F Taroni. Decision theoretic properties of forensic identification: underlying logic and argumentative implications. *Forensic Science International*, 177:120–132, 2008.
- [4] A Biedermann, S Bozza, and F Taroni. The decisionalization of individualization. *Forensic Science International*, 266:29–38, 2016.
- [5] J S Buckleton, J-A Bright, and D Taylor. *Forensic DNA Evidence Interpretation*. CRC Press, Boca Raton, FL, second edition, 2016.
- [6] C Champod. Identification/individualisation, overview and meaning of ID. In J H Siegel, P J Saukko, and G C Knupfer, editors, *Encyclopedia of Forensic Science*, pages 1077–1084. Academic Press, San Diego, 2000.
- [7] C Champod. Fingerprint identification: Advances since the 2009 NAS report. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370:1–10, 2015.
- [8] C Champod, C Lennard, P Margot, and M Stoilovic. *Fingerprints and Other Ridge Skin Impressions*. CRC Press, Boca Raton, second edition, 2016.
- [9] E K Cheng. Reconceptualizing the burden of proof. *The Yale Law Journal*, 122:1254–1279, 2013.
- [10] H Chernoff and L E Moses. *Elementary Decision Theory*. John Wiley & Sons, New York, 1959.
- [11] S A Cole. Individualization is dead, long live individualization! Reforms of reporting practices for fingerprint analysis in the United States. *Law, Probability and Risk*, 13:117–150, 2014.
- [12] I Dror. Cognitive neuroscience in forensic science: understanding and utilizing the human element. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370:1–8, 2015.
- [13] I Dror. Human expert performance in forensic decision making: Seven different sources of bias. *Australian Journal of Forensic Sciences*, in press:1–7, 2017.
- [14] G Edmond, M B Thompson, and J M Tangen. A guide to interpreting forensic testimony: scientific approaches to fingerprint evidence. *Law, Probability and Risk*, 13:1–25, 2014.
- [15] Expert Working Group on Human Factors in Latent Print Analysis. Latent print examination and human factors: Improving the practice through a systems approach. Technical report, National Institute of Standards and Technology, U.S. Department of Commerce, 2012.
- [16] R D Friedman. *The Elements of Evidence*. West Academic Publishing, Saint Paul, Minnesota, fourth edition, 2017.
- [17] J Kaplan. Decision theory and the factfinding process. *Stanford Law Review*, 20:1065–1092, 1968.
- [18] D H Kaye. The validity of tests: Caveant omnes. *Jurimetrics Journal*, 27:349–361, 1987.
- [19] C R Kingston. Application of probability theory in criminalistics. *Journal of the American Statistical Association*, 60:70–80, 1965.
- [20] P Kintz. 2014 consensus for the use of alcohol markers in hair for assessment of both abstinence and chronic excessive alcohol consumption. *Forensic Science International*, 249:A1–A2, 2015.
- [21] P L Kirk. The ontogeny of criminalistics. *Journal of Criminal Law, Criminology and Police Science*, 54:235–238, 1963.
- [22] E S Lander. Response to the ANZFSS council statement on the President’s Council Of Advisors On Science And Technology Report. *Australian Journal of Forensic Sciences*, 49:366–368, 2017.
- [23] D V Lindley. *Making Decisions*. John Wiley & Sons, Chichester, second edition, 1985.
- [24] D V Lindley. Subjective probability, decision analysis and their legal consequences. *Journal of the Royal Statistical Society, Series A*, 54: 83–92, 1991.
- [25] D V Lindley. The philosophy of statistics. *The Statistician*, 49:293–337, 2000.
- [26] D V Lindley. *Understanding Uncertainty*. John Wiley & Sons, Hoboken, revised edition, 2014.
- [27] D A Nance. *The Burdens of Proof. Discriminatory Power, Weight of Evidence, and Tenacity of Belief*. Cambridge University Press, Cambridge, 2016.
- [28] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, D.C., 2009.
- [29] G Parmigiani. *Modeling in Medical Decision Making, A Bayesian approach*. John Wiley & Sons, Chichester, 2002.

- [30] G Parmigiani and L Inoue. *Decision Theory: Principles and Approaches*. John Wiley & Sons, Chichester, 2009.
- [31] PCAST. President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Washington, D.C., 2016.
- [32] President's Council of Advisors on Science and Technology (PCAST). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Washington, D.C., 2016.
- [33] D A Stoney. What made us ever think we could individualize using statistics? *Journal of the Forensic Science Society*, 31:197–199, 1991.
- [34] J M Tangen, M B Thompson, and D J McCarthy. Identifying fingerprint expertise. *Psychological Science*, 22:995–997, 2011.
- [35] 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.
- [36] 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.
- [37] W C Thompson, J Black, A Jain, and J Kadane. *Latent Fingerprint Examination*. Forensic Science Assessments: A Quality and Gap Analysis. American Association for the Advancement of Science, Washington, D.C., 2017.
- [38] A Towler, D White, K Ballantyne, R A Searston, K A Martire, and R I Kemp. Are forensic scientists experts? *Journal of Applied Research in Memory and Cognition*, 7:199–208, 2018.
- [39] BT Ulery, A Hicklin, JA Buscaglia, and MA Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy of Science of the United States of America*, 108:7733–7738, 2011.