

The principle of total evidence reprised

Franco Taroni ^{1,*}, Colin Aitken², Silvia Bozza^{3,1}, Patrick Juchli¹

¹School of Criminal Justice, The University of Lausanne, Lausanne, Switzerland

²School of Mathematics and Maxwell Institute of Mathematical Sciences, The University of Edinburgh, Edinburgh, United Kingdom

³Department of Economics, Ca' Foscari University of Venice, Venice, Italy

*Corresponding author. School of Criminal Justice, The University of Lausanne, Lausanne, Switzerland. E-mail: Franco.Taroni@unil.ch

Abstract

The Principle of Total Evidence, postulated by Carnap in 1947, implies that to achieve the best assignation of a probability, all available information should be considered, subject to cost. For the evaluation of evidence, it is important that the best assignation of probability be made. The benefits of such an assignation are shown to be an increase in the expected utility of any decision made, a decrease in an expectation of an error of an inference that might be made and an increase in the expected weight of evidence for the correct hypothesis. A practical illustration is given with reference to a recent Italian case.

Keywords: Bayes' theorem; Bayes factor; evidence evaluation; expected error; expected utility; principle of total evidence.

1. Introduction: rationale for acquiring new evidence

The Principle of Total Evidence,¹ postulated by Carnap (1947), recommends that, in the assignation of a probability, all available information should be considered.²

In the context of a criminal trial, the trier of fact (judge or jury) should, ideally, be as informed as possible about the case under trial as is practically feasible. The Principle recommends that all evidence be considered. Such a recommendation has to be tempered with the cost of the acquisition of evidence. Decisions about the acquisition of evidence are part of the investigation of a crime. At the trial, when the cases for the prosecution and the defence are fully prepared, all evidence has been acquired. There is no cost to the court associated with the admission of any evidence. There is only a cost of resources within the trial process to be considered when decisions are being taken as to whether or not to admit formally particular items of evidence. Various benefits that arise from the admission of new evidence are discussed. These benefits have been introduced in the philosophical literature and are presented anew here in the context of a criminal case and with numerical verifications of the benefits that arise from the admission of additional evidence in order to bring attention to them to the wider readership of lawyers and forensic scientists.

The case used as an illustration of these ideas is the Taffio Palmi (victim)–Busetto (convicted defendant) (TPB) case in Venice, Italy, where a victim was found stabbed to death in her flat where she lived alone. According to several witnesses, the victim used to wear a golden necklace, which was not found on the body, nor at the crime scene nor anywhere else in the victim's flat. About a month later, a person of interest, PoI, was suspected of the crime and apprehended.

¹ Hereafter, the Principle.

² From a historical point of view, it is of interest that Carnap (1947) (at page 138) refers to Keynes (1921) who wrote 'Bernoulli's second maxim, that we must take into account all the information we have, amounts to an injunction that we should be guided by the probability of that argument, amongst those of which we know the premises, of which the evidential weight is the greatest. But should not this be re-enforced by a further maxim, that we ought to make the weight of our arguments as great as possible by getting all the information we can?' (p. 84). The Principle also underlies Locke (1689)'s statement quoted by Keynes: 'He that judges without informing himself to the utmost that he is capable, cannot acquit himself of judging amiss' (p. 89).

Received: 18 May 2023. Revised: 25 May 2024. Accepted: 13 August 2024

© The Authors (2024). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

During a search of the PoI's flat, a broken necklace was recovered and seized for forensic analysis.

The PoI declared that the necklace was a family heirloom and that it had not been touched in years. The prosecution claimed that the necklace belonged to the victim, and that it was snatched off the victim's neck by the PoI during the commission of the crime. The question about whether that necklace was the victim's or not became crucial evidence in the trial. A presumptive test for the detection of human cells gave a negative result, and no DNA profile was obtained by this analysis. A few months later, a second forensic laboratory repeated the DNA analysis on the necklace and obtained a complete profile corresponding to that of the victim. This second result played a key role at the trial despite the fact that potential contamination in the lab could not be excluded. Indeed, the court established that the necklace was the one that had been seized by the offender from the victim during the offence that led to the death of the victim. The PoI was then given a life sentence for the murder of the victim. The sentence was confirmed by the Court of Appeal and then by the Supreme Court.³

Other forensic elements were retrieved, and information from them was available at the time of the trial. For example, a relevant⁴ red-coloured fingermark was observed and collected from a wall in the victim's apartment; the mark was composed of the victim's blood and its minutiae did not correspond to any of the PoI's fingerprints nor to any of the victim. Shoe-marks were registered at the crime scene. The marks did not correspond to any of the PoI's shoes. Other items belonging to the PoI (e.g. shoes, slippers, carpets, cleaning rags) were analysed by the first laboratory for possible traces of the victim's blood; no such traces were detected. The Court just considered one item of evidence, the DNA result from the second laboratory, the result which declared a correspondence between the victim's DNA profile and that observed (in small quantities) on the necklace.⁵

This scenario is alarming because additional information (i.e. absence of DNA on a series of items, negative evidence for shoe-marks and for the fingermark) is available at no cost (since all the tests had been done) relevant to the uncertain event of judicial interest (i.e. the hypothesis that the PoI stabbed to death the victim). The expected utility for any decision involving new information will not decrease as a result of the new information. This result has long been known by philosophers of science and statisticians. For example, [Salmon \(1992\)](#) affirmed:

An inductive argument gives a high degree of confirmation to its conclusion only if there is no additional evidence available at the time the argument is formulated to change the degree of confirmation. (free translation, at p. 100)

In its original form, [Salmon \(1966\)](#) wrote:

Given that the degree of confirmation of hypothesis h on evidence e is p , and given also the truth of e , we are not allowed to infer h even if p is very near one. Rather, we must use our inductive logic according to certain definite rules of applications. First, there is the requirement of total evidence. If e is the evidence statement we are going to use, it must incorporate *all* (authors' italics) relevant available evidence. This is an important respect in which inductive logic differs from deductive. (p. 76)

Similarly, [Lindley \(1985\)](#) clarified that 'the information is always expected to be of value' (p. 131). The critical point related to the use of limited amount of information by a Court of Justice is well expressed by [Lindley \(1985\)](#):

Our result says that the cost should be the only reason for ruling evidence to be inadmissible: that if evidence is virtually cost-free then it should be admitted, for it is expected to be of value

³ Assise Venezia 22.12.2014, n.1; Assise Appello Venezia 18.11.2016; Cassazione Penale 26.04.2018, n. 37002.

⁴ In this context, the term 'relevant' denotes that the evidence recovered from the crime scene is connected with the crime and hence has been left by the offender, as defined by [Stoney \(1991, 1994\)](#).

⁵ The interested readers can refer to [Taroni, Bozza, and Garbolino \(2018a,b\)](#), [Gennari \(2021\)](#) and [De March and Taroni \(2020\)](#) for further details and some analysis of this criminal case.

in judging the case. This goes against current legal practice. Thus English law does not allow evidence of bad character to be used to increase the probability that the defendant is guilty. Our argument says it should if the evidence is almost free. We are not saying that the law is incorrect: we are saying that the situation needs reconsideration in the light of expected utility theory. It may be, for reasons that are not clear to me, that some material fact has been forgotten in applying the theory to legal practice. (p. 132)

The idea that any scientific or judicial decision should be based on the available evidence with its suggestion of the importance of gathering more evidence has been explored in the past (see e.g. [Good 1985](#) and [Raiffa and Schlaifer 1961](#)). This has been done from various points of view, for example, through a decision making approach calling for a calculus of the expected value of sample information ([Good 1967](#)) and through the expectation of the weights of evidence considered as a criterion for the value of an experimental design ([Good 1979](#)). The idea that it is better to use all the available information can also be viewed as a simple consequence of Bayesian confirmation theory, where posterior probabilities are updated by the acquisition of new information in order to discriminate better between hypotheses.

Historically, it was [Ayer \(1957\)](#) who first considered the importance of making new observations. He related this consideration to [Carnap \(1947\)](#) ‘Principle’ where it was recommended to use all the available information when assigning⁶ a probability; in other words, that ‘the requirement of total evidence [...] says that in evaluating a hypothesis, you should take account of all evidence you have’ ([Barrett and Sober 2020](#): 191). From a Bayesian perspective, one should condition on all the information you have to assess how probable a hypothesis is. This approach represents an example of what Carnap defined as the ‘methodology of induction’. As paraphrased by [Salmon \(1966\)](#):

The methodology of induction contains rules for the application of inductive logic—that is, rules that tell us how to make use of the statements of degree of confirmation in deciding courses of practical action. As I indicated, the requirement of total evidence is one of the important methodological rules, and the rules of maximizing estimated utility is another. These rules tell us how to use the results of inductive logic’. (p. 93)⁷

The main point of interest in this article is not how to measure the support given by any acquired information [whether through the quantification of the Bayes factor after having acquired such information or during a pre-assessment procedure as developed by [Cook et al. \(1998, 1999\)](#)] or to measure the way such a current or future finding may impact on a course of action. These ideas are well-documented. The quantification of the expected utility (or loss)⁸ related to a list of actions and of the gain in the acquisition of new information have been described in literature through practical examples. For example, given that the decision to collect new evidence has to be taken before the evidence is available, the problem is the calculation of the expected gain of this new (unknown) evidence so that the gain can be compared with the cost of the search. Provided that utilities, or the losses, of the outcomes of our decisions can be quantified in such a way that they can be compared to the cost of an experiment to gather and analyse new evidence, Bayesian decision theory explains, in what is known as the principle of

⁶ Note that the term ‘estimating a probability’ was used in the original text. As discussed by [Fischhoff and Beyth-Marom \(1983\)](#), ‘The term ‘assign’ is used rather than ‘estimate’ to emphasize that a probability expresses one’s own feelings rather than an appraisal of a property of the physical world. Thus, there is no ‘right’ probability value for a particular statement’ (p. 240).

⁷ [Salmon \(1966\)](#) refers to [Carnap \(1950\)](#). See, e.g., ‘Certainly I and my friends have learned much from other authors, both in the purely mathematical theory of probability and in the methodology of its applications’ (p. xiv), or the more complete paragraph ‘In the application of inductive logic still another difficulty is involved, which does concern inductive logic itself. This difficulty consists in the fact that, if an observer wants to apply inductive logic to an expectation concerning a hypothesis *b*, he has to take as evidence *e* a complete report of all his observational knowledge. Many authors on probability have not given sufficient attention to the *requirement of total evidence*. They often leave aside a great part of the available information as though it were irrelevant. However, cases of strict irrelevancy are much more rare than is usually assumed’ (at p. 208).

⁸ Utilities and losses are considered complementary. Utility quantifies the desirability of a consequence of the decision based on the acquisition of new information. Loss quantifies the undesirability of a consequence of the decision based on the acquisition of new information.

rationality, how the expected value of information may be calculated taking advantage of the recommendation to maximize the expected utility. This approach is not new in forensic science. As an illustration, consider a scenario involving information provided by the fingerprints not processed in a forensic science laboratory. Gittelsohn *et al.* (2013) examined the question of processing, or of not processing, a fingerprint from a decision-theoretic point of view⁹ and answered this question through a quantified expression of the expected value of information associated with the processed fingerprint which was compared with the cost of processing the mark.

The main point of interest of the article is, rather, the provision of a formal justification for the intuitive idea that it is rational to decide to search for, acquire and use new information as long as the cost of so doing is negligible when compared with the expected gain of this new information. A corollary of this idea is that all the information available at a given time should be used, again assuming there is no additional cost of doing so. The use of as much information as possible, even when it is incomplete, can only be expected to improve the quality of a decision.

The article is structured as follows: Section 2 will briefly introduce the readers to Bayesian decision theory and the quantification of the expected value of information. This section also develops an alternative proof of the Principle. Section 3 repeats for completeness, the proof of a sentence by Horwich (2016) that ‘the expected error in our probability judgements is minimized by the acquisition of new evidence’. This result is then interpreted in the context of the TPB case. A corollary of the previous proofs is the extension to what has been called by Good ‘The theorem of the expected weight of evidence’ (Good 1985); if an alternative hypothesis is false, it is expected that the logarithm of the Bayes factor (weight of evidence) in its favour to be less than 0 in future experiments. This aspect is described in Section 4 with an application to the TPB case. Note that the Bayes factor may be thought of as the *value* of evidence and the logarithm of the Bayes factor may be thought of as the *weight* of evidence. A discussion (Section 5), with some thoughts on the measurement of utility, concludes the article.

2. Bayesian decision theory

2.1 Expected utility (loss) and the expected value of perfect information

A standard procedure in decision theory is the quantification of the expected gain (or profit) of a certain action. Given that generally the decision to acquire new information has to be taken before the information is available, the problem is to calculate the expected gain of this new information, so that the gain can be compared with the cost of the search or acquisition. However, in the TPB case (and in many others that come to trial), the cost of the collection and use of the forensic findings can be ignored because all the findings were already available to the court, so the cost of consideration is negligible.

Provided that the utilities (or the losses) of the outcomes of the possible actions (say, deciding that the PoI stabbed the victim to death, or deciding that some other person stabbed the victim to death) may be quantified in such a way that they can be compared to the cost of the experiment, Bayesian decision theory explains how the expected value of information may be calculated.¹⁰

Imagine the advice to a judge who has to decide which one of two alternative hypotheses, say H_1 and H_2 , to choose and suppose also that these two hypotheses can be considered as if they were exhaustive. Denote by d_1 and d_2 the set of available courses of action, and by C_{ij} the consequence of taking decision d_i ($i = 1, 2$) when hypothesis H_j ($j = 1, 2$) turns out to be true. A utility (loss) function can be introduced to assess the desirability (undesirability) of decision outcomes (consequences) and is denoted by $U(C_{ij})$ ($L(C_{ij})$). The decision matrix is represented in Table 1 when dealing with utilities (panel A), or with losses (panel B). An illustration is given in Table 2 where a utility equal to 1 is associated to the most desirable consequences C_{11} and C_{22}

⁹ The theory of such an approach can be found in Lindley (1985) with a series of forensic applications in Taroni *et al.* (2010) and Gittelsohn (2013).

¹⁰ The interested reader can refer to the seminal book of Lindley (1985) for an extended presentation of Bayesian decision theory and to Taroni, Bozza, and Biedermann (2020) for a description of the role of decision theory in forensic science. A summary of the theory is given in Aitken Taroni, and Bozza (2021) with other forensic scientific examples given, for example, in Biedermann, Bozza, and Taroni (2008, 2016, 2020).

Table 1. Decision matrix with d_1 and d_2 denoting the possible actions, H_1 and H_2 denoting the states of nature (the hypotheses of interest), C_{ij} denoting the consequence of deciding d_i when hypothesis H_j is true, $U(\cdot)$ denoting the utility function quantifying the desirability of decision consequences C_{ij} (panel A), and $L(\cdot)$ the loss function quantifying the undesirability of decision consequences C_{ij} (panel B).

Panel A	H_1	H_2
d_1 : choosing H_1	$U(C_{11})$	$U(C_{12})$
d_2 : choosing H_2	$U(C_{21})$	$U(C_{22})$
Panel B	H_1	H_2
d_1 : choosing H_1	$L(C_{11})$	$L(C_{12})$
d_2 : choosing H_2	$L(C_{21})$	$L(C_{22})$

Table 2. Decision matrix with d_1 and d_2 denoting the possible actions, H_1 and H_2 denoting the states of nature (the hypotheses of interest) and U_{ij} (L_{ij}) denoting the utility (loss) quantifying the desirability (undesirability) of decision consequences C_{ij} . A utility (loss) equal to 1 (0) is associated to the most desirable (most undesirable) consequences.

Panel A	H_1	H_2
d_1 : choosing H_1	1	U_{12}
d_2 : choosing H_2	U_{21}	1
Panel B	H_1	H_2
d_1 : choosing H_1	0	L_{12}
d_2 : choosing H_2	L_{21}	0

(panel A). The loss function that is derived following these choices is characterized by a loss equal to 0 associated to the least undesirable consequences C_{11} and C_{22} (Panel B).

The best decision is that one for which the expected utility (loss) is the maximum (minimum). In symbols—using I to denote background information already available about the two hypotheses, and E to denote the extra information—a rational decision maker will take decision d_i ($i = 1, 2$) as the one corresponding to the i for which

$$\sum_{j=1}^2 U_{ij} Pr(H_j|E, I) \quad (1)$$

is maximized, or

$$\sum_{j=1}^2 L_{ij} Pr(H_j|E, I) \quad (2)$$

is minimized.

There is an argument that this discussion should be considered in terms of only one of loss or utility. However, both terms are used regularly and their inclusion here eases comparison with other work. For those unfamiliar with the terms, they may perhaps be more easily thought of as costs and benefits.

An investigator has to decide whether to carry out an experiment (to acquire new information) the result of which is denoted E . The problem is to determine how much to pay for that experiment. As a first step towards the solution of the problem, it is shown how to calculate the *expected value of perfect information*, that is, how much it would be worth to know with certainty which hypothesis is true. If the true hypothesis were known, the decision with the smallest loss in the column of Table 1 (Panel B) with respect to that hypothesis is the one to take. The expected loss with perfect information is calculated with the multiplication of the minimum loss for each hypothesis by the probability of that hypothesis, followed by the sum all these products:

$$\sum_{j=1}^2 (\min_i L_{ij}) Pr(H_j | E, I). \quad (3)$$

The minimum loss for each hypothesis is zero, so the expected loss of perfect information is zero. The choice before the truth is known is given by Equation (2). The difference between Equations (2) and (3) is the measure of the reduction in the expected loss or, equivalently, the increase in the expected gain that could be obtained with perfect information. In other words, it is the measure of the expected value of perfect information:

$$\min_i \sum_{j=1}^2 L_{ij} Pr(H_j | E, I) - \sum_{j=1}^2 (\min_i L_{ij}) Pr(H_j | E, I). \quad (4)$$

This is the maximum price that should be paid for perfect information. The expected value of perfect information will always be—by definition—greater than zero: indeed, whatever decision is taken without perfect information, the value of Equation (2) will be greater than the value of Equation (3), since every loss L_{ij} in the former is replaced by a loss $(\min_i L_{ij})$ in the latter which cannot be greater.

2.2 Expected value of sample or partial information

If the experiment (or acquisition of information which may be thought of as evidence) were of such kind to tell you the truth, this would be the end of the story. Unfortunately, such an experiment will provide only partial information, changing your probabilities for H_1 and H_2 but without them reaching the extreme values of 1 and 0. The experiment may be short of ideal but still be potentially useful. Denote by $\bar{U}(d_i)$ the expected utility of decision d_i :

$$\bar{U}(d_i) = \sum_{j=1}^2 U_{ij} Pr(H_j | E, I), \quad (5)$$

and by $\bar{L}(d_i)$ the expected loss of decision d_i :

$$\bar{L}(d_i) = \sum_{j=1}^2 L_{ij} Pr(H_j | E, I). \quad (6)$$

The best decision after having conducted the experiment (e.g. a search for a particular type of evidence and observed it, E , or not, \bar{E}) will be the decision that maximizes the expected utility

$$\max_i \bar{U}(d_i), \quad (7)$$

or minimizes the expected loss

$$\min_i \bar{L}(d_i). \quad (8)$$

The optimal decision is often referred to as the *Bayes decision* or *Bayes action*, and it answers the question ‘after observation of the information, how much has been learnt about H_j , and how will this influence one’s choice of action?’ This approach follows the normative theory of decision making based on the principle of maximizing the expected utility (or minimizing the expected loss), which combines, in a mathematical function, both the quantification of the desirability (or undesirability) of decision consequences, expressed in terms of utilities (losses), and the uncertainty about unknown states of nature (e.g. the hypotheses of interest) expressed by probabilities.

The theory of expected utility (loss) can be considered as a theory for rational choices (see, e.g., Jeffrey 1987; Chater and Oaksford 2012; Briggs 2019¹¹). In this respect, if a person fails to

¹¹ Text available at <https://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/>.

prefer acts with higher expected utility (or lower expected loss), then it has been written that that person violates at least one of the axioms of rational preference (Zynda 2000).¹²

It is not known what the outcome of the experiment will be but the likelihoods are known. The problem is ‘how much is expected to be learnt from the information prior its observation, and how will the choice of action be influenced by its knowledge?’ The probabilities for the (extra) evidence can be calculated with an extension of the conversation [Lindley (1991)]:

$$Pr(E|I) = \sum_{j=1}^2 Pr(E|H_j, I)Pr(H_j|I). \quad (9)$$

The action to be chosen is the one that minimizes the loss for any possible result E of the experiment. The expected loss with partial information is the sum of the products of the minimum expected loss for each possible result of the experiment, as shown in Equation (3), and the probability of that result:

$$\sum_E \min_i \sum_{j=1}^2 L_{ij} Pr(H_j|E, I) Pr(E|I). \quad (10)$$

With an application of Bayes’ formula, Equation (10) may be rewritten as

$$\sum_E \min_i \sum_{j=1}^2 L_{ij} Pr(E|H_j, I) Pr(H_j|I). \quad (11)$$

The choice before the experiment is given by the minimum value of Equation (2), the difference between that and Equation (11) is called the *expected value of partial information* (with the sign inverted), also called the *expected value of sample information* (see, e.g., Winkler 2003, Hays and Winkler 1970):

$$\min_i \sum_{j=1}^2 L_{ij} Pr(H_j|I) - \sum_E \min_i \sum_{j=1}^2 L_{ij} Pr(E|H_j, I) Pr(H_j|I). \quad (12)$$

This also represents the maximum price that should be paid to have this partial information.

Similarly, the expected gain in utility is

$$\sum_E \max_i \sum_{j=1}^2 U_{ij} Pr(E|H_j, I) Pr(H_j|I) - \max_i \sum_{j=1}^2 U_{ij} Pr(H_j|I). \quad (13)$$

A comparison of the value obtained from Equation (13) with a given cost of the experiment enables a decision to be made as to whether it is worthwhile to conduct the experiment. If the gain in value is less than the cost of the experiment (acquisition of the information), one would choose not to acquire this information so to maximize the expected net gain, as suggested by Raiffa and Schlaifer (1961). A numerical example is given by Gittelsohn (2013).

It is shown in Appendix 1 that, with the use of the principle of rationality (Good 1967), consideration of new evidence cannot lead to a decrease in utility. Any increase has, of course, to be balanced with the cost of the acquisition of new evidence. Thus, the principle of rationality supports the Principle of Total Evidence (Good 1967) that all the evidence already available should be used, provided that the cost of doing so is negligible.

¹² An argument for expected utility theory relies on so-called representation theorems. Following Zynda [2000] (p. 51), the argument has three premises. Briggs [2019] wrote: ‘The Rationality Condition: The axioms of expected utility theory are the axioms of rational preference. Representability: If a person’s preferences obey the axioms of expected utility theory, then she can be represented as having degrees of belief that obey the laws of the probability calculus [and a utility function such that she prefers acts with higher expected utility]. The Reality Condition: If a person can be represented as having degrees of belief that obey the probability calculus [and a utility function such that she prefers acts with higher expected utility], then the person really has degrees of belief that obey the laws of the probability calculus [and really does prefer acts with higher expected utility].’

3. Expected error and probability assignments

It is shown by Horwich (2016) that ‘expected error in our probability judgements is minimized by the acquisition of new evidence’ (p. 119). It is assumed that the ‘acquisition of new evidence’ is cost free.

The expected error around a probability assignment on a hypothesis (e.g. a prosecution proposition¹³) H based on the background information I only is specified by the quantification of the absolute value of the difference between the assigned probabilistic value of, say, H , $Pr(H|I)$, and of the value for the truth (whose value equals 1) or the falsity (whose value equals 0) of H . The same procedure may be applied to quantify the expected error for probability assignments on H after the acquisition of information about the correspondence (E) or lack of correspondence (\bar{E}) of new evidence $E = \{E, \bar{E}\}$ with a feature of the PoI, for example, a fingerprint.¹⁴ Analogous reasoning may be applied to consideration of an alternative hypothesis (e.g. a defence proposition) \bar{H} .

3.1 Expected error on $Pr(H|I)$

$$\begin{aligned} E_e(H|I) &= \\ &Pr(H|I)\{1 - Pr(H|I)\} + \{1 - Pr(H|I)\}\{Pr(H|I) - 0\} \\ &= Pr(H|I)\{1 - Pr(H|I)\} + \{1 - Pr(H|I)\}Pr(H|I) \\ &= 2Pr(H|I)\{1 - Pr(H|I)\}. \end{aligned} \quad (14)$$

3.2 Expected error on $Pr(H|E, I)$

$$\begin{aligned} E_e(H|(E, \bar{E}), I) &= \\ &Pr(E|I)[Pr(H|E, I)\{1 - Pr(H|E, I)\} + \{1 - Pr(H|E, I)\}\{Pr(H|E, I) - 0\}] + \\ &Pr(\bar{E}|I)[Pr(H|\bar{E}, I)\{1 - Pr(H|\bar{E}, I)\} + \{1 - Pr(H|\bar{E}, I)\}\{Pr(H|\bar{E}, I) - 0\}] \\ &= Pr(E|I)[2Pr(H|E, I)\{1 - Pr(H|E, I)\}] + \\ &Pr(\bar{E}|I)[2Pr(H|\bar{E}, I)\{1 - Pr(H|\bar{E}, I)\}] \\ &= 2Pr(E|I)Pr(H|E, I)\{1 - Pr(H|E, I)\} + \\ &2Pr(\bar{E}|I)Pr(H|\bar{E}, I)\{1 - Pr(H|\bar{E}, I)\}. \end{aligned} \quad (15)$$

It can be proved that $E_e(H|(E, \bar{E}), I) \leq E_e(H|I)$ (Horwich 2016). The proof is provided in Appendix 2. New information is always expected to be of value and it is rational to decide to acquire it. The error (as considered by Horwich 2016) in the hypothesis assessment can never be increased with the addition of new evidence. Table 3 gives some illustrative values for $Pr(E|H, I)$ and $Pr(E|\bar{H}, I)$ for a series of fixed values of $Pr(H|I)$. Equal values for the expected errors are observed when $Pr(E|H, I) = Pr(E|\bar{H}, I)$ [see the last column of Table 3 and Equation (14)] where the value of the evidence is neutral (the Bayes factor equals 1).

3.3 The Taffio Palmi - Busetto case (TPB)

These ideas about expected error can be applied to the TPB case. A second forensic analysis found a correspondence in DNA profiles for the DNA on the necklace found in Busetto’s apartment and the DNA profile of the victim, Taffio Palmi. This was the evidence considered by the court and may be thought of as the background information, I . Other evidence, that was available at no extra cost, was not considered by the court, and it is of interest to have some idea of the size of the reduction in the expected error on the probability of the guilt of Busetto if this other evidence was considered.

The propositions for consideration are

- H : Busetto (B) murdered Taffio Palmi (TP).

¹³ The term ‘hypothesis’ is a generic term to clarify that there are two (or more) points of view from parties at a trial. The term ‘proposition’ is the text that specifies the content of the hypothesis. For example, there is a prosecutor’s hypothesis and its proposition is that the PoI is the source of a stain.

¹⁴ In Section 2, E denoted extra information with an implicit assumption of the presence of evidence. In Section 3, account is taken of the possible absence of evidence.

Table 3. Expected error on $Pr(H|I)$, $E_e(H|I)$ and on $Pr(H|E, I)$, $E_e(H|E, I)$, for changing values of $Pr(H|I)$, $Pr(E|H, I)$ and $Pr(E|\bar{H}, I)$.

$Pr(H I)$	0.1	0.1	0.1	0.5	0.5
$Pr(E H, I)$	0.99000	1.00000	0.50000	0.99000	1.00000
$Pr(E \bar{H}, I)$	0.10000	0.01000	0.80000	0.10000	0.01000
$E_e(H I)$	0.18000	0.18000	0.18000	0.50000	0.50000
$E_e(H E, I)$	0.09628	0.01651	0.17177	0.10072	0.00990
$Pr(H I)$	0.5	0.9	0.9	0.9	0.9
$Pr(E H, I)$	0.50000	0.99000	1.00000	0.50000	0.50000
$Pr(E \bar{H}, I)$	0.80000	0.10000	0.01000	0.80000	0.50000
$E_e(H I)$	0.50000	0.18000	0.18000	0.18000	0.18000
$E_e(H E, I)$	0.45055	0.03614	0.00200	0.17415	0.18000

- \bar{H} : Busetto did not murder Taffio Palmi.

The expected error on $Pr(H|I)$ is given by Equation (14). Other available evidence that was not considered by the court includes the following.

- The fingerprint on the wall of TP 's apartment did not correspond to any of the fingerprints of B . (Denote this E_F .) The complement, \bar{E}_F , is that the fingerprint did correspond to one of the fingerprints of B .
- Shoe marks found at the crime scene did not correspond to the treads of any shoes found associated with B (E_S) The complement, \bar{E}_S , is that the shoe marks did correspond to the treads of a shoe associated with B .
- There was an absence of corresponding DNA on a series of items, such as slippers, carpets, and cleaning rags associated with B (E_D) The complement, \bar{E}_D , is that there was corresponding DNA on the series of items associated with B .

For any E of $\{E_F, E_S, E_D\}$, the expected error on $Pr(H|E, I)$, where E denotes (E, \bar{E}) , is given by Equation (15).

For ease of notation, denote $Pr(H|I)$ as p_0 , and $Pr(E|H, I)$ as p and $Pr(E|\bar{H}, I)$ as q . For determination of the expected error associated with additional evidence E , there are three (possibly subjective) probabilities that need to be assigned. These are the prior probability (p_0) for the prosecution proposition H and then the probabilities (p, q) for evidence E , conditional on H and conditional on \bar{H} . The probability of $(E|I)$ may be written as

$$Pr(E|I) = Pr(E|H, I)Pr(H|I) + Pr(E|\bar{H}, I)Pr(\bar{H}|I) = pp_0 + q(1 - p_0),$$

and then

$$\begin{aligned}
 E_e(H|(E, \bar{E}), I) &= 2Pr(E|I)Pr(H|E, I)\{1 - Pr(H|E, I)\} \\
 &\quad + 2Pr(\bar{E}|I)Pr(H|\bar{E}, I)\{1 - Pr(H|\bar{E}, I)\} \\
 &= 2Pr(E|H, I)Pr(H|I)\left\{1 - \frac{Pr(E|H, I)Pr(H|I)}{Pr(E|I)}\right\} \\
 &\quad + 2Pr(\bar{E}|H, I)Pr(H|I)\left\{1 - \frac{Pr(\bar{E}|H, I)Pr(H|I)}{Pr(\bar{E}|I)}\right\} \\
 &= 2[p p_0 \left\{1 - \frac{p p_0}{p p_0 + q(1 - p_0)}\right\}] \\
 &\quad + 2[(1 - p) p_0 \left\{1 - \frac{(1 - p) p_0}{(1 - p) p_0 + (1 - q)(1 - p_0)}\right\}].
 \end{aligned} \tag{16}$$

Table 4. Expected error $E_e(H|E, I)$, denoted E_e with the consideration of evidence **E**. The prior probability p_0 of $(H|I)$ is chosen to be 0.7, 0.8, and 0.9 and the corresponding prior odds are given in parentheses. The probabilities for the presence (E) of the evidence conditional on H and conditional on the defence proposition \bar{H} are denoted p and q , respectively, with corresponding likelihood ratio q/p ($Pr(E|\bar{H}, I)/Pr(E|H, I)$) in favour of the defence. The corresponding posterior odds are $Pr(H|E, I)/Pr(\bar{H}|E, I)$, denoted O_{pst} , calculated as $O_{pst} = \frac{p_0}{1-p_0} \times \frac{p}{q}$.

$p_0 = 0.7(2.3)$					$p_0 = 0.8(4.0)$					$p_0 = 0.9(9.0)$				
p	q	E_e	$\frac{q}{p}$	O_{pst}	p	q	E_e	$\frac{q}{p}$	O_{pst}	p	q	E_e	$\frac{q}{p}$	O_{pst}
$E_e(H I) = 0.42$ $2p_0(1-p_0)$					$E_e(H I) = 0.32$ $2p_0(1-p_0)$					$E_e(H I) = 0.18$ $2p_0(1-p_0)$				
0.1	0.7	0.26	7.0	0.33	0.1	0.7	0.21	7.0	0.57	0.1	0.7	0.14	7.0	1.29
0.2	0.7	0.32	3.5	0.67	0.2	0.7	0.26	3.5	1.14	0.2	0.7	0.16	3.5	2.57
0.3	0.7	0.36	2.3	1.00	0.3	0.7	0.29	2.3	1.71	0.3	0.7	0.17	2.3	3.86
0.1	0.8	0.22	8.0	0.29	0.1	0.8	0.18	8.0	0.50	0.1	0.8	0.12	8.0	1.13
0.2	0.8	0.29	4.0	0.58	0.2	0.8	0.24	4.0	1.00	0.2	0.8	0.15	4.0	2.25
0.3	0.8	0.33	2.7	0.88	0.3	0.8	0.27	2.7	1.50	0.3	0.8	0.16	2.7	3.38
0.1	0.9	0.17	9.0	0.26	0.1	0.9	0.15	9.0	0.44	0.1	0.9	0.11	9.0	1.00
0.2	0.9	0.24	4.5	0.52	0.2	0.9	0.21	4.5	0.89	0.2	0.9	0.14	4.5	2.00
0.3	0.9	0.29	3.0	0.78	0.3	0.9	0.24	3.0	1.33	0.3	0.9	0.15	3.0	3.00

Some numerical examples are given in Table 4. For reference, the expected error associated with the prior probability $Pr(H|I)$ is $2p_0(1-p_0)$. This has a maximum of 0.5 when p_0 equals 1/2 and there is maximum uncertainty associated with H , and a minimum of 0 when p_0 equals 1 or 0 and there is no uncertainty associated with H ; it is known with certainty either to be true or to be false.

The prior odds $Pr(H|I)/Pr(\bar{H}|I)$ are given in parentheses alongside the three values of $p_0 = Pr(H|I)$ chosen. In addition, the reciprocal, q/p of the likelihood ratio or value of the evidence, $Pr(E|H, I)/Pr(E|\bar{H}, I)$, is given. As each chosen p is less than q , the reciprocal q/p of the evidential value is presented and is to be understood as the support the evidence gives for the defence proposition. Finally, the posterior odds $Pr(H|E, I)/Pr(\bar{H}|E, I)$, denoted O_{pst} , are given. These differ for the three values of $Pr(H|I)$.

The choice of $p < q$ has been made by considering no or little relationship between the corresponding evidence and the PoI. Thus, for each of the three items of evidence E_F , E_S , and E_D , the evidence is more probable under the defence hypothesis (\bar{H}) than the prosecution hypothesis (H).

- The further p_0 is from 0.5, the lower the expected error $E_e(H|E, I)$ is.
- The closer p is to 0.5, the higher the expected error $E_e(H|E, I)$ is.
- The closer q is to 0.5, the higher the expected error $E_e(H|E, I)$ is.

The changes in the posterior odds O_{pst} are as expected for changes in p_0, p and q :

- For fixed p_0 (prior odds) and q , O_{pst} increases as p (and hence p/q) increases.
- For fixed p_0 and p , O_{pst} decreases as q increases (and hence p/q decreases).
- For fixed p and q , O_{pst} increases as p_0 increases.

These aspects can be clearly observed in Fig. 1.

The prior odds $p_0/(1-p_0)$ in the three examples are 2.3, 4.0, and 9.0, respectively. Consideration of evidence with likelihood ratio p/q can lead to a reduction in the expected error and a decrease in the odds in favour of the prosecution hypothesis of up to a factor of 9. As well as consideration of the expected error, the possible change in odds through evaluation of new evidence is another factor to be accounted for in the court's consideration of evidence.

The theory and the numerical examples show that $E_e(H|I) \geq E_e(H|E, I)$. The failure to consider the available evidence did not reduce the error around the probability of the prosecutor's hypothesis that B murdered TP . This is not a coherent way to deal with available evidence.

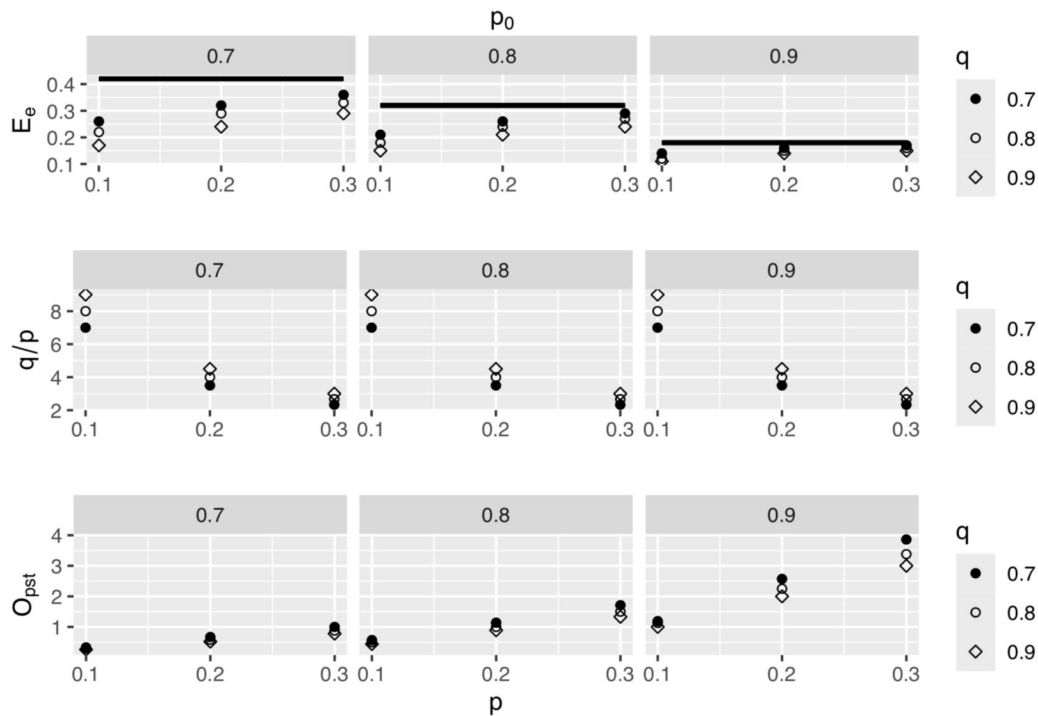


Figure 1. Diagrammatic illustration of examples from Table 4 of the relationships among the expected error on the posterior probability, the likelihood ratio, and the posterior odds, with changes in the prior probability p_0 and the probability of the evidence conditional on the defence proposition, q . The prior probability p_0 of $(H|I)$ is chosen to be 0.7, 0.8, and 0.9, with corresponding prior odds equal to 2.3, 4.0, and 9.0. The probabilities for the presence (E) of the evidence conditional on the prosecution proposition H and conditional on the defence proposition \bar{H} are denoted p (with illustrative values 0.1, 0.2, 0.3) and q (with illustrative values 0.7, 0.8, 0.9) and are indicated by the symbols \bullet , \circ , \diamond . Top: Expected error $E_e(H|E, I)$ (Eq. 15), denoted E_e with the consideration of evidence E . The solid line indicates the Expected error on $Pr(H|I)$, $E_e(H|I) = 2p_0(1 - p_0)$ that equals 0.42, 0.32 and 0.18, respectively. Middle: Likelihood ratio q/p in favour of the defence. Bottom: Posterior odds O_{pst} in favour of the prosecution proposition, $Pr(H|E, I)/Pr(\bar{H}|E, I)$, calculated as $O_{pst} = \frac{p_0}{1-p_0} \times \frac{p}{q}$.

4. Expected value and weight of evidence

It was noticed by Good (1950), reiterated in Good (1979) and Good (1985), that A.M. Turing remarked that the expectation of the Bayes factor against a true hypothesis is equal to 1 and the expected weight of the evidence quantified through the logarithm of the Bayes factor in favour of a true hypothesis is non-negative (Good 1979: 395). These results have been expressed in the following terms (Good 1950):

The expected factor for a wrong hypothesis in virtue of any experiment is equal to 1. [...] the expected weight of evidence for right hypotheses is positive and for wrong hypotheses is negative. (p. 72)

These two results refer to expectations, the first for the value and the second for the weight of evidence. These are considerations for investigators when thinking whether or not to consider a certain type of evidence. They are not applicable once the evidence has been considered and results have been obtained.

The value of evidence has been defined to be the Bayes factor. This definition is not just abstract mathematical terminology. The Bayes factor is the multiplicative factor that converts the odds in favour of a hypothesis, such as a prosecution hypothesis, before evidence is assessed, to the odds in favour of the hypothesis after the evidence is assessed. A factor that has such a function has a good claim to be called the ‘value’ of the evidence.

The weight of the evidence has been defined to be the logarithm of the Bayes factor. Again, this definition is not just abstract terminology. The logarithmic function converts the multiplicative expression of the relationship of the prior odds, the Bayes factor, and the posterior odds into an additive expression. The logarithm of the posterior odds is the sum of the logarithm of the Bayes factor and the logarithm of the prior odds. This is a very good analogy with the scales of justice. The logarithm of the numerator of the Bayes factor, say $\log \{Pr(E|H, I)\}$, is added to the scale of the prosecution containing the logarithm of the prior probability $\log \{Pr(H|I)\}$ to give the logarithm of the posterior probability $\log \{Pr(H|E, I)\}$. The logarithm of the denominator of the Bayes factor, say $\log \{Pr(E|\bar{H}, I)\}$, is added to the scale of the defence containing the logarithm of the prior probability $\log \{Pr(\bar{H}|I)\}$ to give the logarithm of the posterior probability $\log \{Pr(\bar{H}|E, I)\}$. Again a term that has such a function has a good claim to be called the ‘weight’ of the evidence.

The combined weight of a set of independent items of evidence is the sum of the individual weights of evidence. For example, let E_1 and E_2 be two independent items of evidence. Then

$$\begin{aligned} \log \left(\frac{Pr(E_1 E_2 | H)}{Pr(E_1 E_2 | \bar{H})} \right) &= \log \left(\frac{Pr(E_1 | H) Pr(E_2 | H)}{Pr(E_1 | \bar{H}) Pr(E_2 | \bar{H})} \right) \\ &= \log \left\{ \left(\frac{Pr(E_1 | H)}{Pr(E_1 | \bar{H})} \right) \left(\frac{Pr(E_2 | H)}{Pr(E_2 | \bar{H})} \right) \right\} \\ &= \log \left(\frac{Pr(E_1 | H)}{Pr(E_1 | \bar{H})} \right) + \log \left(\frac{Pr(E_2 | H)}{Pr(E_2 | \bar{H})} \right) \end{aligned}$$

where I has been omitted for ease of notation.

A similar, but more complex, argument can be made for the combined weight of dependent items of evidence. Thus, consideration of additional evidence will lead to a change in the expected weight of evidence. This consideration also supports the Principle in that failure to include all available evidence means a failure to achieve the best value for the expected weight of evidence.

4.1 Expected value of evidence

Good demonstrated these results as follows. First, it is shown that the expected value of the evidence for a wrong proposition is 1.

The proof of the result that the expected value of the evidence for a wrong proposition is 1 has two parts, one when H is true and one when \bar{H} is true. When H is true, the wrong proposition is \bar{H} . Given background information I , the value of evidence E_r , the observed outcome of evidence E , with $r = 1, \dots, n$ possible outcomes, for example, a type of tyre or a type of shoe, in support of \bar{H} is $Pr(E_r | \bar{H}, I) / Pr(E_r | H, I)$, with probability $Pr(E_r | H)$. Thus, the expected value of evidence E , before it is observed and noted to have a particular outcome E_r , in support of \bar{H} , when H is true, is

$$\sum_{r=1}^n Pr(E_r | H, I) \frac{Pr(E_r | \bar{H}, I)}{Pr(E_r | H, I)} = \sum_{r=1}^n Pr(E_r | \bar{H}, I), \quad (17)$$

where $\sum_r Pr(E_r | \bar{H}, I) = Pr(E_1 \cup E_2 \cup \dots \cup E_n | \bar{H}, I) = 1$.

Similarly, the expected value of evidence E , before it is observed and noted to have a particular outcome E_r , in support of H , when \bar{H} is true, is

$$\sum_{r=1}^n Pr(E_r | \bar{H}, I) \frac{Pr(E_r | H, I)}{Pr(E_r | \bar{H}, I)} = \sum_{r=1}^n Pr(E_r | H, I) = 1. \quad (18)$$

This result has an implication for the consideration of the value of the evidence. The value of a particular outcome E_r of an evidential type E is the Bayes factor. By definition the value in support of \bar{H} is $Pr(E_r | \bar{H}, I) / Pr(E_r | H, I)$. If H is the true proposition, the evidence should support H ,

the posterior odds in favour of H should be greater than the prior odds and so the Bayes factor $Pr(E_r|\bar{H}, I)/Pr(E_r|H, I)$ should be less than 1. However, it has been shown that the expected value of this term is equal to 1. Thus, for some possible outcomes E_r in $\{E_1, \dots, E_n\}$, this Bayes factor will be greater than one and the posterior odds in favour of \bar{H} will be greater than the prior odds in favour of \bar{H} which is not a desirable result. The result that the expected value of the evidence for a wrong proposition is 1 means that there will be occasions when the outcome is seen to support the wrong proposition. Good (1985) notes that '[t]he only way to get an expected value of 1 is if the distribution of the Bayes factor is skewed to the right, that is, when the factor against the truth exceeds 1 it can be large' (p. 255). Thus, there has to be careful consideration of evidence, awareness of this possibility, and acceptance of the support given to the proposition by the evidence should not be given without careful thought. Good (1985) admits that '[i]t is disturbing that one can get a large factor against the truth' (p. 255). There is further discussion of this result later in Good (1985) which is beyond the scope of this article.

As an illustration of the result in Equation (17), consider the following forensic scientific example. A biological stain is recovered on a crime scene. A person of interest, the PoI, is considered as the donor of the stain. A laboratory will perform DNA analyses on the stain and on the PoI's reference sample. The evidence presented in court is either a report of a correspondence (E_1) between the two genetic profiles, or a report of a non-correspondence (E_2) between them. There are only two possible results so $n = 2$ and, for ease of notation, denote E_1 as E and E_2 as \bar{E} .

The Court is interested in the evaluation of the findings under two hypotheses at the source level: H , the PoI is the source of the stain, and \bar{H} , the PoI is not the source of the stain (an unrelated person is the source of that stain). The value of the finding is expressed through a Bayes factor.

The Bayes factor is either:

$$\frac{Pr(E|H, I)}{Pr(E|\bar{H}, I)} \text{ or } \frac{Pr(\bar{E}|H, I)}{Pr(\bar{E}|\bar{H}, I)}. \quad (19)$$

- The terms in the numerators, $Pr(E|H, I)$ and $Pr(\bar{E}|H, I)$, are the probabilities that the analyst will report a correspondence or non-correspondence, respectively, if the PoI is the source of the crime scene stain. These are the two numerators of the possible Bayes factors in Equation (19). The probability of reporting a correspondence if the PoI is the source of the crime stain is the product of the probability there is a correspondence, which is 1, and the probability that it is reported. Denote the probability of reporting a correspondence $Pr(E|H, I)$ by α and its complement, the probability of failing to report a correspondence when there is one, as $(1 - \alpha)$, respectively.
- The terms in the denominator, $Pr(E|\bar{H}, I)$ and $Pr(\bar{E}|\bar{H}, I)$, are the probabilities that the analyst will report a correspondence or non-correspondence, respectively, if the PoI is not the source of the crime scene stain. The probability that a correspondence is reported given that the PoI is not the source of the crime stain is the probability of a random correspondence and a correct report of that random correspondence or the probability of a non-correspondence which is incorrectly reported as a correspondence. Denote the probability of a random correspondence as γ and the probability of a report of a non-correspondence as a correspondence as $(1 - \beta)$. The probability of the report of a non-correspondence as a non-correspondence is then β . Thus, the probability of a report of a correspondence (E) when the PoI is not the source of the crime stain is $\alpha\gamma + (1 - \beta)(1 - \gamma)$. The probability of a report of a non-correspondence (\bar{E}) when the PoI is not the source of the crime stain is then $(1 - \alpha)\gamma + \beta(1 - \gamma)$.
- If the evidence is the report of a correspondence between the DNA profiles of the PoI and the crime stain (E), the Bayes factor is

$$\frac{Pr(E|H, I)}{Pr(E|\bar{H}, I)} = \frac{\alpha}{\alpha\gamma + (1 - \beta)(1 - \gamma)}.$$

- If the evidence is the report of no correspondence between the DNA profiles of the PoI and the crime stain (\bar{E}), the Bayes factor is

$$\frac{Pr(\bar{E}|H, I)}{Pr(\bar{E}|\bar{H}, I)} = \frac{1 - \alpha}{(1 - \alpha)\gamma + \beta(1 - \gamma)}.$$

Assume, for example, that $\alpha = 0.99$, $\gamma = 0.000001$ and $\beta = 0.999$, then

$$\begin{aligned} \frac{Pr(E|H, I)}{Pr(E|\bar{H}, I)} &= \frac{\alpha}{\alpha\gamma + (1 - \beta)(1 - \gamma)} \\ &= \frac{0.99}{(0.99 \times 0.000001) + (0.001 \times 0.999999)} \\ &= 989.0219. \end{aligned}$$

and

$$\begin{aligned} \frac{Pr(\bar{E}|H, I)}{Pr(\bar{E}|\bar{H}, I)} &= \frac{1 - \alpha}{(1 - \alpha)\gamma + \beta(1 - \gamma)} \\ &= \frac{0.01}{(0.01 \times 0.000001) + (0.999 \times 0.999999)} \\ &= 0.01001002. \end{aligned}$$

Consider that hypothesis H (the PoI is the donor of the stain) is true. The probability of a reported correspondence is $Pr(E|H, I) = \alpha = 0.99$ with a corresponding Bayes factor of approximately 990. The probability of a reported non-correspondence is $Pr(\bar{E}|H, I) = (1 - \alpha) = 0.01$ with a corresponding Bayes factor of approximately 0.010.

Therefore, the expected Bayes factor (before evidence is obtained¹⁵) in favour of \bar{H} when the PoI is the donor of the stain (H true) is

$$\begin{aligned} Pr(E|H, I) \frac{Pr(E|\bar{H}, I)}{Pr(E|H, I)} + Pr(\bar{E}|H, I) \frac{Pr(\bar{E}|\bar{H}, I)}{Pr(\bar{E}|H, I)} &= \\ = \alpha \left\{ \frac{\alpha\gamma + (1 - \beta)(1 - \gamma)}{\alpha} \right\} + (1 - \alpha) \left\{ \frac{(1 - \alpha)\gamma + \beta(1 - \gamma)}{1 - \alpha} \right\} &= 0.99 \times \frac{1}{989} + 0.01 \times \frac{1}{0.010} \\ = 1.001. & \end{aligned} \quad (20)$$

The final result is not exactly 1 because of rounding errors. Use of a computer with greater precision gives the exact result of 1. Study of Equation (20) shows equivalence with Equation (17).

4.2 Expected weight of evidence

Good (1950) also showed that the expected weight of evidence for the correct proposition is non-negative and for the wrong proposition is non-positive. The proof depends on the mathematical result that the geometric mean of a set of numbers is less than or equal to the arithmetic mean, with equality only if the set of numbers are all equal. First, some notation is defined.

Consider two mutually exclusive hypotheses H and \bar{H} and a type of evidence E with mutually exclusive outcomes E_1, E_2, \dots, E_n . For a particular outcome, $E_r, r = 1, \dots, n$, the following notation is introduced for ease of presentation.

- $p_r = Pr(E_r|\bar{H})$,
- $f_r = Pr(E_r|H)/Pr(E_r|\bar{H})$,
- $p_r f_r = Pr(E_r|H)$,

and I has been omitted for ease of notation.

¹⁵ Of course, before the evidence is obtained, the value of γ is not known. The value given here is for illustration; the result holds for any value of γ .

It is assumed that $p_r > 0$ and $f_r > 0$ for all $r = \{1, 2, \dots, n\}$. Also $\sum p_r = 1$ and $\sum p_r f_r = 1$. The weight of evidence $E_r, r = 1, \dots, n$, in support of proposition H is

$$\log \{Pr(E_r|H)/Pr(E_r|\bar{H})\}.$$

If H is false (i.e. \bar{H} is true), the expected weight of evidence of type E in support of H is

$$\sum_{r=1}^n Pr(E_r|\bar{H}) \log \left(\frac{Pr(E_r|H)}{Pr(E_r|\bar{H})} \right) = \sum p_r \log f_r = \sum \log (f_r^{p_r}) = \log \prod f_r^{p_r}.$$

$\prod f_r^{p_r}$ is the geometric mean of f_r with weights p_r (remember $\sum p_r = 1$). Thus, from the result that the geometric mean is less than or equal to the arithmetic mean,

$$\prod f_r^{p_r} \leq \frac{\sum p_r f_r}{\sum p_r} = 1.$$

Thus:

$$\sum_{r=1}^n Pr(E_r|\bar{H}) \log \left(\frac{Pr(E_r|H)}{Pr(E_r|\bar{H})} \right) = \log \prod f_r^{p_r} \leq \log 1 = 0. \quad (21)$$

The result for the expected weight of evidence of type E in support of H if H is true can be shown analogously to be greater than or equal to zero.¹⁶ Remember $p_r f_r = Pr(E_r|H)$. For a slight easing of notation, denote $p_r f_r$ as q_r with $\sum q_r = 1$. Then

$$\sum_{r=1}^n Pr(E_r|H) \log \left(\frac{Pr(E_r|\bar{H})}{Pr(E_r|H)} \right) = \sum q_r \log f_r^{-1} \leq 0,$$

from Equation (21). Now:

$$\sum q_r \log f_r^{-1} \leq 0 \Rightarrow - \sum q_r \log f_r \leq 0 \Rightarrow \sum q_r \log f_r \geq 0,$$

and, hence,

$$\sum_{r=1}^n Pr(E_r|H) \log \left(\frac{Pr(E_r|\bar{H})}{Pr(E_r|H)} \right) \geq 0. \quad (22)$$

As with the result for the value of the evidence, this result has an implication for the consideration of the weight of the evidence. The weight of a particular outcome E_r of an evidential type E is the logarithm of the Bayes factor. By definition the weight in support of \bar{H} is $\log \{Pr(E_r|\bar{H}, I)/Pr(E_r|H, I)\}$. As H is the true proposition, the evidence should support H , the posterior odds in favour of H should be greater than the prior odds and so the logarithm of the Bayes factor, $\log \{Pr(E_r|\bar{H}, I)/Pr(E_r|H, I)\}$, should be less than 0. It has been shown that the expected value of this term is less than or equal to 0. This may be thought a more pleasing result than for the value of the evidence as it allows for the possibility that the weight will never be greater than 0. However, the previous result shows that there will be a value greater than 1 and hence a weight greater than 0. As before, there has to be careful consideration of evidence,

¹⁶ Note that the expected value of the evidence, the mean of a ratio, being equal to 1, does not mean that mean of the logarithm of the ratio is equal to zero. The logarithm of a mean of a ratio that is equal to 1 would be 0. The logarithm of a mean is not the same as the mean of a logarithm.

Table 5. Probabilities of not finding a corresponding fingermark (E) if B is guilty (H) and if B is not guilty (\bar{H}), the respective likelihood ratios and the expected weight EW (in natural logarithms) of the evidence E from (23).

$Pr(E H)$	$Pr(E \bar{H})$	$\frac{Pr(E H)}{Pr(E \bar{H})}$	$\frac{Pr(\bar{E} H)}{Pr(\bar{E} \bar{H})}$	EW
0.01	0.98	0.010	49.5	3.82
0.02	0.99	0.020	98.0	4.42
0.05	0.98	0.051	47.5	3.52
0.10	0.95	0.105	18.0	2.38
0.10	0.90	0.111	9.0	1.76
0.001	0.995	0.001	199.8	5.29

awareness of this possibility, and not unthinking acceptance of the support given to the proposition by the evidence.

Consider the TPB case. Proposition H is that the PoI (B) is guilty and \bar{H} is that B is not guilty. Consider the evidence that the fingermark on the wall of Taffio Palmi's (TP 's) apartment did not correspond to any of B 's fingerprints, denote this E .¹⁷ Thus, E is supportive of proposition \bar{H} . The complement of E , \bar{E} , is evidence that the fingermark on the wall did correspond to one of B 's fingerprints and is supportive of H . In the notation of the general result, $n = 2$, E_1 corresponds to E and E_2 corresponds to \bar{E} . The expected weight of evidence for the correct proposition is, with Equation (22) written with $n = 2$, E replacing E_1 and \bar{E} replacing E_2 ,

$$Pr(E|H) \log \left(\frac{Pr(E|H)}{Pr(E|\bar{H})} \right) + Pr(\bar{E}|H) \log \left(\frac{Pr(\bar{E}|H)}{Pr(\bar{E}|\bar{H})} \right). \quad (23)$$

The value of evidence E , that the fingermark on the wall of TP 's apartment did not correspond to any of B 's fingerprints, in support of H is $Pr(E|H)/Pr(E|\bar{H})$. The value of evidence \bar{E} , that the fingermark on the wall of TP 's apartment did correspond to one of B 's fingerprints, in support of H is $Pr(\bar{E}|H)/Pr(\bar{E}|\bar{H})$. Investigators of the crime have to elicit values for the probabilities of not finding a corresponding fingermark if B is guilty and if B is not guilty ($Pr(E|H)$ and $Pr(E|\bar{H})$, respectively). Table 5 gives values for the respective likelihood ratios of not finding or finding a corresponding fingermark and the associated expected weight of evidence. In the TPB case, it is reasonable to assume that the absence of a corresponding fingermark to any of B 's fingerprints is suggestive that the true hypothesis is that B is innocent.

All these results support the theoretical result that the expected weight of the evidence in support of the hypothesis that B is not guilty is positive.

The likelihood ratio for the evaluation of evidence has an intuitively attractive meaning in that it is the factor that converts the odds in favour of a proposition before the evidence is considered to posterior odds in favour of the proposition after the evidence is considered. The result that the expected weight of evidence in favour of a true hypothesis is non-negative is confirmation that the statistic, logarithm of the likelihood ratio, is an intuitively attractive definition for the weight of evidence. The *expected* weight of a particular type of evidence is a function which is considered before a search is made for the evidence. In the TPB case where a fingermark was observed on the wall of the victim's apartment, consideration can be given to the expected weight of the evidence of the absence or presence of a corresponding fingerprint. It illustrates the relevance of the fingermark to the investigation through the requirement to consider the probabilities of finding or not finding a fingermark at the crime scene that corresponds to a fingerprint from Busetto, if she is innocent or if she is guilty. It is not suggested that precise numerical values are given to these probabilities. Verbal equivalents could be used: for example, it is (very) unlikely that a corresponding fingermark would be found if B were innocent and (very) likely that a corresponding fingermark would be found if B were guilty.

¹⁷ Here, E refers to a non-correspondence. In the discussion of the expected value of evidence, Section 4.1, E refers to the *report* of a correspondence. This difference in definition has no influence on the results presented.

5. Discussion and conclusion

Satisfaction of Carnap's Principle of Total Evidence (Carnap 1947) enables jurists and scientists to reach logical and justifiable conclusions. The Principle is a fundamental prerequisite for any valid methodology for inferences and decisions because it justifies the acquisition and use of all available items of evidence. However, lawyers may argue that mathematics and numerical assignments (quantification) have no place in adjudication.

Appendix 1 shows that consideration of new evidence cannot lead to a decrease in utility. Appendix 2 shows that consideration of evidence reduces the expected error. In Section 2, it is commented that '[p]rovided that the utilities ... of the outcomes of the possible actions ... can be quantified ... Bayesian decision theory explains how the expected value of information may be calculated'. Some may argue that since outcomes of actions cannot be quantified, the discussion of this section of the article is rendered irrelevant. However, the proof that consideration of new evidence cannot lead to a decrease in utility does not depend on quantification. Similarly, the proof that consideration of evidence reduces expected error does not rely on quantification. For those who are wary of mathematics and quantification, these proofs may be considered as representations of relative likelihoods; the outcome of a particular action is more or less likely than the outcome of another action.

The use of symbols in mathematical arguments may give the impression that the argument is abstract and divorced from reality. However, in many applications, these symbols have verbal interpretations. In the context of the administration of justice, these verbal interpretations concern evidence, propositions and assignments of uncertainty. The symbols, equations, and algebraic manipulations are all shorthand for verbal descriptions. The use of the symbols enables lengthy verbal descriptions to be made concisely and unambiguously.

Whilst consideration of the Principle of Total Evidence is important, attention has to be paid to the cost of the acquisition of new evidence. It is not practical to collect evidence indefinitely, not least because the cost of collection would increase indefinitely. For example, in a case of comparative genetic analysis between the profile of a biological trace found on a crime scene and that of a person of interest, the number of DNA markers considered cannot increase indefinitely. Consider, for sake of illustration, a scenario involving results from 16 DNA markers that show a correspondence between the profiles. The defence could argue that additional markers should be studied as they could show a potential non-correspondence with their client and hence exonerate them as the donor of the stain.

If—as mentioned in Section 1—information was costless and there were no deadlines for the taking of decisions then the argument is compelling. However, these conditions do not hold in the context of a criminal investigation and trial. Cost plays an important role. The cost of the new information will allow the calculation of the net gain.

The gain in utility given by Equation (13) has to be compared with the cost of an experiment or, more generally, of the acquisition of new information, before a decision is made as to whether it is worthwhile to acquire the information.

Table 4 and Fig. 1 show the expected error associated with evidence compared with the expected error before the evidence was assessed. The accompanying text gives an interpretation of the results to show that it is not the exact values that are of particular importance; rather, it is the relative values. Table 4 and Fig. 1 also show the posterior odds for various values of the prior odds in favour of the prosecution proposition and for various probabilities associated with the acquisition of the evidence for each of the propositions of involvement (or not) of the person of interest. Again, it is not the exact values of these probabilities that are important, it is their relative value. The three items of evidence (fingermarks, E_B , shoe marks, E_S and DNA, E_D) that were not considered by the Court in the TPB case, had been acquired. There was no additional cost for their acquisition. The costs to be accounted for by the Court when compared to the utility of their consideration are those of the court's time.

More generally, there is much to think about in the consideration of expected errors, utilities and losses. It is easy to give a theoretical result that shows an increase in utility through the acquisition of additional information. It is not so easy to determine how to implement the result. First, consider the expected error associated with evidence Equation (16). Three (possibly subjective) probabilities are required, the prior probability for the prosecution proposition, and the

two probabilities for the evidence, one considered conditional on the prosecution proposition, one considered conditional on the defence proposition.

The second consideration, that of the expected value of partial or perfect information and the associated values of utilities and losses, is more difficult. This is particularly the case during an investigation. It is easier during a trial when it may be thought all the evidence has been acquired and assessed by the investigators. The cost of its consideration by the court is small compared with the utility of increasing the probability of a correct verdict.

The prior probabilities and evidential probabilities are needed as before. It may be thought controversial for an investigator to consider prior probabilities. However, the investigator's opinion on these will not be presented in court. It will be useful during the assessment of the value of the acquisition of additional evidence. The advantage of considering subjective values for the evidential probabilities and hence likelihood ratios has been described by [Jackson, Aitken, and Roberts \(2014\)](#).

Various factors have to be considered in the choice of utilities. The cost of the search for additional evidence, and for its assessment if found, may be easy to determine relative to the intangible cost of a miscarriage of justice. Also, the cost of losing, for example, can have different values for different protagonists in a civil trial. In a criminal trial, there is the age-old conundrum that tries to determine the number of guilty people acquitted that equate to the one innocent person convicted. Such miscarriages of justice are impossible to quantify. Perhaps the best that can be attempted is to consider probabilities and utilities in qualitative terms. Consider the effect of very small or very large probabilities and very large losses or utilities on the expected utility of partial or perfect information. Such considerations may provide helpful guidance as to whether to gather more evidence or not. Of course, in these days of sophisticated software packages it may be possible for one to be developed that will provide guidance given inputs from the investigators at very little expenditure of time or money.

Not all available items of evidence were considered in the TPB case. Thus, even though quantification may not be possible, it can be argued that utility was not maximized and expected error was not minimized. There is a good argument to be made that those responsible for the administration of criminal justice should require courts to pay attention to Carnap's Principle. Whilst this may seem an esoteric requirement, it is, in reality, no more than a requirement to consider all available evidence, subject to cost, in order to maximize utility, minimize expected error, and hence improve the quality of justice.

Acknowledgements

The authors thank the Swiss National Science Foundation for its support through grant number 100011_204554/1 (The Anatomy of Forensic Inference and Decision). Additionally, we also acknowledge The Consortium of Swiss Academic Libraries for supporting open access to this article.

Conflict of interest statement. None declared.

References

- AITKEN, C., TARONI, F., and BOZZA, S. (2021) *Statistics and the Evaluation of Evidence for Forensic Scientists*, 3rd edn. Chichester: John Wiley & Sons.
- AYER, A. (1957) 'The Conception of Probability as a Logical Relation', in S. Körner (ed.) *Observation and interpretation*, pp. 12–30. London: Butterworths.
- BARRETT, M., and SOBER, E. (2020) 'The Requirement of Total Evidence: A Reply to Epstein's', *Philosophy of science*, **87**: 191–203.
- BIEDERMANN, A., BOZZA, S., and TARONI, F. (2008) 'Decision-Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications', *Forensic Science International*, **177**: 120–132.
- BIEDERMANN, A., BOZZA, S., and TARONI, F. (2016) 'The Decisionalization of Individualization', *Forensic Science International*, **266**: 29–38.
- BIEDERMANN, A. et al. (2020) 'Computational Normative Decision Support Structures of Forensic Interpretation in the Legal Process', *Scripted*, **17**: 83–124.
- BIGGS, R. A. (2019) 'Normative Theories of Rational Choice: Expected Utility', in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Stanford University: Metaphysics Research Lab.

- CARNAP, R. (1947) 'On the Application of Inductive Logic', *Philosophy and Phenomenological Research*, 8: 133–148.
- CARNAP, R. (1950) *Logical Foundations of Probability*, Chicago: University of Chicago Press
- CHATER, N., and OAKSFORD, M. (2012) 'Normative Systems: Logic, Probability, and Rational Choices', in K. J. Holyoak and R. G. Morrison (eds) *The Oxford handbook of thinking and reasoning*, pp. 11–21. Oxford: Oxford University Press.
- COOK, R. et al. (1998) 'A Model for Case Assessment and Interpretation', *Science & Justice*, 38: 151–6.
- COOK, R. et al. (1999) 'Case Pre-Assessment and Review in a Two-Way Transfer Case', *Science & Justice*, 39: 103–11.
- DE MARCH, I., and TARONI, F. (2020) 'Bayesian Networks and Dissonant Items of Evidence: A Case Study', *Forensic Science International: Genetics*, 44: 102172.
- FISCHHOFF, B., and BEYTH-MAROM, R. (1983) 'Hypothesis Evaluation from a Bayesian Perspective', *Psychological Review*, 90: 239–60.
- GENNARI, G. (2021) 'Errore giudiziario e prova scientifica', in L. Luparia (ed.) *L'errore giudiziario*, pp. 227–58. Giuffrè Francis Lefebvre.
- GITTELSON, S. (2013) 'Evolving from Inferences to Decisions in Forensic Science', PhD thesis, The University of Lausanne—School of Criminal Justice, Lausanne.
- GITTELSON, S. et al. (2013) 'Decision-Theoretic Reflections on Processing a Fingerprint', *Forensic Science International*, 226: e42–e47.
- GOOD, I. J. (1950) *Probability and the Weighing of Evidence*. Charles Griffin and Company Limited.
- GOOD, I. J. (1967) 'On the Principle of Total Evidence', *The British Journal for the Philosophy of Science*, 17: 319–321.
- GOOD, I. J. (1979) 'Studies in the History of Probability and Statistics. XXXVII A.M. Turing's Statistical Work in World War II', *Biometrika*, 66: 393–96.
- GOOD, I. J. (1985) 'Weight of Evidence: A Brief Survey (with Discussion)', in J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds) *Bayesian Statistics*, 2, pp. 249–70. Amsterdam: North Holland.
- HAYS, W., and WINKLER R. (1970) *Statistics: Probability, Inference and Decision—Volume I*. Holt, Rinehart and Winston, Inc., New York.
- HORWICH, P. (2016) *Probability and Evidence*. Cambridge: Cambridge University Press.
- JACKSON, G., AITKEN, C. G. G., and ROBERTS, P. (2014) 'Case assessment and interpretation of expert evidence'. Technical report, The Royal Statistical Society. <https://rss.org.uk/RSS/media/File-library/Publications/rss-case-assessment-interpretation-expert-evidence.pdf>.
- JEFFREY, R. (1987) 'Risk and Human Rationality', *The Monist*, 70: 223–236.
- KEYNES, J. (1921) *A Treatise on Probability*. London: MacMillan and Co., Limited.
- LINDLEY, D. V. (1985) *Making Decisions*, 2nd edn. Chichester: John Wiley & Sons.
- LINDLEY, D. V. (1991) 'Probability', in C. G. Aitken and D. A. Stoney (eds) *The Use of Statistics in Forensic Science*, pp. 27–50. New York: Ellis Horwood.
- LOCKE, J. (1689) 'Essay concerning human understanding—Book ii: Ideas'. Chapter xxi: Power. Section 67. <https://www.earlymoderntexts.com/assets/pdfs/locke1690book2.pdf>.
- RAIFFA, H. and SCHLAIFER, R. (1961) *Applied Statistical Decision Theory*. Cambridge, Massachusetts: The M.I.T. Press.
- SALMON, W. (1992) '40 anni di spiegazione scientifica: scienza e filosofia 1948-1987', Franco Muzzio Editore, Padova, Italian edition of Four decades of scientific explanation (1989) edition.
- SALMON, W. C. (1966) *The Foundations of Scientific Inference*. Pittsburgh, PA: University of Pittsburgh Press.
- STONE, D. (1991) 'Transfer Evidence', in C. Aitken and D. Stoney (eds) *The Use of Statistics in Forensic Science*, pp. 107–38. New York: Ellis Horwood.
- STONE, D. A. (1994) '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.
- TARONI F. et al. (2010) *Data Analysis in Forensic Science. A Bayesian Decision Perspective*. Chichester: John Wiley & Sons.
- TARONI, F., BOZZA, S., and GARBOLINO, P. (2018a) 'Contaminazioni di un reperto con il DNA. quando la prova genetica porta direttamente alla condanna', *Diritto Penale Contemporaneo*, 2: 1–14.
- TARONI, F. et al. (2018b) 'Prova genetica del DNA e risultati dissonanti: come valutare congiuntamente gli elementi scientifici di prova', *Diritto Penale Contemporaneo*, 11: 77–94.
- TARONI, F., BOZZA, S., and BIEDERMANN, A. (2020) 'Decision Theory', in D. Banks, K. Kafadar, D. Kaye, and M. Tackett (eds) *Handbook of Forensic Statistics*, pp. 103–130. Boca Raton: Chapman & Hall/CRC.
- WINKLER, R. (2003) *Bayesian Inference and Decision*, 2nd edn. Gainesville: Probabilistic Publishing.
- ZYND, L. (2000) 'Representation Theorems and Realism about Degrees of Belief', *Philosophy of Science*, 67: 45–69.

Appendix

Appendix 1: Proof that consideration of new evidence cannot lead to a decrease in utility

The proof that follows shows that it is worthwhile to take account of new evidence in that such account cannot lead to a decrease in utility. Any increase has, of course, to be balanced with the cost of acquiring the new evidence. It is an expansion of the proof given in [Good \(1967\)](#).

Suppose there are r mutually exclusive and exhaustive hypotheses, H_1, H_2, \dots, H_r and a choice of s acts, or classes of acts, A_1, A_2, \dots, A_s . Let the utility of act A_i if H_j is true be $U(A_i|H_j) = u_{ij}$. Suppose that, on some evidence, E , there are initial probabilities $p_j = \Pr(H_j|E)$. If just E is taken into account, then the (expected) utility of act A_i is $\sum_j p_j u_{ij}$ and the principle of rationality recommends the choice of $i = i_0$, the value of i that maximizes this expression. Therefore, the (expected) utility in the rational use of E is

$$\max_i \sum_j p_j u_{ij}.$$

Thus, in summary, the notation is as follows:

- i : identifier for actions $A_i : A_1, \dots, A_s$.
- j : identifier for propositions $H_j : H_1, \dots, H_r$.
- u_{ij} : utility $U(A_i|H_j)$ for action A_i when H_j true.
- The prior probability of H_j with initial evidence E : p_j .
- The expected utility of action A_i is $\sum_{j=1}^r p_j u_{ij} = \sum_{j=1}^r U(A_i|H_j) \Pr(H_j|E)$.

For an application, choices need to be made for the values of the prior probabilities and the utilities.

The rational choice of action is the action A_i which maximizes the expected utility, suppose that action is A_{i_0} . The maximum expected utility is $\max_i \sum_j p_j u_{ij}$. Thus, it can be written that the expected utility of A_{i_0} is

$$\max_i \sum_j p_j u_{ij} = \sum_j p_j u_{i_0 j}. \quad (24)$$

The questions then are: *Is it worthwhile taking a new observation, taking account of new evidence? Is there an increase in utility if new evidence is considered?*

The new evidential type is assumed to have one of t possible outcomes, denoted E_1, \dots, E_t , indexed by $k = \{1, \dots, t\}$ and assumed mutually exclusive and exhaustive and independent of E ; a subscript is associated with the new evidence E_k to differentiate it from the prior evidence E on which the prior probability $\Pr(H_j|E)$ is assigned. The probabilities associated with the E_k , $\Pr(E_k|H_j)$ depend on the proposition H_j . Denote these probabilities p_{jk} . Since the E_k are mutually exclusive and exhaustive,

$$\sum_{k=1}^t \Pr(E_k|H_j) = \Pr(E_1|H_j) + \dots + \Pr(E_t|H_j) = \sum_{k=1}^t p_{jk} = 1. \quad (25)$$

The posterior probability of $(H_j|E, E_k)$ is $Pr(H_j|E, E_k)$. Denote this as q_{jk} . Then

$$\begin{aligned}
 q_{jk} &= Pr(H_j|E, E_k) \quad (j = 1, \dots, r; k = 1, \dots, t) \\
 &= Pr(E_k|H_j, E)Pr(H_j|E)/Pr(E_k|E) \\
 &= Pr(E_k|H_j)Pr(H_j|E)/Pr(E_k) \\
 &\quad \text{since } E \text{ and } E_k \text{ independent} \\
 &= Pr(E_k|H_j)Pr(H_j|E)/\sum_j Pr(E_k|H_j)Pr(H_j|E) \\
 &= p_j p_{jk} / \sum_j p_j p_{jk},
 \end{aligned}$$

the final (posterior) probability of H_j if E_k occurs (and given E has occurred).

The expected utility of A_i given E_k , and combined with E , is

$$\sum_{j=1}^r q_{jk} u_{ij} = \sum_j U(A_i|H_j) Pr(H_j|E, E_k).$$

The principle of rationality recommends the choice of A_i for which the expected utility of A_i , given E_k combined with E is maximized, ie

$$\max_i \sum_j q_{jk} u_{ij}.$$

Thus, before a new observation is made, the maximum expected utility is the weighted average of the maximum expected utilities associated with each outcome E_k , with the weights being the probabilities of E_k , that is,

$$\sum_{k=1}^t Pr(E_k) \left[\max_i \sum_j (q_{jk} u_{ij}) \right].$$

Now, as shown above,

$$\begin{aligned}
 Pr(E_k) &= \sum_j Pr(E_k|H_j, E)Pr(H_j|E) \\
 &= \sum_j Pr(E_k|H_j)Pr(H_j|E) \\
 &\quad \text{since } E \text{ and } E_k \text{ are independent} \\
 &= \sum_j p_{jk} p_j, \\
 \text{and } q_{jk} &= p_j p_{jk} / \sum_j p_j p_{jk} = p_j p_{jk} / Pr(E_k).
 \end{aligned}$$

Then

$$\begin{aligned}
\sum_k Pr(E_k) \left[\max_i \sum_j (q_{jk} u_{ij}) \right] &= \sum_k Pr(E_k) \left[\max_i \left(\sum_j \frac{p_j p_{jk}}{Pr(E_k)} u_{ij} \right) \right], \\
&= \sum_k \left[Pr(E_k) \left(\frac{1}{Pr(E_k)} \right) \max_i \left(\sum_j p_j p_{jk} u_{ij} \right) \right] \\
&\quad \text{since } Pr(E_k) \text{ is independent of } i \text{ and } j \\
&\quad \text{and can be taken outside the sum as a common} \\
&\quad \text{factor and outside the term } \max_i \\
&= \sum_k \left[\max_i \sum_j p_j p_{jk} u_{ij} \right].
\end{aligned}$$

Thus, the expected utility of an action based on new evidence is

$$\sum_k Pr(E_k) \left[\max_i \sum_j (q_{jk} u_{ij}) \right] = \sum_k \left[\max_i \sum_j p_j p_{jk} u_{ij} \right]. \quad (26)$$

This is the sum over all possible outcomes $\{E_1, \dots, E_t\}$ of the new evidence and is the expected utility before the new evidence is observed.

The rational choice of A_i if no new evidence (24) is taken is

$$\max_i \sum_j p_j u_{ij}. \quad (27)$$

Thus, it is desired to prove that (26) is no less than (is greater or equal to) (27). In mathematical terms, it is to be shown that

$$\sum_k \max_i \sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} \geq \max_i \sum_j Pr(H_j|E) u_{ij}. \quad (28)$$

The right-hand side of the inequality is the maximum utility without considering new evidence E_k . The left-hand side is the expected utility when E_k is considered.

The expression

$$\sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} = \sum_j p_j p_{jk} u_{ij}$$

is a summation over all r possible hypotheses/propositions and is therefore independent of any particular value of j . The expression varies in value depending on the action A_i taken and the observation E_k made, and only on that action and that observation. The expression is said to be a function f of i and k and thus it can be written as $f(i, k)$:

$$f(i, k) = \sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} = \sum_j p_j p_{jk} u_{ij}$$

and

$$\sum_k f(i, k) = \sum_k \sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} = \sum_k \sum_j p_j p_{jk} u_{ij} \quad (29)$$

is a function of the actions A_i , $\{i = 1, \dots, s\}$ only, since dependence on k has also been eliminated through summation.

Let i_0 be a value that maximizes (29), i.e.

$$f(i_0) = \max_i \sum_k f(i, k), \quad (30)$$

the maximum over i of the sums $\sum_k f(i, k)$.

Consider the $f(i, k)$, $j = 1, \dots, s$, individually for a particular value of k . Then $\max_i f(i, k)$ is the maximum value of all the $f(i, k)$, $i = 1, \dots, s$.

Thus, by definition,

$$\max_i f(i, k) \geq f(i_0, k),$$

since the left-hand side is the maximum of $f(i, k)$ over all i and in particular i_0 .

Sum both sides over k .

$$\sum_k \max_i f(i, k) \geq \sum_k f(i_0, k).$$

However, again by definition, i_0 is the value of i that maximizes $\sum_k f(i, k)$; see (30). Thus

$$\sum_k f(i_0, k) = \max_i \sum_k f(i, k).$$

Thus

$$\sum_k \max_i f(i, k) \geq \max_i \sum_k f(i, k). \quad (31)$$

From (28), if the result

$$\sum_k \max_i \sum_j p_j p_{jk} u_{ij} \geq \max_i \sum_j p_j u_{ij}$$

can be proved, then the questions ‘Is it worthwhile taking a new observation, taking account of new evidence?’ and ‘Is there an increase in utility if new evidence is considered?’ can be answered in the affirmative: ‘It is worthwhile taking a new observation, taking account of new evidence’ and ‘There is an increase in utility if new evidence is considered’.

Consider the right-hand side of (31).

$$\begin{aligned} \sum_k f(i, k) &= \sum_k \sum_j p_j p_{jk} u_{ij} \\ &= \sum_k \sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} \\ &= \sum_j \sum_k Pr(H_j|E) Pr(E_k|H_j) u_{ij} \\ &\quad \text{swapping order of summation; then } Pr(H_j|E) \\ &\quad \text{and } u_{ij} \text{ are common factors within the} \\ &\quad \text{summation over } k \\ &= \sum_j Pr(H_j|E) u_{ij} \sum_k Pr(E_k|H_j) \\ &= \sum_j Pr(H_j|E) u_{ij} \\ &\quad \text{since } \sum_k Pr(E_k|H_j) = 1, \text{ from (25)} \\ &= \sum_j p_j u_{ij}. \end{aligned}$$

Thus, the right-hand side of (31) equals $\max_i [\sum_j p_j u_{ij}]$.

Remember that $f(i, k) = \sum_j Pr(H_j|E) Pr(E_k|H_j) u_{ij} = \sum_j p_j p_{jk} u_{ij}$.

The left-hand side of (31), $\sum_k \max_i f(i, k)$ may then be written as

$$\sum_k \max_i f(i, k) = \sum_k \max_i \sum_j p_j p_{jk} u_{ij}.$$

This establishes the result

$$\sum_k \max_i \sum_j p_j p_{jk} u_{ij} \geq \max_i \sum_j p_j u_{ij}.$$

Thus: ‘It is worthwhile taking a new observation, taking account of new evidence’ and ‘There is an increase in utility if new evidence is considered’.

The result proves that it never leads to a decrease in utility to take new evidence. It is not a result which says what the decision should be once the evidence is taken. The result precedes the observation of new evidence. Once the new evidence has been taken and the observation made, say the observation is E_{k_0} , then the rational decision A_{i_1} is the one for which i_1 is the $\max_i \sum_j p_j p_{jk_0} u_{ij}$.

Appendix 2: Proof that $E_e(H|(E, \bar{E}), I) \leq E_e(H|I)$

The proof is taken from Horwich (2016: 119–20).

$$\begin{aligned} \text{Let } x &= Pr(H|I), y = Pr(H|E, I), \\ z &= Pr(H|\bar{E}, I), w = Pr(E). \end{aligned}$$

From Equations (14) and (15), the expected errors before and after observing E are

$$\begin{aligned} E_e(H|I) &= 2Pr(H|I)\{1 - Pr(H|I)\} = 2x(1 - x), \\ E_e(H|(E, \bar{E}), I) &= 2Pr(E|I)[Pr(H|E, I)\{1 - Pr(H|E, I)\}] \\ &\quad + 2Pr(\bar{E}|I)[Pr(H|\bar{E}, I)\{1 - Pr(H|\bar{E}, I)\}] \\ &= 2wy(1 - y) + 2(1 - w)z(1 - z). \end{aligned}$$

Note that

$$Pr(H) = Pr(E)Pr(H|E, I) + Pr(\bar{E})Pr(H|\bar{E}) \Rightarrow x = wy + (1 - w)z.$$

Then

$$\begin{aligned} \frac{1}{2} \{E_e(H|I) - E_e(H|(E, \bar{E}), I)\} &= \\ &= x(1 - x) - wy(1 - y) - (1 - w)z(1 - z) \\ &= x - x^2 - wy + wy^2 - (1 - w)z + (1 - w)z^2 \\ &= \{x - wy - (1 - w)z\} + \{wy^2 + (1 - w)z^2 - x^2\} \\ &= wy^2 + (1 - w)z^2 - x^2 \\ &= wy^2 + (1 - w)z^2 - \{wy + (1 - w)z\}^2 \\ &= wy^2 + (1 - w)^2 - w^2y^2 - (1 - w)^2z^2 - 2w(1 - w)yz \\ &= w(1 - w)y^2 + (1 - w)\{z^2 - (1 - w)z^2\} - 2w(1 - w)yz \\ &= w(1 - w)(y - z)^2 \\ &\geq 0, \end{aligned}$$

with equality if and only if $y = z$, $Pr(H|E, I) = Pr(H|\bar{E}, I)$ or the meaningless $w = Pr(E) = 1$ or 0 . Thus, the expected error, $E_e(H|I)$, before the observation of E is never less than the expected error $E_e(H|(E, \bar{E}), I)$ after the observation of E .