Authors' post-print:

Taroni F., Bozza S., Biedermann A. 2020, Decision theory, in: Handbook of Forensic Statistics, Banks D. L., Kafadar K., Kaye D.H., Tackett M. (Eds.), Chapman & Hall/CRC, Handbooks of Modern Statistical Methods, 103–130.

5

Decision theory

CONTENTS

Franco Taroni
School of Criminal Justice, University of Lausanne, Switzerland (franco.taroni@unil.ch)

Silvia Bozza
Department of Economics, Ca'Foscari University Venice, Venice, Italy (silvia.bozza@unive.it)
School of Criminal Justice, University of Lausanne, Switzerland (silvia.bozza@unil.ch)

Alex Biedermann
School of Criminal Justice, University of Lausanne, Switzerland (alex.biedermann@unil.ch)

5.1 Introduction

Forensic scientists, lawyers and other participants of the legal process are routinely faced with problems of making decisions under circumstances of uncertainty. Uncertainty relates to propositions of interest that are not completely known by the decision-maker at the time when a decision needs to be made. Propositions may relate to the source or nature of forensic traces, marks and objects. For example, with friction ridge marks, propositions of interest may be 'Does this fingermark come from the person of interest (POI) or from some unknown person?'. In forensic document examination, a scientist may ask 'Is this a genuine document or has it been modified (e.g., page substitution)?'. In forensic anthropology the question 'Are these human remains?' may arise, and so on. Replying in one way or another to such questions may be perceived as uncomfortable since knowledge about the relevant underlying truth-state of the world is incomplete to some extent. For example, in typical real-world applications of forensic science it is not known with certainty, when *deciding* to consider a POI as the source of a particular fingermark, whether the POI is in fact the source of the fingermark. Similarly, at an advanced stage of the legal process, the question of whether to convict or acquit a POI (i.e. the verdict) needs to be made in the presence of incomplete knowledge about whether or not the POI truly is the offender. There are analogies between the above questions, in terms of their logical underpinnings, that can be studied, analysed and described using formal methods, such as decision theory, which will be the main aim of this chapter.

Around the middle of the past century, discussions intensified and several fields of study emerged on decision-making concerning, for example, contexts where decisions have monetary consequences. These developments gravitated around questions such as how decisions should be made in order to be considered rational (Pratt et al., 1964). Though an important area, economics was not the only branch with strong interests in decision-making and decision analysis. Entire fields of study developed and interacted with each other in various ways, including psychology, mathematics and statistics, the law and philosophy of science, among others.

This chapter will primarily rely on statistical decision theory[1] as developed by Leonard Savage (1954) and in subsequent treatises (e.g., Lindley, 1985; Luce and Raiffa, 1958; Raiffa, 1968) as the framework for studying the formal structure of decision problems arising in forensic science and the law. Before proceeding with this presentation, an important preliminary needs to be considered. It deals with the question of how to understand decision analysis and the notion of theory of decision. To this point, the field of judgment and decision making, a branch of applied psychology, has contributed considerably by crystallizing three main perspectives and approaches, known as the descriptive, the normative and the prescriptive view (Baron, 2008; French et al., 2009). For a review of the history of these terms, see Baron (2006). Broadly speaking, the descriptive approach focuses on peoples' observable decision behaviour and extends to the development of psychological theories intended to explain how individuals make decisions. Such research is valuable in that it allows one to better understand the conditions under which decision behaviour departs towards incoherence or, worse, logical error. However, revealing such departures requires reference points against which observable decision behaviour can be compared. The provision of such reference points, also called normative standards, is the object of study of the normative approach. Decision theory and decision criteria (or, norms) derived from it, fall into this

[1] In later parts of this chapter, the discussion will be extended to the notion of *Bayesian* statistical decision theory, emphasizing the idea of using Bayesian inference procedures to inform decision makers, for example based on experimental information (Parmigiani, 2001).

category of study. It is mainly pursued by mathematicians, statisticians and philosophers of science. The third perspective, the prescriptive approach, addresses the question of what recommendations ought to be derived from normative insights in order to improve practical decision making. For example, some strict normativists, such as Lindley (1985), consider that the normative concept of probability – that is, a standard for reasoning under uncertainty – and decision theory as its extension, are also prescriptive in the sense that they provide direct prescriptions on how to arrange one's reasoning and acting. Properly distinguishing the different intentions and goals of these kinds of decision science research is important for an informed discourse about notions of decision and decision analysis in forensic science applications (Biedermann et al., 2014).

This chapter is structured as follows. Section 5.2 outlines standard elements of statistical decision theory that will be exemplified in Section 5.3 for decision problems arising in the law in general (Section 5.3.1) and forensic science in particular (Section 5.3.2). This exposition will include examples such as decisions following forensic inference of source (i.e., identification/individualization; Section 5.3.2.1). Discussion and conclusions will be presented in Section 5.4. Further readings on applications of decision theory in forensic science and treatments of decision theory in general are given in Section 5.5.

5.2 Concepts of statistical decision theory

5.2.1 Preliminaries: basic elements of decision problems

Decision theory is a mathematical theory of how to make decisions when there is uncertainty about the true state of nature. The presence of uncertainty implies that a choice among the alternative courses of action leads to uncertainty regarding which consequences will effectively take place. In statistical terms, the states of nature may also be referred to as parameters, commonly denoted by θ, which may be discrete or continuous. The collection of all possible states of nature is denoted by Θ, the parameter space, and represents a first element of the formalization of decision problems. A second basic element is the feasible decisions (or courses of action), denoted by d. The space of all decisions, called the decision space, is denoted by $\mathcal{D}$. The third basic element is the consequences c. They are defined as the outcome following the combination of a decision d taken when the actual state of nature is θ, formally written $c(d, \theta)$. The space of all consequences is denoted by $\mathcal{C}$. Before proceeding, in Section 5.2.2, with presenting a formal approach to qualifying and quantifying the relative merit of rival courses of action, given the basic elements of the decision problem, it is useful to devote a few more comments to the description of the decision space and the parameter space.

Regarding the decision space, it is important for the decision maker to draw up an exhaustive list of m decisions that are available, say $d_1, d_2, ..., d_m \in \mathcal{D}$. As noted by Lindley: "(...) it would not be a properly defined decision problem in which the only decision was whether to go to the cinema, because if the decision were not made (that is, one did not go to the cinema) one would have to decide whether to stay at home and read, or go to the public-house, or indulge in other activities. All the possible decisions, or actions, must be included (...)" (Lindley, 1965, p. 63). Further, it is convenient to make the requirement of exclusivity, meaning that only one of the decisions can be selected. As noted by Lindley: "Hence, the decisions are both exclusive and exhaustive: one of them *has* to be taken, and at most one of them *can* be taken." (Lindley, 1985, p. 6).

The second task for a decision maker is to draw up a list of n exclusive and exhaustive events or states of nature, say $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$. Regarding the latter list, the decision maker may distinguish between situations of certainty and uncertainty. In the former case, certainty, the decision maker has complete knowledge about the states of nature. Hence, each alternative course of action leads to one and only one foreseeable consequence, and a choice among alternatives is equivalent to a choice among related consequences. In the latter case, uncertainty, the decision maker does not know which state of nature actually holds, or what the future will be. Consequently, each available course of action will have one of several consequences. It is possible, however, to measure uncertainty about the states of nature using a suitable probability distribution Pr over Θ. Note that in some fields, such as business decision analysis and operations research , this situation is called 'decision making under risk' and the expression 'decision making under uncertainty' is reserved to situations in which the decision maker is unable to provide a list of all possible outcomes and/or a probability distribution for the various outcomes. In this Chapter, however, this interpretation will not be pursued.

5.2.2 Utility theory

The principal issue in decision making under uncertainty is the selection of a member in the list of available decisions without knowing which state of nature is truly the case. The aim, therefore, is to create a framework that allows decision makers to assess the consequences of alternative courses of action in order to compare them and avoid irrational choices or behavior.

The formulation of such a decision framework involves, first, the assumption that the decision maker can express preferences amongst possible consequences. It is in fact assumed that the space of consequences has a partial pre-ordering, denoted by $\preceq$, meaning that the decision maker must be able to specify, at any point, which consequence is suitable or whether they are equivalent (Piccinato, 1996). When comparing any pair of consequences $(c_1, c_2) \in \mathcal{C}$, $c_1 \prec c_2$ indicates that the consequence c_2 is strictly preferred to consequence c_1, $c_1 \sim c_2$ indicates that c_1 and c_2 are equivalent (or equally preferred), while $c_1 \preceq c_2$ indicates that c_1 is not preferred to c_2, that is either $c_1 \prec c_2$, or $c_1 \sim c_2$ holds. The measurement of preferences among decision outcomes is operated by a function, called a utility function, denoted by $\mathrm{U}(\cdot)$ that associates a utility value $\mathrm{U}(d, \theta)$ to each one of the possible consequences $c(d, \theta)$, also denoted $\mathrm{U}(c)$; it specifies the desirability of each consequence on some numerical scale.

Second, the decision maker's uncertainty about the states of nature, when they are discrete, is expressed in terms of a probability mass function $\Pr(\theta \mid I)$, where I denotes the relevant information available at the time when the probability assessment is made. Combining the utilities $\mathrm{U}(d, \theta)$ for decision consequences and the probabilities for states of nature leads to a measure of the desirability of alternative courses of action d in terms of their expected utility (EU)[2]:

$$\mathrm{EU}(d) = \sum_{\theta \in \Theta} \mathrm{U}(d, \theta) \Pr(\theta \mid I).$$

A standard decision rule, based on EU, instructs one to select the action with the maximum

[2]The same idea can be applied when θ is continuos and takes values in Θ_c, $\theta \in \Theta_c$. The probability mass function $\Pr(\theta \mid I)$ is replaced by a probability density function $f(\theta \mid I)$ and the expected utility of decision d is:

$$\mathrm{EU}(d) = \int_{\Theta_c} \mathrm{U}(d, \theta) f(\theta \mid I) d\theta.$$

expected utility (see also Section 5.2.3). Hereafter, information I will be omitted to simplify the notation, though it is important to keep in mind that it conditions all probability assignments.

Some further conditions (axioms) must be imposed on the preference system in order for there to exist a function U, the utility function, such that for any pair $(c_1, c_2) \in \mathcal{C}$, the relationship $c_1 \preceq c_2$ holds if and only if $\mathrm{U}(c_1) \leq \mathrm{U}(c_2)$.

A.1 The first axiom requires that the preference system is *complete*. This amounts to assume that for any pair of consequences (c_1, c_2) of the space of consequences $\mathcal{C}$, it must always be possible to express a preference or indifference among them (one of the following relations must hold: $c_1 \prec c_2, c_2 \prec c_1, c_1 \sim c_2$).

A.2 The second axiom requires that the preference system is *transitive*. This means that for any $(c_1, c_2, c_3) \in \mathcal{C}$, if one prefers c_2 to c_1 ($c_1 \prec c_2$) and c_3 to c_2 ($c_2 \prec c_3$), then one prefers c_3 to c_1 ($c_1 \prec c_3$). In the same way, if one is indifferent between c_1 and c_2 ($c_1 \sim c_2$), and is indifferent between c_2 and c_3 ($c_2 \sim c_3$), then one is indifferent between c_1 and c_3 ($c_1 \sim c_3$). Not all the consequences are equivalent to each other, that is, for at least a pair of consequences (c_1, c_2), either $c_1 \prec c_2$ or $c_2 \prec c_1$ holds.

A.3 The third axiom requires that the ordering of preferences is invariant with respect to compound gambles. For any pair of consequences $(c_1, c_2) \in \mathcal{C}$, such that $c_1 \preceq c_2$, then, for any other consequence $c_3 \in \mathcal{C}$, and any probability α, the gamble that offers probability α of winning c_2, and probability $(1-\alpha)$ of winning c_3 is preferred (or it is equivalent) to the gamble that offers probability α of winning c_1 and probability $(1-\alpha)$ of winning c_3. Denote by $(c_i, c_j; \alpha, 1-\alpha)$ the gamble offering c_i with probability α, and c_j with probability $(1-\alpha)$, $i \neq j$. This axiom can then be formulated as follows: $c_1 \preceq c_2$ if and only if $(c_1, c_3; \alpha, 1-\alpha) \preceq (c_2, c_3; \alpha, 1-\alpha)$, for any $\alpha \in [0, 1]$ and any $c_3 \in \mathcal{C}$.

A.4 The forth axiom requires that there are not (i) infinitely desirable or (ii) infinitely undesirable consequences. Let $(c_1, c_2, c_3) \in \mathcal{C}$ be any three consequences such that c_1 is preferred to c_2 and c_2 is preferred to c_3 ($c_3 \prec c_2 \prec c_1$). Then there exist probabilities α and β, such that (i) c_2 is preferred to the gamble $(c_1, c_3; \alpha, 1-\alpha)$; (ii) the gamble $(c_1, c_3; \beta, 1-\beta)$ is preferred to c_2.

If (i) does not hold, then one will always prefer the possibility of obtaining the best consequence c_1, no matter how small is the probability of obtaining it, to c_2; that is, one believes that c_1 is infinitely better than c_2 (and c_3). If (ii) does not hold, then one will prefer c_2, no matter how small the probability of obtaining the worse consequence c_3 is; that is, one believes that c_3 is infinitely worse than c_2 (and c_1).

If these four conditions are satisfied, then one can prove the *expected utility theorem*, according to which there exists a function U on the space of consequences $\mathcal{C}$ such that for any d_i and d_k belonging to the decision space $\mathcal{D}$, d_i is preferred (or equivalent) to d_k if and only if the expected utility of d_i, $\mathrm{EU}(d_i)$, is greater (or equal) than the expected utility of d_k, $\mathrm{EU}(d_k)$, that is, assuming θ discrete, if

$$\sum_{\theta \in \Theta} \mathrm{U}(d_i, \theta) \Pr(\theta) \geq \sum_{\theta \in \Theta} \mathrm{U}(d_k, \theta) \Pr(\theta).$$

Consider, next, any consequence c and a pair $(c_1, c_2) \in \mathcal{C}$ such that $c_1 \prec c_2$, and $c_1 \preceq c \preceq c_2$. Following the stated conditions, it may be proved (see De Groot (1970)) that there exists a unique number $\alpha \in [0, 1]$ such that

$$c \sim [\alpha c_1 + (1-\alpha) c_2], \tag{5.1}$$

and that

$$U(c) = \alpha U(c_1) + (1-\alpha)U(c_2). \tag{5.2}$$

It can also be proved that the utility function is invariant under linear transformations. This means that if $U(c)$ is a utility function, then for any $a > 0$, $aU(c) + b$ is also a utility function preserving the same pattern of preferences.

Utility functions can be constructed in different ways. One possibility starts with a pair of non-equivalent consequences $(c_1, c_2) \in \mathcal{C}$ and assigns them a utility value. This will fix the origin and the scale of the utility function. The desirability of each consequence $c \in \mathcal{C}$ of interest will then be compared with those of c_1 and c_2. Given that utility functions are invariant under linear transformation, the choice of c_1 and c_2, and the choice of the scale of the utility, are not relevant. They are, however, generally identified with the worst and the best consequence, respectively. It is assumed, for example, that the utility of the worst consequence is zero, $U(c_1) = 0$, and the utility of the best consequence is one, $U(c_2) = 1$. The utilities of the remaining intermediate consequences are computed using Equation (5.2). This will be discussed further in Section 5.3.

5.2.3 Implications of the expected utility maximisation principle

Consider taking a decision d when the true state of nature is θ, so that the consequence is $c(d,\theta)$. It is possible to show, using relation (5.1), that there exists some α such that the consequence $c(d,\theta)$ is equivalent to a hypothetical gamble offering the worst consequence c_1 with probability α and the best consequence c_2 with probability $(1-\alpha)$

$$c(d,\theta) \sim [\alpha c_1 + (1-\alpha)c_2], \qquad c_1 \preceq c(d,\theta) \preceq c_2.$$

The utility $U(d,\theta)$ of the consequence $c(d,\theta)$ can then be calculated using Equation (5.2) as follows:

$$U(d,\theta) = \alpha \underbrace{U(c_1)}_{0} + (1-\alpha)\underbrace{U(c_2)}_{1} = 1-\alpha.$$

According to this, for any d and any θ, selecting decision d is equivalent to assigning a probability $U(d,\theta) = 1-\alpha$ to the occurrence of the most favorable consequence. This hypothetical gamble can always be played. It can be played, in particular, after that decision d has been taken and it is known which state of nature θ holds. The term $U(d,\theta)$ can be understood as the conditional probability of obtaining the consequence c_2, given decision d has been taken and the state of nature θ occurred: $\Pr(c_2 \mid d,\theta) = U(d,\theta)$. Note that probability $\Pr(c_2 \mid d)$ can be written in extended form as

$$\Pr(c_2 \mid d) = \sum_{\theta\in\Theta} \Pr(c_2 \mid d,\theta)\Pr(\theta). \tag{5.3}$$

Therefore, (5.3) can be rewritten as

$$\Pr(c_2 \mid d) = \sum_{\theta\in\Theta} U(d,\theta)\Pr(\theta), \tag{5.4}$$

namely, the expected utility that quantifies the probability of obtaining the best consequence once decision d is taken (Lindley, 1985). The decision rule which instructs decision makers to select the decision which maximizes the expected utility (MEU criterion) is optimal because it is the decision which has associated with it the highest probability of obtaining the most favorable consequence.

States of nature:	θ_1	θ_2
Decisions: d_1	C_{11}	C_{12}
d_2	C_{21}	C_{22}

TABLE 5.1
A simple decision matrix with two decisions d_1 and d_2, two states of nature θ_1 and θ_2 and corresponding decision consequences C_{ij} (for $i, j = \{1, 2\}$).

5.2.4 The loss function

An alternative way to express preferences among decision consequences $c(d, \theta)$ is the use of non-negative loss functions. When a utility function is available, the loss function can be derived as follows (Lindley, 1985):

$$\mathrm{L}(d, \theta) = \max_{d \in \mathcal{D}} \mathrm{U}(d, \theta) - \mathrm{U}(d, \theta) \tag{5.5}$$

The loss $\mathrm{L}(d, \theta)$ for a given consequence $c(d, \theta)$ thus is defined as the difference between the utility of the best consequence under the state of nature at hand and the utility for the consequence of interest. That is, the loss measures the penalty for choosing a non-optimal action, also called opportunity loss (Press, 1989, p. 26-27): the difference between the utility of the best consequence that could have been obtained and the utility of the actual one received.

Note that following Equation (5.5), losses cannot, by definition, be negative because $\mathrm{U}(d, \theta)$ will be smaller or at best equal to $\max_{d \in \mathcal{D}} \mathrm{U}(d, \theta)$. The expected loss, $\mathrm{EL}(d)$, thus characterises the undesirability of each possible decision, and can be quantified as follows:

$$\mathrm{EL}(d) = \sum_{\theta \in \Theta} \mathrm{L}(d, \theta) \Pr(\theta).$$

When using losses instead of utilities, the decision rule of maximising expected utility becomes the rule instructing the selection of the decision that minimizes the expected loss $\mathrm{EL}(d)$. It might be objected that assuming a non-negative loss function is too restrictive. Note, however, that the loss function represents error due to an non-optimal choice. It thus makes sense to consider that even the most favorable decision will induce at best a zero loss.

5.2.5 Particular forms of the expected utility maximisation principle

For the remainder of this Chapter, it will be important to anticipate two particular forms in which the MEU principle may be formulated. Consider, first, the utility-based perspective of a two-action decision problem involving two states of nature, θ_1 and θ_2. The decision maker's probabilities for these states of nature are $\Pr(\theta_1 \mid \cdot)$ and $\Pr(\theta_2 \mid \cdot)$, respectively, such that $\Pr(\theta_1 \mid \cdot) + \Pr(\theta_2 \mid \cdot) = 1$. Note that $\mid \cdot$ is shorthand notation for the conditioning on any relevant evidence E or background information I. The two possible decisions are d_1 and d_2, representing the decision maker's acceptance of, respectively, θ_1 and θ_2 as the true states of nature. Hereafter, write C_{ij} to denote the consequence $c(d_i, \theta_j)$ of taking decision d_i when θ_j is the actual state of nature and denote the corresponding utility by $\mathrm{U}(C_{ij})$. The decision problem is summarized in Table 5.1.

According to the principle of maximization of expected utility, the decision maker should select decision d_1 rather than d_2 if $\mathrm{EU}(d_1) > \mathrm{EU}(d_2)$. This will be the case if

$$\mathrm{U}(C_{11}) \Pr(\theta_1 \mid \cdot) + \mathrm{U}(C_{12}) \Pr(\theta_2 \mid \cdot) > \mathrm{U}(C_{21}) \Pr(\theta_1 \mid \cdot) + \mathrm{U}(C_{22}) \Pr(\theta_2 \mid \cdot), \tag{5.6}$$

which can be rearranged to give

$$\frac{\Pr(\theta_1 \mid \cdot)}{\Pr(\theta_2 \mid \cdot)} > \frac{\mathrm{U}(C_{22}) - \mathrm{U}(C_{12})}{\mathrm{U}(C_{11}) - \mathrm{U}(C_{21})}. \tag{5.7}$$

The term $\mathrm{U}(C_{22}) - \mathrm{U}(C_{12})$ in the numerator on the right-hand side of (5.7) is the additional utility involved in making the correct decision when θ_2 turns out to be the correct state of nature. An alternative way to look at this term is to consider it as the potential regret: it is the potential loss in utility when erroneously deciding d_1 instead of d_2. The term $\mathrm{U}(C_{11}) - \mathrm{U}(C_{21})$ similarly deals with the potential regret of deciding d_2 when the true state of nature is θ_1. Relation (5.7) thus states that decision d_1 should only be taken if the odds in favour of θ_1 are sufficient to outweigh any extra potential regret associated with incorrectly deciding d_1 (Spiegelhalter et al., 2004).

Consider now the loss-based account. Recall, from Section 5.2.4, that the loss $\mathrm{L}(d_i, \theta_j) = \mathrm{L}(C_{ij})$ for a decision consequence C_{ij} is the difference between the utility of the outcome of the best decision under the state of nature at hand, and the utility of the outcome of the actual decision d_i under the same state of nature. Therefore, the decision that minimizes the expected loss is the same as the decision that maximizes the expected utility. Continuing the example introduced above, assume that there is a positive loss L incurred when falsely choosing a proposition that is not actually the case, that is $\mathrm{L}(C_{ij}) > 0$ if $i \neq j$, and there is no loss when accepting a proposition that is actually the case, that is $\mathrm{L}(C_{ij}) = 0$ if $i = j$. The loss can be symmetric, $\mathrm{L}(C_{ij}) = \mathrm{L}(C_{ji})$, or asymmetric, $\mathrm{L}(C_{ij}) \neq \mathrm{L}(C_{ji})$, $i \neq j$. The decision criterion depicted in (5.6) will become to select d_1 rather then d_2 if $\mathrm{EL}(d_1) < \mathrm{EL}(d_2)$, that is if

$$\mathrm{L}(C_{11}) \Pr(\theta_1 \mid \cdot) + \mathrm{L}(C_{12}) \Pr(\theta_2 \mid \cdot) < \mathrm{L}(C_{21}) \Pr(\theta_1 \mid \cdot) + \mathrm{L}(C_{22}) \Pr(\theta_2 \mid \cdot), \tag{5.8}$$

and the expected loss of deciding d_i will be:

$$\mathrm{EL}(d_i \mid \cdot) = \underbrace{\mathrm{L}(C_{ii})}_{0} \Pr(\theta_i \mid \cdot) + \mathrm{L}(C_{ij}) \Pr(\theta_j \mid \cdot) = \mathrm{L}(C_{ij}) \Pr(\theta_j \mid \cdot), \qquad i \neq j.$$

Considering the principle of minimizing expected loss and given that

$$\mathrm{EL}(d_1 \mid \cdot) < \mathrm{EL}(d_2 \mid \cdot) \text{ if and only if } \mathrm{L}(C_{12}) \Pr(\theta_2 \mid \cdot) < \mathrm{L}(C_{21}) \Pr(\theta_1 \mid \cdot),$$

the decision problem involves a comparison of odds with the ratio of losses associated with erroneous decisions. Specifically, deciding d_1 rather than d_2 is optimal if and only if:

$$\frac{\Pr(\theta_1 \mid \cdot)}{\Pr(\theta_2 \mid \cdot)} > \frac{\mathrm{L}(C_{12})}{\mathrm{L}(C_{21})} \tag{5.9}$$

or, equivalently

$$\frac{\Pr(\theta_2 \mid \cdot)}{\Pr(\theta_1 \mid \cdot)} < \frac{\mathrm{L}(C_{21})}{\mathrm{L}(C_{12})}. \tag{5.10}$$

The loss ratio on the right-hand side in (5.9) and (5.10) fixes a threshold for odds. The relation (5.9) specifies that if the odds in favor of θ_1 exceed the loss incurred from incorrectly choosing decision d_1 divided by the loss incurred from incorrectly choosing decision d_2, then the decision maker should take decision d_1.

5.2.6 Likelihood ratios in the decision framework

So far it has been considered that the decision maker's probabilities for the state of nature are conditional probabilities written $\Pr(\theta_1 \mid \cdot)$ and $\Pr(\theta_2 \mid \cdot)$, incorporating all relevant evidence E and background information I available at the time when the decision needs to be made. The odds in (5.9) can therefore be interpreted as *posterior* odds. It is useful to emphasize that likelihood ratios, commonly used in forensic science for quantifying the value of forensic results (e.g., Aitken and Taroni, 2004), play an important role in the inference process preceding the decision. Recalling that the posterior odds can be written as the product of the prior odds and the likelihood ratio for the forensic results E, the relation (5.9) can thus be rewritten as:

$$\frac{\Pr(\theta_1 \mid I, E)}{\Pr(\theta_2 \mid I, E)} = \underbrace{\frac{\Pr(\theta_1 \mid I)}{\Pr(\theta_2 \mid I)}}_{\text{prior odds}} \times \underbrace{\frac{\Pr(E \mid \theta_1, I)}{\Pr(E \mid \theta_2, I)}}_{\text{likelihood ratio}} > \underbrace{\frac{L(C_{12})}{L(C_{21})}}_{\text{loss ratio}} . \tag{5.11}$$

Relation (5.11) defines the conditions under which the decision d_1 is preferable to d_2, that is when the relative losses on the right are smaller than the product on the left, containing the likelihood ratio. Thus, it is now possible to reformulate the decision criterion, minimizing expected loss (Section 5.2.5), with an emphasis on the likelihood ratio, as follows:

> *The decision d_1 is to be preferred to decision d_2 if the product of the likelihood ratio and the prior odds is larger than the ratio of the losses associated with adverse decision consequences.*

A more intuitive form of (5.11) can be obtained when working with logarithms (e.g., Good, 1950):

$$\log\left[\frac{\Pr(\theta_1 \mid I)}{\Pr(\theta_2 \mid I)}\right] + \log\left[\frac{\Pr(E \mid \theta_1, I)}{\Pr(E \mid \theta_2, I)}\right] > \log\left[\frac{L(C_{12})}{L(C_{21})}\right] . \tag{5.12}$$

By re-arranging the terms one can isolate the log-likelihood ratio as follows:

$$\log\left[\frac{\Pr(E \mid \theta_1, I)}{\Pr(E \mid \theta_2, I)}\right] > \log\left[\frac{L(C_{12})}{L(C_{21})}\right] - \log\left[\frac{\Pr(\theta_1 \mid I)}{\Pr(\theta_2 \mid I)}\right] . \tag{5.13}$$

Note that following Good (1950), the logarithm of the likelihood ratio, the term on the left, is commonly referred to as the weight of evidence. The decision criterion minimizing expected loss (Section 5.2.5) thus becomes:

> *The decision d_1 is to be preferred to decision d_2 if and only if the weight of evidence is* greater *than the difference between the logarithm of the ratio of the losses associated with adverse consequences and the logarithm of the prior odds in favor of proposition θ_1.*

5.3 Decision theory in the law and forensic science

5.3.1 Legal applications

Decision theory offers a formal framework for thinking analytically about decision problems, but this perspective – in particular the use of probability – is controversial both among some legal scholars and practitioners. Although formal decision theoretic discourses can be dated back about half a century ago (Kaplan, 1968), there has been a concentration of several article collections since the mid-1980s. See for example, Tillers and Green (1988), especially the paper by Kaye (1988), and the collection of articles in the *Internatinal Journal of Evidence & Proof* (Vol. 1, 1997) entitled 'Bayesianism and Juridical Proof', edited by R. J. Allen and M. Redmayne.

A generic outline of applying the decision-theoretic elements presented in Section 5.2 to legal decision problems proceeds among the following lines. Assume that in a case of interest there are only two possible decisions: decision d_1, finding for the plaintiff, and decision d_2, finding for the defense. Considering this decision problem in terms of expected utilities requires the specification of probabilities of the states of nature and utilities for decision consequences. Let θ_1 and θ_2 denote versions of the case wherein the plaintiff or defendant, respectively, is entitled to judgment. Let $\Pr(\theta_j)$ be the decision maker's probability, at a given time, for a given state of nature θ_j. Deciding in favour of, respectively, the plaintiff and the defendant may lead to accurate consequences, namely C_{11} and C_{22}, or adverse outcomes, namely C_{12} and C_{21}. Then, comparing the relative merit of decisions d_1 and d_2 comes down to criterion (5.6), that is deciding in favour of the plaintiff rather than for the defendant would require the expected utility of decision d_1 to be greater than the expected utility of decision d_2. The immediate question following this observation is 'When is this the case?'. As noted in Section 5.2.5, it can be helpful to illustrate the logic of the decision-theoretic result by formulating the MEU principle in an alternative form, such as relation (5.7), separating the thinking about probabilities from thinking about the utilities of the various decision consequences. For a discussion of relation (5.7) see, for example, Friedman (1997, 2017).

The decision-theoretic criterion may be more insightful if it is considered through a loss-based perspective (Section 5.2.5) using, for example, relation (5.9). Let $\mathrm{L}(C_{12})$ denote the loss associated with wrongly deciding in favour of the plaintiff, and $\mathrm{L}(C_{21})$ denote the loss associated with wrongly deciding in favour of the defendant. It is then clear to see that with a symmetric $0-k$ loss function, that is with $\mathrm{L}(C_{12}) = \mathrm{L}(C_{21}) = k$ for adverse decision outcomes, and zero loss $\mathrm{L}(C_{11}) = \mathrm{L}_(C_{22}) = 0$ for accurate decision outcomes, deciding in favour of, for example, the plaintiff is warranted if and only if the probability $\Pr(\theta_1 \mid \cdot)$ is greater than 0.5. This result is sometimes associated with the notions of 'balance of probabilities' and 'more probable than not' standards, translating common ideas in civil litigation according to which a correct judgment for the plaintiff is as preferable as a correct judgment for the defendant, and that erroneous verdicts for either side are equally undesirable (e.g., Kaye, 1999).

The expression (5.9) can also capture the logical structure of the decision problem of the typical criminal case where the prosecution has the burden of proving its case with respect to a particular standard. In such situations, $\mathrm{L}(C_{12})$ is the loss of falsely declaring the defendant guilty, whereas $\mathrm{L}(C_{21})$ is the loss associated with a false acquittal. A common viewpoint is the preference ordering $C_{12} \prec C_{21}$ according to which a false conviction is more undesirable than a false acquittal. Consequently, this amounts to consider the loss $\mathrm{L}(C_{12})$ of a false conviction to be greater, often considerably greater, than the loss $\mathrm{L}(C_{21})$ associated with a false acquittal. For any loss ratio thus defined, depending on the nature of the case and the stakes involved, the criterion (5.9) provides the minimal odds of liability required in order for a conviction, decision d_1, to be preferred to an acquittal, decision d_2. In this context, reference is sometimes made to Blackstone's 10 to 1 criterion according to which it is better that 10 truly liable defendants go free than 1 innocent defendant being wrongly convicted. It has been noted, however, that this criterion does not easily map to assignments of losses in a given singular case, but rather seems to refer to actual error rates across multiple distinct trials (Kaye, 1999).

It should be kept in mind that criterion (5.9) is an analytical result that can be thought through in two different ways: either starting from a preference structure (i.e., a loss ratio for adverse decision consequences) and derive the lower limit of the odds necessary to warrant a conviction, or starting with a given value for the odds of liability and then work out the loss ratio corresponding to these odds so that a given decision is warranted. Over the past decades, this normative account has stimulated considerable empirical research on, for example, what various subjects (e.g., judges, citizens, etc.) consider as required levels of probability before deciding in one way or another. See Dane (1985) and Simon and Mahan (1971), for example, and Hastie (1993) for a review. The quantitative values observed in such studies, using various elicitation procedures and methodologies, vary over broad ranges and depend largely on experimental conditions. Note, however, that this mismatch between, on the one hand, consistency requirements for utilities and probabilities implied by the theory, and, on the other hand, peoples' intuitive feelings about these assignments and their interdependency, does not invalidate the formal mathematical results. Similarly, arithmetics is not abandoned simply because practically operating individuals might reply, for example, with an answer other than 4 to the question of how much is $2 + 2$ (see, e.g., Lindley in de Finetti (1974)). This is an instance of the difference between the normative and the descriptive perspectives to decision mentioned in Section 5.1.

5.3.2 Forensic science applications

5.3.2.1 Forensic identification

Preliminaries

One of the most well known (Champod, 2000) but also most widely challenged (Cole, 2014) notions in forensic science is 'individualization', or 'identification'. It is a conclusion following inference of source and involves the claim to reduce a pool of potential donors of a forensic trace to a single source. This section presents a decision-theoretic account of forensic individualization based on analyses previously given in Biedermann et al. (2008, 2016). For presentations in the wider context of Bayesian data analysis and Bayesian decision networks, see Taroni et al. (2010, 2014). More generally, examiners' conclusions in forensic identification practice now are, increasingly often, referred to as decisions (Cole and Biedermann, 2020). See, for example, the reports issued by the PCAST (2016) and the AAAS (Thompson et al., 2017). There is an interest, thus, to devote attention to the ways in which decision may be understood and conceptualised from a scientific point of view. Decision theory provides a mathematically rigorous account for this.

Suppose extraneous material (e.g., blood) or a mark (e.g., finger- or toolmark) is collected at a crime scene and an individual is apprehended or a tool – called a potential source – is found. Similarly, one may imagine a litigation case in which a contested signature is present on a questioned document (e.g., a contract) and the question is whether or not the signature is from the POI. For the purpose of the current discussion, assume two uncertain events defined as 'The crime mark comes from the suspect' (θ_1), and 'The crime mark comes from an unknown person' (θ_2). These two states of nature are discrete and form the parameter space Θ. Assume further that 'identifying' (sometimes also called 'individualizing') an individual as being the source of a crime mark can be considered as a decision (d_1) made by a person authorized to do so. For the remainder of the analysis, it is not necessary to specify whether this authorized person is a (forensic) scientist or some other participant in the legal process. As alternative decisions, consider the conclusions 'inconclusive' (d_2) and 'exclusion' (d_3). These forms of conclusion are currently used by many forensic practitioners. Combining these elements leads to the decision matrix shown in Table 5.2. The outcome of an 'identification' ('exclusion') conclusion, decision d_1 (d_3), is

Decisions	States of nature θ_1: POI is donor	θ_2: An unknown person is donor
d_1: identification	C_{11}: correct identification	C_{12}: false identification
d_2: inconclusive	C_{21}: neutral	C_{22}: neutral
d_3: exclusion	C_{31}: false exclusion	C_{32}: correct exclusion

TABLE 5.2
Decision matrix for a forensic identification problem with d_i, $i = 1, 2, 3$, denoting decisions, θ_j, $j = 1, 2$, denoting states of nature and C_{ij} denoting the consequence of taking decision d_i when θ_j turns out to be the true state of nature (Biedermann et al., 2008).

an accurate outcome if the POI is truly (is truly not) the origin of the crime mark. These consequences are referred to as 'correct identification' and 'correct exclusion', respectively. The outcome of an 'identification' ('exclusion') conclusion, decision d_1 (d_3), can be adverse if the POI is truly not (is truly) the origin of the crime mark. These consequences are listed as 'false identification' and 'false exclusion', respectively. Because the statement 'inconclusive', decision d_2, does not convey any information that tends to associate or otherwise the POI with the issue of the source of the crime mark, the consequences following d_2 are referred to as 'neutral'.

Preference ordering and construction of the utility function

Consider the following ordering of consequences:

$$C_{12} \prec C_{31} \prec C_{21} \sim C_{22} \prec C_{32} \sim C_{11}. \tag{5.14}$$

This preference ordering states that the most preferred consequences are a correct identification (C_{11}) and a correct exclusion (C_{32}), and the worst consequence is a false identification (C_{12}). To construct the utility function, after having chosen the scale, one starts by assigning the maximum utility value to the best consequence, in this case the couple C_{32} and C_{11}, and the minimum utility value to the worst consequence, C_{12}. Therefore, if a $(0, 1)$ scale is chosen, $\mathrm{U}(C_{11}) = \mathrm{U}(C_{32}) = 1$ and $\mathrm{U}(C_{12}) = 0$.

The next steps consist in assigning an utility value to the intermediate consequences. Consider the consequence called 'neutral', C_{21}, and the above preference ranking

$$C_{12} \prec C_{21} \prec C_{11}.$$

In particular, it has been observed – see (5.1) and (5.2) – that if the preference system respects given conditions, there exists, for the decision maker, a unique number $0 \le \alpha \le 1$ such that the consequence C_{21} is equivalent to a hypothetical gamble where the worst consequence, C_{12}, is obtained with probability α, and the best consequence, C_{11} is obtained with probability $(1 - \alpha)$:

$$C_{21} \sim [\alpha C_{12} + (1 - \alpha) C_{11}], \tag{5.15}$$

and the utility of C_{21} can be computed as

$$\mathrm{U}(C_{21}) = \alpha \underbrace{\mathrm{U}(C_{12})}_{0} + (1 - \alpha) \underbrace{\mathrm{U}(C_{11})}_{1} = 1 - \alpha.$$

Note that the utility of consequence C_{21} turns out to be the probability $(1 - \alpha)$ of finishing with the best consequence in the space of all possible consequences. In particular, note the equivalence of utility and probability in the latter sentence. Finding such an α

is the most difficult part of the utility elicitation procedure. It involves answering the question what would make the decision maker indifferent between a neutral consequence, and a situation in which a false identification might occur. Specifically, the decision maker must specify the value α so that the sure consequence C_{21} appears equivalent to the gamble in which the worst consequence is obtained with probability α and the best consequence is obtained with probability $1 - \alpha$.

When thinking about the above question, one must be careful *not* to use as values for α the assigned probabilities $\Pr(\theta_1)$ and $\Pr(\theta_2)$ for the propositions of interest. In fact, the number α is a limiting value, asking decision makers to crystallize what would be, in general, their highest probability of running the risk of making the worse mistake they are willing to exchange with the consequence of rendering an 'inconclusive' statement. Suppose, for instance, as an extreme position, that the answer is zero, meaning that the decision maker never wants to run such a risk. This would mean, however, that no matter how high the probability of a correct identification is, the decision maker would consider that a neutral conclusion is as good as a correct identification. It is also worth noting that since $C_{32} \sim C_{11}$, one can substitute C_{32} for C_{11} in (5.15), so obtaining another hypothetical gamble:

$$C_{21} \sim [\alpha C_{12} + (1 - \alpha)C_{32}].$$

To ensure coherence, the decision maker should again consider a neutral conclusion as much worth as a correct exclusion, no matter how high the probability of a correct exclusion is. Thus, if the decision maker would indeed consider that the highest probability for incurring the worst consequence, in exchange with the consequence of providing an 'inconclusive' statement, is strictly zero, then this belief cannot be coherent with the preference ranking (5.14). It can be coherent only with the following ordering:

$$C_{12} \prec C_{31} \prec C_{21} \sim C_{22} \sim C_{32} \sim C_{11}. \tag{5.16}$$

Note that there is a change in the fourth preference sign from the left.

Consider the logical implications of the preference ordering (5.16) through the decision matrix in Table 5.2. The preference ranking (5.16) implies that, if the proposition θ_1 is true, then decisions d_1 ('identification') and d_2 ('inconclusive') are equally preferred and both better than decision d_3 ('exclusion'). In turn, if the proposition θ_2 is true, then the decisions d_2 and d_3 are equally preferred and both better than decision d_1. Thus, decision d_2 is the best decision overall, because, if the proposition θ_1 is true, it is better than d_3, and, if the proposition θ_2 is true, it is better than d_1. Therefore, if the decision maker considers a strictly zero probability for the event of incurring the worst consequence, in exchange to the consequence of rendering an 'inconclusive' statement, then the decision maker should *always* take the decision 'inconclusive' (d_2).

This does not correspond, however, to the way in which decision makers behave in practice. Therefore, there *must exist* for these decision makers a unique value $0 < \alpha < 1$ such that the hypothetical gamble (5.15) does make sense for them, despite the inherent challenge to find the limiting value α. For the purpose of illustration, assume that the decision maker considers that $\alpha = 0.001$ is appropriate. Then

$$\mathrm{U}(C_{21}) = \alpha \mathrm{U}(C_{12}) + (1 - \alpha)\mathrm{U}(C_{11}) = 1 - \alpha = 0.999.$$

Likewise, the utility of the consequence C_{31} can be elicited and quantified in comparison with $\mathrm{U}(C_{12})$, the utility of the worst consequence, and $\mathrm{U}(C_{11})$, the utility of the best consequence,

$$\mathrm{U}(C_{31}) = \alpha^* \mathrm{U}(C_{12}) + (1 - \alpha^*)\mathrm{U}(C_{11}) = 1 - \alpha^*.$$

Here, $1 - \alpha^*$ represents the decision maker's highest probability for incurring the worst consequence in exchange with the consequence of rendering a 'false exclusion' statement. For behaviour to be coherent, this limiting value α^* must necessarily be higher then the previous limiting value $\alpha = 0.001$. The reason for this is that the decision maker is facing, on the right-hand-side, a gamble with the same consequences as before and, on the left-hand-side, a less preferred consequence: recall the ranking $C_{31} \prec C_{21}$. Assume, for example, that $\alpha^* = 0.01$ is felt to be correct. Then $\mathrm{U}(C_{31}) = 0.99$. This value means that the decision maker is indifferent between a false exclusion (C_{31}) and a gamble in which the worst consequence, a false identification (C_{12}), is obtained with probability $\alpha^* = 0.01$, and a correct identification with probability $1 - \alpha^* = 0.99$. Note that the order relation in the space of consequences is preserved. However, this is not the end of the matter. It is a good idea at this stage to check the appropriateness of the so-built utility function because there is no guarantee that the quantified utility values are coherent (Berger, 1988). This question can be examined by comparing different combinations of consequences as:

$$C_{31} \prec C_{21} \prec C_{11} \qquad \text{or} \qquad C_{12} \prec C_{31} \prec C_{21}$$

Consider the case on the left, for instance. There must exist, at this stage, a unique value α' such that

$$C_{21} \sim [\alpha' C_{31} + (1 - \alpha')C_{11}]. \tag{5.17}$$

According to the illustrated gambling scheme, and the quantified utilities,

$$\begin{aligned} \mathrm{U}(C_{21}) &= \alpha'\mathrm{U}(C_{31}) + (1 - \alpha')\mathrm{U}(C_{11}) \\ 0.999 &= \alpha' 0.99 + (1 - \alpha'). \end{aligned}$$

When solving this equation one obtains $\alpha' = 0.1$. Now, if one believes that this value is correct, in the sense that one is indifferent between a neutral consequence and a gamble where a false exclusion may occur with probability 0.1, then the utility function is coherent. Otherwise, one needs to go back and check previous assessments. For further discussion on such comparisons, see also Biedermann et al. (2008).

Computing expected utilities

Consider Table 5.3 for a summary of the utility values derived above. Assume that there is scientific evidence, denoted by E, available and used to inform the probabilities of the competing propositions θ_1 and θ_2, leading to posterior probabilities $\Pr(\theta_1 \mid E)$ and $\Pr(\theta_2 \mid E)$ at the time when the decision is made[3]. Start by considering the computation of the expected utility of the decision d_1 ('identification'),

$$\mathrm{EU}(d_1) = \mathrm{U}(C_{11}) \Pr(\theta_1 \mid E) + \mathrm{U}(C_{12}) \Pr(\theta_2 \mid E).$$

Given the assigned utility values, it is immediately seen that the expected utility of decision d_1 reduces to: $\mathrm{EU}(d_1) = \Pr(\theta_1 \mid E)$. So, assuming a $(0, 1)$ utility function, and recalling result (5.4) where the expected utility of a decision d is equated with the probability of obtaining the best consequence once decision d is taken, it follows that

$$\mathrm{EU}(d_1) = \Pr(\text{correct identification} \mid d_1) = \Pr(\theta_1 \mid E).$$

In the same way, one can compute the expected utilities of the decisions 'inconclusive' (d_2) and 'exclusion' (d_3):

$$\begin{aligned} \mathrm{EU}(d_2) &= \mathrm{U}(C_{21}) \Pr(\theta_1 \mid E) + \mathrm{U}(C_{22})[1 - \Pr(\theta_1 \mid E)] \\ &= \mathrm{U}(C_{21}) = \mathrm{U}(C_{22}). \end{aligned}$$

[3]Note again that information I is omitted to simplify the notation, though it is important to keep in mind that it conditions all probability assignments.

Decisions	Uncertain events θ_1	θ_2
d_1: identification	1	0
d_2: inconclusive	0.999	0.999
d_3: exclusion	0.99	1

TABLE 5.3
Illustrative values for utilities $\mathrm{U}(C_{ij}) = \mathrm{U}(d_i, \theta_j)$, as discussed in the text, for a case of forensic identification. The propositions of interest are θ_1 'The crime stain comes from the POI' and θ_2 'The crime stain comes from an unknown person' (Taroni et al., 2010, 2014).

$$\begin{aligned} \mathrm{EU}(d_3) &= \mathrm{U}(C_{31})\Pr(\theta_1 \mid E) + \mathrm{U}(C_{32})[1 - \Pr(\theta_1 \mid E)] \\ &= \mathrm{U}(C_{31})\Pr(\theta_1 \mid E) + [1 - \Pr(\theta_1 \mid E)]. \end{aligned}$$

The decision with the highest expected utility, that is the optimal decision, depends on the relative magnitude of $\Pr(\theta_1 \mid E)$, $\mathrm{U}(C_{21})$ and $\mathrm{U}(C_{31})$.

Recall, from the above analysis, that the probability α in (5.1) is actually a limiting value. Intuitively, this implies that, if the probability of θ_2 is higher than this limiting value, then 'identification' cannot be the best decision. To examine this aspect, consider the utility values given in Table 5.3 and assume that the probability of θ_2 is 0.0011 (i.e., slightly greater that α):

$$\Pr(\theta_2 \mid E) = 0.0011 > \alpha = 0.001.$$

The following expected utility values can then be calculated:

$$\begin{aligned} \mathrm{EU}(d_1) &= (1 \times 0.9989) = 0.9989 \\ \mathrm{EU}(d_2) &= (0.999 \times 0.9989) + (0.999 \times 0.0011) = 0.999 \\ \mathrm{EU}(d_3) &= (0.99 \times 0.9989) + (1 \times 0.0011) = 0.990011. \end{aligned}$$

Thus, $\mathrm{EU}(d_2) > \mathrm{EU}(d_1) > \mathrm{EU}(d_3)$ and the decision 'inconclusive' (d_2) is better than the decision 'identification' (d_1). 'Inconclusive' is the decision with the highest expected utility.

Comments

It is readily seen that the decision with the highest expected utility depends on the interplay between probabilities and utilities, though merely looking at formulae in isolation might not be helpful to get a sense of the interaction among the relevant factors. It may thus be useful to graphically display the expected utilities of the various decisions as a function of, for example, the probability of θ_1, the proposition according to which the crime mark comes from the POI. Following the above computations, it is clear that $\mathrm{EU}(d_1)$ is an increasing linear function, corresponding to the probability of θ_1. In turn, $\mathrm{EU}(d_2)$ is a constant, leading to a horizontal line at $y = \mathrm{U}(C_{21}) = \mathrm{U}(C_{22})$. Finally, $\mathrm{EU}(d_3)$ is also a linear function, but decreasing. This is illustrated in Figure 5.1 (left). Note that, for illustrative purposes and improving readability, the computations plotted in Figure 5.1 are based on slightly modified values for α and α^*, that is 0.05 instead of 0.001 and 0.2 instead of 0.01, respectively. The bold solid lines in Figure 5.1 (right) highlight the decision with the maximum expected utility. Intersections between expected utility functions represent transition points, indicating a change in the decision with maximum expected utility when further increasing or decreasing the probability of θ_1. So, it can be easily observed that, by choosing $\alpha = 0.05$ and $\alpha^* = 0.2$,[4] for $\Pr(\theta_1 \mid E)$ smaller than 0.25 the optimal decision is

[4]This amounts to have $\mathrm{U}(C_{21}) = \mathrm{U}(C_{22}) = 0.95$ and $\mathrm{U}(C_{31}) = 0.8$.

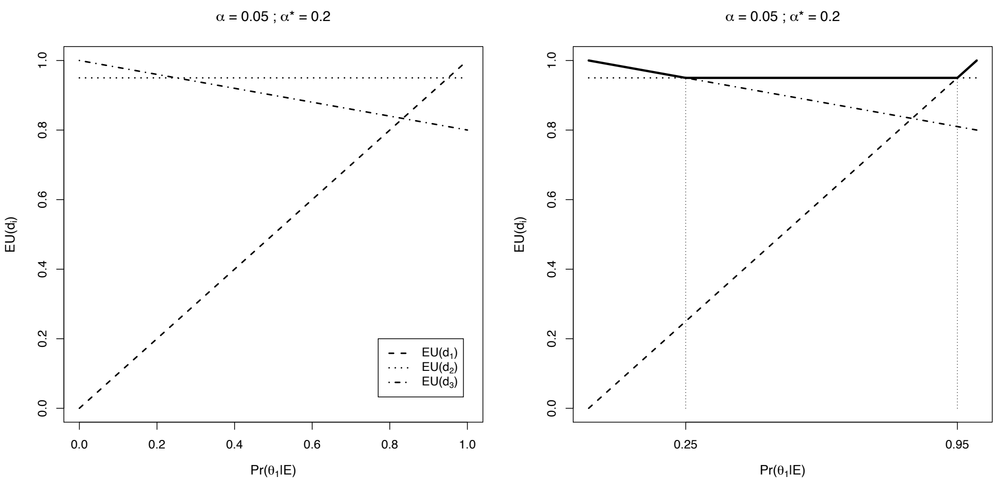

FIGURE 5.1
Illustrative expected utilities for a case of forensic identification with the propositions θ_1 'The crime stain comes from the POI' and θ_2 'The crime stain comes from an unknown person' (left). The available courses of action are 'identification' (d_1), 'inconclusive' (d_2) and 'exclusion' (d_3). The expected utilities are computed as a function of the probability of θ_1 (x-axis), using the utility values obtained by choosing $\alpha = 0.05$ and $\alpha^* = 0.2$. The bold solid lines (right) highlight, for each possible value of the probability of θ_1, the decision with the maximum expected utility.

d_3 (exclusion), and for $0.25 \leq \Pr(\theta_1 \mid E) \leq 0.95$ the optimal decision is d_2 (inconclusive). Finally, decision d_1 (identification) is optimal only whenever $\Pr(\theta_1 \mid E)$ is larger than 0.95.

It may be objected that these values may not be reasonable, and in fact decision makers will build their utility function coherently with their preference systems. Expected utility functions of decisions d_1, d_2 and d_3 obtained with the original choices $\alpha = 0.001$ and $\alpha^* = 0.01$ are plotted in Figure 5.2, where the $x-$ and $y-$ axes focus on the transition points (i.e., on the left, the transition point from d_3 to d_2 is highlighted, fixed at $\Pr(\theta_1 \mid E) = 0.1$; on the right the transition point from d_2 to d_1 is shown, fixed at $\Pr(\theta_1 \mid E) = 0.999$). In particular, it is pointed out that according to this preference system, the probability of θ_1 must be at least equal to 0.999 to have d_1 (identification) as the optimal decision, or equivalently, the probability of θ_2 must be smaller than 0.001, the limiting value.

Note also that the decision-theoretic analysis of forensic identification can also be conducted with only two decisions, such as d_1 ('identification') and d_2 ('do not identify') and/or the quantification of decision consequences in terms of losses instead of utilities (Biedermann et al., 2016). The decision criterion of minimizing expected loss then comes down to criterion (5.9) with inferential properties as discussed previously in Section 5.2.5 and 5.3.1. For a discussion on the application of the decision-theoretic account to forensic identification in the particular area of fingermarks, see Champod et al. (2016). Identification decisions following database searches are discussed in Gittelson et al. (2012). For a formulation of the expected loss minimization decision criterion, and an explicit representation of the role of the likelihood ratio in the inference process preceding the decision, the development given in Section 5.2.6 applies, in particular (5.12) and (5.13). Consider, for example, the following

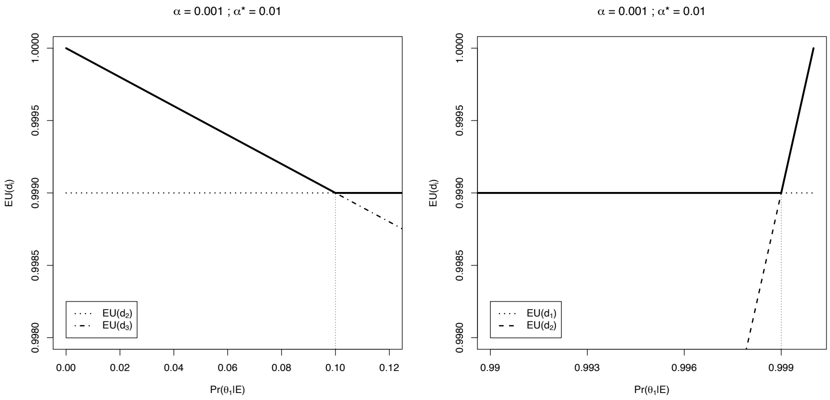

FIGURE 5.2
Illustrative expected utilities for in a case of forensic identification with the propositions θ_1 'The crime stain comes from the POI' and θ_2 'The crime stain comes from an unknown person'. The available courses of action are 'identification' (d_1), 'inconclusive' (d_2) and 'exclusion' (d_3). The expected utilities are computed as a function of the probability of θ_1 ($x-$axis), using the utility values given in Table 5.3. The bold solid lines highlight, for each possible value of the probability of θ_1, the decision with the maximum expected utility.

abbreviated form of the criterion (5.12):

$$\log(\text{PO}) + \log(\text{LR}) > \log(\text{RL}),$$

that is the requirement that the sum of the logarithm of the prior odds (PO) and the logarithm of the likelihood ratio (LR) must exceed the logarithm of the ratio of the losses of adverse decision consequences (RL). Table 5.4 illustrates numerical examples of combinations of values, in particular limiting values that the likelihood ratio must exceed, in order to make decision d_1 (individualization) preferable to decision d_2 (not individualizing the POI).

5.3.2.2 Understanding probability assignment as a decision: the use of proper scoring rules

So far in this Section, probability has been encountered as a concept to express uncertainty about the truth or otherwise of propositions or events that are not entirely known to the decision maker at the time when a decision needs to be made. Interpreting probability as a decision maker's personal degree of belief, it will be unreasonable then to say that decision makers do not know their own states of mind. Throughout this chapter, it is tacitly assumed that the decision makers have, each of them in their own way, and at any point in time, their own probability, depending on their extent of knowledge and background information. However, a legitimate question that may arise is what it means for a given person to assign a so-called personal probability. In this context proper scoring rules allow one to clarify the probability elicitation process, pointing out that it can itself be understood as question of decision making.

To illustrate the notion of proper scoring rule, consider a person who is asked to state the probability that he or she assigns to a given event E. Assume further that part of the question is the information that the declared probability will be scored with respect

$PO = \Pr(\theta_1 \mid I)/\Pr(\theta_2 \mid I)$	LR	RL	log(PO)	log(LR)	log(RL)
$1/10 = 0.1$	100	10	-1	2	1
$1/10 = 0.1$	1000	100	-1	3	2
$1/1000 = 0.001$	10^5	100	-3	5	2
$1/1000 = 0.001$	10^6	1000	-3	6	3

TABLE 5.4
Numerical examples, presented in Biedermann et al. (2016), of minimum likelihood ratio (LR) values for satisfying the expected loss minimization decision criterion necessary to make the decision d_1 (individualising the POI) preferable to d_2 (not individualizing the POI), criteria (5.12) and (5.13), with PO denoting prior odds (odds in favour of the proposition that the POI of interest is the source of the crime stain) and RL denoting the relative losses (i.e., the ratio of the losses of adverse decision consequences). The values in columns four to six are the logarithms (base 10) of the values presented in the first three columns.

to the actual truth or falsity of E, denoted by, respectively, 1 and 0. The purpose of this added constraint is to motivate people to report their actual beliefs, rather than a deliberately chosen value (i.e., a value that is different from the one they have in mind), in a way that is made precise shortly below. Distorted probability assertions are sometimes encountered in the context of forensic identification, for example when experts round off small probabilities to zero, or high probabilities to 1. Another typical example is the notion of relevance of trace material which expresses the relationship between a given trace or stain and the offender. Typically, one cannot categorically assert relevance, as observed by Stoney, because relevance "(...) may range from very likely to practically nil (...)" (Stoney, 1994, at p. 18). Formally, relevance refers to a proposition of the kind 'the stain or mark comes from (or, was left by) the offender' and appears as a factor in various likelihood ratio developments (e.g., Evett et al., 1998). Uncertainty about propositions regarding evidential relevance is sometimes suppressed and evaluators declare a probability of relevance $p' = 1$, thus rounding up to 1 a probability actually smaller than one. Scoring rules allow one to show that this distortion of probabilities is not advisable.

The notion of score, in the decision-theoretic context, refers to the square of the difference between the probability p' for the event E, as stated by the person, and the actual truth-value of E, zero or one. Because of this squared difference the rule is also called 'quadratic scoring rule'. As an example, assume that the scientist, or any other person being asked to state their probability for event E, declares the value $p' = 0.8$. Thus, in the analysis here, statement $p' = 0.8$ is interpreted as a decision and the question is how to decide in an optimal way (i.e., what value p' to report). One can then distinguish two cases. In one case, E is true and thus has the truth value 1. This situation leads to the score $(1-0.8)^2 = 0.2^2$. In the other case, the event E is not true and thus assumes the truth value 0, leading to the score $(0-0.8)^2 = (-0.8)^2$. However, these score calculations, leading to expressions of *actual* penalty, are only hypothetical. Given that the scientist is uncertain about whether or not E is true, the scientist cannot know the actual score. At best the scientist can consider a prevision of the scoring and then seeking a way of proceeding that minimizes the expected penalty. This leads to the notions of 'prevision of the scoring', 'expected penalty' and 'expected loss'. These notions are based on the idea of combining the possible scores and the probabilities p and $(1-p)$ with which the scores may be produced. Recall that the general concept of expected value was introduced in Section 5.2.2.

In general, the expected loss for reporting probability p' given the scientists' actual belief

p is

$$\mathrm{EL}(p'_i) = \mathrm{L}(p'_i, E = true)\,\Pr(E = true) + \mathrm{L}(p'_i, E = false)\,\Pr(E = false),$$

or $(1-p')^2p + p'^2(1-p)$ for short. One can readily see that the expected loss is minimal when the reported belief p' corresponds to the scientist's actual belief p. In the example considered here, with $p' = p = 0.8$, the expected loss is:

$$\begin{aligned}(1-p')^2p + p'^2(1-p) &= (1-0.8)^2 \times 0.8 + 0.8^2 \times (1-0.8)\\ &= 0.16.\end{aligned}$$

It is left as an exercise for readers to verify that for any reported value $p' \neq p$, the expected penalty would be larger. It is thus in the interest of scientists to report their actual belief. Note that the quadratic scoring rule can also be used for the elicitation of conditional probabilities (e.g., for an event E given another event H) when supposing that the penalties only apply if the conditioning event holds (e.g., de Finetti, 1972).

The quadratic scoring rule is a *proper scoring rule* because it implies that whatever one's true belief p, one should sincerely choose this value as one's reported probability; it is the choice with the minimal expected penalty. Other scoring rules may not necessarily exhibit this property. For example, the simple difference between the actual truth value of the event for which a probability is to be assessed and the reported probability p' does not imply optimality for sincerely reported beliefs. Historically, the quadratic scoring rule appears in many writings of the subjective probabilist Bruno de Finetti (e.g. de Finetti, 1962, 1982). However, following a paper by Brier (1950) on an application in meteorology, the rule is also known as 'Brier's Rule'. The quadratic scoring rule is one of the most simple proper scoring rules, while others exist based on, for example, the logarithm (Good, 1952). See also Parmigiani and Inoue (2009) for further details on scoring rules and related concepts, such as calibration. Biedermann et al. (2013) discuss scoring rules in the context of forensic individualization and the use of influence diagrams (Bayesian decision networks) for implementation. The implications of this viewpoint for the subjectivist interpretation of probability is considered in Biedermann (2015), Biedermann et al. (2017) and Biedermann and Vuille (2018).

5.3.2.3 Other forensic decision problems: consignment inspection

Secion 5.3.2.2 considered a problem in which the space of decisions covered any value in the interval between zero and one, including these endpoints, while the states of nature were binary. Consider now a situation in which the states of nature are continuous, but the space of decisions is binary. Suppose an unknown proportion, denoted by θ, such as the proportion of a consignment of individual items that are of a certain kind (e.g., contain an illegal substance) is the object of interest.

The proportion θ may take values in the range $[0, 1]$, including the endpoints. Suppose that the decision maker faces two available decisions, denoted by d_1 and d_2. The first decision, d_1, amounts to accepting the view that the proportion θ of 'positive' units (i.e., units with illegal content) in the consignment is not greater than some specified value θ^*, for example $\theta^* = 0.95$. The second decision, d_2, is the view according to which the proportion θ of positive units in the consignment is greater than the specified value θ^*. Formally, the two decisions d_1 and d_2 can be conceptualised as decisions to accept one of the two composite hypotheses $H_1 : \theta \leq \theta^*$ and $H_2 : \theta > \theta^*$. A decision d_i is accurate if the true value of the unknown proportion θ lies in the range of values defined by such a hypothesis H_i. Otherwise, it is an incorrect decision.

Assume that the undesirability of decision consequences is quantified in terms of losses. There is no positive loss associated with accurate decisions, though incorrect decisions have an associated positive loss. Formally, denote by $\mathrm{L}(d_i, \Theta_j)$ the loss associated with the decision d_i, for $i = 1, 2$, while Θ_j, for $j = 1, 2$, is the true state of affairs. The term Θ_j is defined as follows: $\Theta_1 = [0, \theta^*]$ and $\Theta_2 = (\theta^*, 1]$. Thus, considering that accurate decisions consequences have zero loss is expressed by $\mathrm{L}(d_i, \Theta_j) = 0$, for $i, j \in \{1, 2\}$ and $i = j$. Further, let l_i denote the loss $\mathrm{L}(d_i, \Theta_j)$ associated with an erroneous conclusion when decision d_i is taken, $i, j \in \{1, 2\}$ and $i \neq j$. The so-built loss function is also called a $0 - l_i$ loss function.

Generally, there are different ways to implement a loss function considering both monetary and non-monetary components. Unlike a $0 - 1$ loss function, derived from the $0 - 1$ utility function presented above, the analysis here will interpret the losses l_i as purely monetary values, with l_1 representing the loss when the decision maker falsely regards a case as one in which $\theta \leq \theta^*$. Similarly, l_2 represents the loss from falsely considering the proportion θ of a consignment to be greater than θ^*. The monetary interpretation of losses is based on the following considerations. Suppose that the decision maker is a member of an investigative authority facing the practical problem of high workload. Thus, there may be an interest to focus primarily on cases where the proportion of seized items is above a certain threshold. The loss l_1, then, could consist of the funds or monetary value of property that could have been confiscated by the investigative authority as a penalty, and given to the public treasury. In turn, for assessing l_2, it is relevant to inquire about the consequence of pursuing a case that is not 'important' enough (i.e., falsely considering $\theta > \theta^*$). Here, the investigative authority might generate expenses which, when compared to the reduced funds that may be seized in a non-priority case with $\theta \leq \theta^*$, could represent a net loss. Also, the loss could represent the amount of compensation to be allocated to an erroneously pursued individual. As an example, consider $l_1 = l_2 = 100\text{K USD}$, but readers may choose their own values, including asymmetric values $l_1 \neq l_2$.

To assign probabilities for the composite hypotheses $H_1 : \theta \in \Theta_1$ and $H_2 : \theta \in \Theta_2$, it is necessary to specify a probability distribution for θ. When the number of units in the consignment is large, it is possible to assume a continuous probability density function, such as a beta distribution with parameters α and β. Note that a beta distribution is chosen here because it allows one to readily incorporate sampling information regarding the number of positive units (i.e., units with illegal content), using a standard updating rule (Bernardo and Smith, 1994). For further discussion in forensic contexts see, for example, Aitken (1999), Aitken and Taroni (2004) and Taroni et al. (2010).

Initially, before inspecting any items of the consignment, suppose that all possible values the proportion θ may assume are considered equally probable, expressed in terms of a so-called uniform prior probability distribution. The probability of the hypothesis $H_1 : \theta \leq 0.95$ is then given by the integral of the uniform beta density with parameters $\alpha = \beta = 1$, with endpoints 0 and $\theta^* = 0.95$:

$$\Pr(\theta \leq 0.95) = \int_0^{0.95} f(\theta \mid \alpha = 1, \beta = 1) d\theta = 0.95 \tag{5.18}$$

From this result it follows that a priori $\Pr(\theta > 0.95) = 0.05$. On the basis of these probabilities one can calculate the expected loss of each decision d_i, for $i = 1, 2$. The result is written, for short, $\mathrm{EL}^0(d_i)$, where the superscript 0 refers to the 'initial' point of time, that is before any items are inspected. This loss is also sometimes called *prior expected loss*, whereas the posterior expected loss (i.e, after considering sampling information) is based on the posterior probability density for θ. In the case considered here, prior to considering sampling information, the assumed uniform probability distribution implies that the decision d_1 to accept the proposition H_1: $\theta \leq 0.95$ involves a 0.95 probability for a zero loss, and a 0.05 probability for a loss l_1. Therefore, the prior expected loss of decision d_1 is

$$\begin{aligned} \mathrm{EL}^0(d_1) &= \Pr(\theta \le 0.95) \times \mathrm{L}(d_1, \Theta_1) + \Pr(\theta > 0.95) \times \mathrm{L}(d_1, \Theta_2) \\ &= 0.95 \times 0 + 0.05 \times l_1 = 0.05 \times l_1 \ . \end{aligned} \tag{5.19}$$

The prior expected loss of decision d_2 to accept the proposition H_2: $\theta > 0.95$ is obtained in the same way:

$$\begin{aligned} \mathrm{EL}^0(d_2) &= \Pr(\theta \le 0.95) \times \mathrm{L}(d_2, \Theta_1) + \Pr(\theta > 0.95) \times \mathrm{L}(d_2, \Theta_2) \\ &= 0.95 \times l_2 + 0.1 \times 0 = 0.95 \times l_2 \ . \end{aligned} \tag{5.20}$$

The optimal decision d^0_{opt} will be the one which minimizes $\mathrm{EL}^0(d_i)$. The losses $l_1 = l_2 =$ 100K USD defined above thus lead to the following result:

$$\mathrm{EL}^0(d_1) = 5\ 000,\ \mathrm{EL}^0(d_2) = 95\ 000,\ \Rightarrow d^0_{opt} = d_1.$$

This result means that in a situation in which (i) the decision maker has not yet inspected any items of the consignment, (ii) prior beliefs about θ are based on an uniform prior distribution, and (iii) the loss function is symmetric (therefore the decision is based entirely on the prior beliefs about θ), it is preferable to conclude d_1, that is the proportion is smaller than 0.95.

It is important to note that the above result is crucially dependent on the decision maker's prior beliefs about the proportion θ and on the choice of the loss function. The optimal decision may change depending on the assigned probabilities and losses. To illustrate this dependency, suppose now that the initial beliefs of the decision maker (based on knowledge about previous consignments, domain expertise, and other sources of information) are represented by a beta$(3, 0.3)$ distribution, a distribution that places more density to high values of θ. Specifically, this distribution implies that approximately 60% of prior belief weight is given to values of θ that are greater than 0.95:

$$\Pr(\theta > 0.95) = \int_{0.95}^{1} f(\theta \mid 3, 0.3) d\theta = 0.6.$$

It follows from this that $\Pr(\theta \le 0.95) = 0.4$ and the prior expected losses of decisions d_1 and d_2 become:

$$\begin{aligned} \mathrm{EL}^0(d_1) &= 0.6 \times 100\mathrm{K} = 60\mathrm{K}, \\ \mathrm{EL}^0(d_2) &= 0.4 \times 100\mathrm{K} = 40\mathrm{K},\ \Rightarrow d^0_{opt} = d_2. \end{aligned}$$

The result now is that it is advisable for the decision maker to decide d_2, that is the decision to consider $\theta > 0.95$, because the expected loss associated with this decision is lower than that for d_1.

A particular assumption in the above example is that the loss function is taken to be symmetric. There is no requirement for this, however, and it is possible to formulate the decision theoretic criterion more generally. In particular, decision d_1 should be taken when the expected loss $\mathrm{EL}(d_1)$ is minimal, that is when

$$\Pr(\theta \in \Theta_2) \times l_1 < \Pr(\theta \in \Theta_1) \times l_2.$$

By rearranging terms, this criterion can be reformulated as follows: decide d_1 whenever $\Pr(\theta \in \Theta_2) < l_2/(l_1 + l_2)$. Thus, when the losses for adverse outcomes are considered equal, the action with the minimum expected loss is the one in which the associated parameter values have the higher probability.

Besides computing expected losses and determining optimal decisions, before or after taking into account sampling information, the above framework presents a starting point for a variety of further analyses and the development of additional concepts. For example, the expected loss of the a priori optimal action is also sometimes referred to as the *expected value of perfect information* (EVPI) about the true state of θ. Although, in many situations, it may not be possible to obtain perfect information about the true state of nature, the EVPI may be a useful measure to think about the decision problem. In particular, it allows one to indicate the maximum amount of money that one should be willing to pay for (expert) information that is such that it would allow one to determine the true state of nature with certainty. In the above example, where the a priori optimal decision d^0_{opt} had an expected loss of 40K, the decision maker should accept additional information about the true proportion only if the cost for that information does not exceed 40K.

Another notion of interest in this context is the *value of sample information* (VSI), defined as the difference between the expected losses of the optimal actions before and after considering sampling information. Yet another concept are pre-posterior analyses that take into account the probabilities for various outcomes of item inspection (i.e., the proportion of inspected items that are of a certain kind). Such analyses may be conducted for fixed or variable sample sizes, and by taking into account or not the cost of inspecting items from the consignment. This leads to further notions, such as the *expected value of sample information* (EVSI) and the *expected net value of sample information* (ENVSI). See, for example Biedermann et al. (2012), Taroni et al. (2014) and Gittelson (2013) for a discussion of these concepts and the use of graphical models, such as decision trees and Bayesian decision networks (influence diagrams) for practically implementing these approaches. More generally, decision-theoretic approaches to sampling can be found, for example, in Schlaifer (1959) and Raiffa and Schlaifer (1961).

Note also that the discussion in this section concentrated on a binary decision among two composite hypotheses regarding a parameter space. A different decision problem is to consider the whole parameter space as the space of possible decisions. That is, given a parameter space and a posterior probability distribution over the possible values that the parameter may take, the question is which single value to select is, in some sense, optimal. This amounts to considering a problem of parameter estimation, that is inference, as a decision. In this case, other mathematically tractable non-constant loss functions, such as the quadratic loss or the piecewise linear loss function, can be implemented (Press, 2003). For general theory on this perspective, see Berger (1985), and for forensic applications Taroni et al. (2010).

5.4 Discussion and conclusions

The core topic of this book, the application of statistics in forensic science, is primarily concerned with questions of inference – that is the reasonable reasoning in the face of uncertainty. This involves procedures for the coherent use of relevant data for revising and informing beliefs about competing propositions of interest. Propositions may be formulated at different levels regarding, for example, the source of particular (forensic) trace material (e.g., DNA, fibres, etc.), or at a more advanced stage, regarding alleged activities of particular individuals, such as the defendant (Cook et al., 1998). Two complications arise when taking this abstract account too literally. One is the occasionally raised claim that statistics can only be applied if data are available. This is not so. Reasoning methods, in particular the specification of conditions for coherence, reply to general questions of formal analysis that can be approached with whatever amount of information and new data – whereas the latter may be nil – there may be. The second complication is the flawed idea that inference is the end of the matter. Again, this is not so. As noted by Lindley (2000), inference is a preliminary to decision, an idea that started to expand more widely since the middle of the last century:

> "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." (Chernoff and Moses, 1959, at p. 1)

The neat connection between, on the one hand, informing beliefs through collected evidence and data, and addressing the question of how to decide, on the other hand, is also at the heart of applications in forensic science and the law. In these areas, decision theory has, over the past few decades, been considered by many authors as an analytical framework for studying selected decision problems. Though not as expanded as in economics and operations research, existing studies in the law are more numerous and also more controversially debated than applications to forensic science problems, which have a more recent history. It is widely acknowledged that decision theory provides a rigorous framework for thinking about decision problems, but it is also widely uncontested that the direct application in practice may not be immediate. The main reason for this is that additional argument needs to be invoked in order to interpret the formal elements of the theory with respect to the defining features of practical decision problems. This is a general challenge, however, that is equally encountered with other formal concepts, such as probability and Bayes' theorem. In the particular context of forensic science, decision theory is currently used to critically review understandings of traditional concepts, such as individualisation/identification, and practice thereof. Insight that is gained from such analyses helps better understand where current forensic practice makes assumptions that go above and beyond forensic examiners' areas of competence (Biedermann et al., 2016). Typical examples for this include assumptions about probabilities for propositions of interest and utilities/losses for decision consequences. The current role of decision theory in forensic science thus is advisory, by providing a reference point – in a normative sense (Section 5.1) – for delineating areas that require attention in ongoing reform efforts (Cole and Biedermann, 2020).

5.5 Further readings

Forensic science

Practising forensic scientists may encounter questions of decision making at various stages in their daily work. For example, scientists may need to decide about whether or not to use a particular analytical apparatus, apply a particular chemical substance or search for a particular type of transferred trace material. Several publications have addressed such questions, in particular in the context of forensic DNA analyses. Taroni et al. (2005) considered the question of whether or not to perform DNA analyses in a case of questioned kinship involving two individuals who are uncertain about whether they are full siblings or unrelated. Another question related to DNA analysis, addressed by Taroni et al. (2007), concerns the number of DNA loci that ought to be analyzed. Planning problems in forensic

DNA analyses are also addressed in Mazumder (2010). Gittelson et al. (2014) used a decision-theoretic approach for the topic of genotype designation, that is a decision problem characterised by complications due to phenomena such as drop-in and drop-out. Decision-theoretic computations can readily become complex, in particular when computations need to be extended beyond one-stage analyses. An example for a staged decision analysis regarding the question of processing or not processing a fingermark is given in Gittelson et al. (2013), focusing on the expected value of information (EVOI) and the cost of processing the fingermark. Other examples of sequential decisions are presented in Taroni et al. (2010, 2014). General forensic applications of Bayesian decision theoretic criteria, for example in the context of kinship analyes and handwriting examinations, are presented in Biedermann et al. (2018). A Bayesian classification criteria is presented in Bozza et al. (2014) to address the problem of determining plant's chemotype.

General

Ramsey (1931) is credited with pioneering expected utility, along with de Finetti (1937), followed by works of von Neumann and Morgenstern (1953) and Marschak (1950) on the axiomatization in terms of 'gambles'. Classical textbooks are Savage (1954), Luce and Raiffa (1958), Raiffa (1968) and De Groot (1970). More recent texbooks are Berger (1988), French (1988), Smith (1988, 2010), Parmigiani and Inoue (2009) and Robert (2007).

5.6 Acknowledgments

Franco Taroni and Alex Biedermann gratefully acknowledge the support of the Swiss National Science Foundation through grants No. IZSEZ0_19114 and No. BSSGI0_155809. Some of the writing of this chapter was carried out during a visiting research stay of Alex Biedermann at New York University School of Law.

Bibliography

C G G Aitken. Sampling - How big a sample? *Journal of Forensic Sciences*, 44:750–760, 1999.

C G G Aitken and F Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, second edition, 2004.

J Baron. President's column: normative, descriptive, and prescriptive. Newsletter of the Society for Judgment and Decision Making, December 2006.

J Baron. *Thinking and Deciding*. Cambridge University Press, New York, 4th edition, 2008.

J O Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1985.

J O Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1988.

J M Bernardo and A F M Smith. *Bayesian Theory*. John Wiley & Sons, Chichester, 1994.

A Biedermann. The role of the subjectivist position in the probabilization of forensic science. *Journal of Forensic Science and Medicine*, 1:140–148, 2015.

A Biedermann and J Vuille. The decisional nature of probability and plausibility assessments in juridical evidence and proof. *International Commentary on Evidence*, 16:1–30, 2018.

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.

A Biedermann, S Bozza, P Garbolino, and F Taroni. Decision-theoretic analysis of forensic sampling criteria using Bayesian decision networks. *Forensic Science International*, 223: 217–227, 2012.

A Biedermann, P Garbolino, and F Taroni. The subjectivist interpretation of probability and the problem of individualisation in forensic science. *Science & Justice*, 53:192–200, 2013.

A Biedermann, F Taroni, and C Aitken. Liberties and constraints of the normative approach to evaluation and decision in forensic science: a discussion towards overcoming some common misconceptions. *Law, Probability and Risk*, 13:181–191, 2014.

A Biedermann, S Bozza, and F Taroni. The decisionalization of individualization. *Forensic Science International*, 266:29–38, 2016.

A Biedermann, S Bozza, F Taroni, and C Aitken. The consequences of understanding expert probability reporting as a decision. *Science & Justice, Special Issue on Measuring and Reporting the Precision of Forensic Likelihood Ratios*, 57:80–85, 2017.

A Biedermann, S Bozza, and F Taroni. Analysing and exemplifying forensic conclusion criteria in terms of Bayesian decision theory. *Science & Justice*, 58:159–165, 2018.

S Bozza, J Broséus, P Esseiva, and F Taroni. Bayesian classification criterion for forensic multivariate data. *Forensic Science International*, 244:295–301, 2014.

G W Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78:1–3, 1950.

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.

C Champod, C Lennard, P Margot, and M Stoilovic. *Fingerprints and Other Ridge Skin Impressions*. CRC Press, Boca Raton, second edition, 2016.

H Chernoff and L E Moses. *Elementary Decision Theory*. John Wiley & Sons, New York, 1959.

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.

S A Cole and A Biedermann. How can a forensic result be a "decision"? A critical analysis of ongoing reforms of forensic reporting formats for federal examiners. *Houston Law Review*, 57:551–592, 2020.

R Cook, I W Evett, G Jackson, P J Jones, and J A Lambert. A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice*, 38:231–239, 1998.

F C Dane. In search of reasonable doubt: a systematic examination of selected quantification approaches. *Law and Human Behavior*, 9:141–158, 1985.

B de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* **7** 1-68 (english translation). In H E Kyburg and H E Smokler, editors, *Studies in Subjective Probability* (1980), pages 93–158. Dover, New York, 2 edition, 1937.

B de Finetti. Does it make sense to speak of 'Good Probability Appraisers'? In I J Good, editor, *The Scientist Speculates: An Anthology of Partly-Baked Ideas*, pages 357–364. Basic Books, New York, 1962.

B de Finetti. *Probability, Induction and Statistics, The Art of Guessing*. John Wiley & Sons, New York, 1972.

B de Finetti. *Theory of Probability, A Critical Introductory Treatment, Volume 1*. John Wiley & Sons, London, 1974.

B de Finetti. The proper approach to probability. In G Koch and K Spizzichino, editors, *Exchangeability in Probability and Statistics*, pages 1–6. North-Holland Publishing Company, Amsterdam, 1982.

M H De Groot. *Optimal Statistical Decisions*. Wiley Classics Library Edition Published 2004. John Wiley & Sons, Inc., Hoboken, New Jersey, 1970.

I W Evett, J A Lambert, and J S Buckleton. A Bayesian approach to interpreting footwear marks in forensic casework. *Science & Justice*, 38:241–247, 1998.

S French. *Decision Theory, An Introduction to the Mathematics of Rationality.* Ellis Horwood Limited, Chichester, 1988.

S French, J Maule, and N Papamichail. *Decision Behaviour, Analysis and Support.* Cambridge University Press, Cambridge, 2009.

R D Friedman. Answering the Bayesioskeptical challenge. *The International Journal of Evidence & Proof, Special Issue*, 1:276–291, 1997.

R D Friedman. *The Elements of Evidence.* West Academic Publishing, Saint Paul, Minnesota, fourth edition, 2017.

S Gittelson. *Evolving from inferences to decisions in forensic science.* PhD thesis, The University of Lausanne, School of Criminal Justice, Lausanne, 2013.

S Gittelson, A Biedermann, S Bozza, and F Taroni. The database search problem: A question of rational decision making. *Forensic Science International*, 222:186–199, 2012.

S Gittelson, S Bozza, A Biedermann, and F Taroni. Decision-theoretic reflections on processing a fingermark. *Forensic Science International*, 226:e42–e47, 2013.

S Gittelson, A Biedermann, S Bozza, and F Taroni. Decision analysis for the genotype designation in low-template-DNA profiles. *Forensic Science International: Genetics*, 9: 118–133, 2014.

I J Good. *Probability and the weighing of evidence.* Griffin, London, 1950.

I J Good. Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14:107–114, 1952.

R Hastie. Algebraic models of juror decision making. In R Hastie, editor, *Inside the Juror. The Psychology of Juror Decision Making*, pages 84–115. Cambridge University Press, New York, 1993.

J Kaplan. Decision theory and the factfinding process. *Stanford Law Review*, 20:1065–1092, 1968.

D H Kaye. What is Bayesianism? In P Tillers and E D Green, editors, *Probability and Inference in the Law of Evidence, The Uses and Limits of Bayesianism (Boston Studies in the Philosophy of Science)*, pages 1–19. Springer, Dordrecht, 1988.

D H Kaye. Clarifying the burden of persuasion: what Bayesian decision rules do and do not do. *The International Journal of Evidence & Proof*, 3:1–29, 1999.

D V Lindley. *Introduction to Probability and Statistics from a Bayesian Viewpoint: Part 2 Inference.* Cambridge University Press, Cambridge, U.K., 1965.

D V Lindley. *Making Decisions.* John Wiley & Sons, Chichester, second edition, 1985.

D V Lindley. The philosophy of statistics. *The Statistician*, 49:293–337, 2000.

R D Luce and H Raiffa. *Games and decisions : introduction and critical survey.* Wiley, New York, 1958.

J Marschak. Rational behavior, uncertain prospects, and measurable utility. *Econometrica*, 18:111–141, 1950.

A Mazumder. *Planning in Forensic DNA Identification Using Probabilistic Expert Systems.* PhD thesis, University of Oxford, Department of Statistics, Oxford, 2010.

G Parmigiani. Decision theory: Bayesian. In N Smelser and P Baltes, editors, *International Encyclopedia of Social and Behavioral Sciences*, pages 3327–3334. Elsevier, Oxford, 2001.

G Parmigiani and L Inoue. *Decision Theory: Principles and Approaches.* John Wiley & Sons, Chichester, 2009.

L Piccinato. *Metodi per le decisioni statistiche.* Springer-Verlag, Milano, 1996.

J W Pratt, H Raiffa, and R Schlaifer. The foundations of decisions under uncertainty: an elementary exposition. *Journal of the American Statistical Association*, 59:353–375, 1964.

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.

S J Press. *Bayesian Statistics: Principles, Models, and Applications.* John Wiley & Sons, Ltd, New York, 1989.

S J Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications.* Wiley-Interscience, Hoboken, second edition, 2003.

H Raiffa. *Decision Analysis, Introductory Lectures on Choices under Uncertainty.* Addison-Wesley, Reading, Massachusetts, 1968.

H Raiffa and R Schlaifer. *Applied Statistical Decision Theory.* The M.I.T. Press, Cambridge, Massachusetts, 1961.

F P Ramsey. Truth and probability. In R B Braithwaite, editor, *The Foundations of Mathematics and Other Logical Essays*, pages 156–198. Routledge & Kegan Paul Ltd, 1931. Reprinted in *Studies in Subjective Probability* 2 edn. (1980) (ed. Kyburg HE and Smokler HE) Dover New York pp. 61-92.

C P Robert. *The Bayesian Choice, A Decision-Theoretic Motivation.* Springer, New York, second edition, 2007.

L J Savage. *The Foundations of Statistics.* Wiley, New York, 1954.

R Schlaifer. *Probability and Statistics for Business Decisions.* McGraw-Hill Book Company, Inc., New York, 1959.

R J Simon and L Mahan. Quantifying burdens of proof: A review from the bench, the jury, and the classroom. *Law and Society Review*, 5:319–330, 1971.

J Q Smith. *Decision analysis: a Bayesian approach.* Chapman and Hall, London, 1988.

J Q Smith. *Bayesian Decision Analysis, Principles and Practice.* Cambridge University Press, Cambridge, 2010.

D J Spiegelhalter, K R Abrams, and J P Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* John Wiley & Sons, Chichester, 2004.

D A Stoney. Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *Journal of the Forensic Science Society*, 34:17–21, 1994.

F Taroni, S Bozza, and C G G Aitken. Decision analysis in forensic science. *Journal of Forensic Sciences*, 50:894–905, 2005.

F Taroni, S Bozza, M Bernard, and C Champod. Value DNA tests: a decision perspective. *Journal of Forensic Sciences*, 52:31–39, 2007.

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.

F Taroni, A Biedermann, S Bozza, P Garbolino, and C G G Aitken. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. Statistics in Practice. John Wiley & Sons, Chichester, second edition, 2014.

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.

P Tillers and E D Green, editors. *Probability and Inference in the Law of Evidence. The uses and limits of Bayesianism*. Kluwer Academic Publishers, Dordrecht, 1988.

J von Neumann and O Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 3rd edition, 1953.