AUTHORS' POSTPRINT
Taroni, F., Garbolino, P. and Bozza, S.
Coherently updating degrees of belief: Radical Probabilism, the generalization of Bayes' Theorem and its consequences on evidence evaluation. Law, Probability and Risk 19 (2020) 293–316. doi:10.1093/lpr/mgab001

Coherently updating degrees of belief: Radical Probabilism, the generalization of Bayes' Theorem and its consequences on evidence evaluation

Franco Taroni[a,*], Paolo Garbolino[b], Silvia Bozza[c]

[a] *University of Lausanne, School of Criminal Justice, Lausanne, Switzerland*
[b] *Department of Architecture and Arts, IUAV University, Venice, Italy*
[c] *Department of Economics, University Ca' Foscari of Venice, Venice, Italy*

Abstract

The Bayesian perspective is based on conditioning related to reported evidence that is considered to be certain. What is called 'Radical Probabilism' replaces such an extreme view by introducing uncertainty on the reported evidence. How can such equivocal evidence be used in further inferences about a main hypothesis? The theoretical ground is introduced with the aim of offering to the readership an explanation for the generalization of the Bayes' Theorem. This extension – that considers uncertainty related to the reporting of evidence – also has an impact on the assessment of the value of evidence through the Bayes' factor. A generalization for such a logical measure of the evidence is also presented and justified.

Keywords: Bayes' Theorem, Radical Probabilism, Bayesian Conditionalization, Probability Kinematics, Bayes' Factor, Evidence Evaluation.

> [...] be the sensible man who tailors his beliefs to the available evidence.
> [20] at p. 311

1. Introduction

The core of the Bayesian perspective to the scientific method can be concisely described in the following terms. Scientific rationality - related to the criteria used to check scientific hypotheses and their plausibility - should refer to a probabilistic framework. We fully endorse this point of view and have, as expressed by Galavotti [14] (at p. 253), the 'conviction that the entire edifice of human knowledge rests on

*Corresponding author
Email address: franco.taroni@unil.ch (Franco Taroni)

probability judgments, not on certainties'[1]. This means that to rationally assess the plausibility of different hypotheses considered at a certain historical moment, say t, scientists should determine the probability of each hypothesis in the light of the totality of information available at that time and update these probabilities as new information are acquired, for example through laboratory experiments or observations. The study of the rational principles governing the attribution of probabilities to hypotheses and observations, and their updating according to new available information is therefore fundamental.

Two articles published in this journal [42, 43] have endorsed such a subjectivist interpretation of probability. In [43] attention was stressed upon the fact that such subjective probabilities can be correctly informed by relative frequencies, whenever available, though this does not equate at all to the acceptance of a frequentist interpretation of probability.[2]. Briefly, given a subjectivist interpretation of probability, opinions of a given subject at a given time can be represented in probabilistic form by specifying their degrees of belief; these beliefs must obey the principles of probability calculus in order to be called coherent. This is generally illustrated through the so-called *Dutch Book argument*, that is, based on the idea that a given subject can avoid bets that would make him suffer a certain loss if, and only if, his degrees of belief are probabilities and therefore managed by the rules of probabilities. Nevertheless, the requirement of coherence at a particular time says nothing about how a given individual should modify their beliefs in light of newly acquired evidence.

The study of the rational principles underlying the updating of personal beliefs can be seen as complementary and a natural consequence to the arguments developed in [42, 43]. One's probabilities should be updated on the basis of appropriate principles in order to guarantee a rational approach. Please note that updating one's mind in the light of newly acquired evidence does not mean changing one's opinion. de Finetti [8] illustrated this aspect in the following terms:

> If we reason accordingly to Bayes' theorem we do not change opinion. We keep the same opinion and we update it to the new situation. If yesterday I said 'Today is Wednesday', today I say 'It is Thursday '. Yet, I have not changed my mind, for the day following Wednesday is indeed Thursday. (at p. 100)

The cardinal principle for the updating of personal beliefs is the one known under the name of *Conditioning principle* which - broadly speaking - states that if the initial opinions of a subject are represented by a probability function $\Pr(\cdot)$, subsequently, the subject acquires empirical evidence or observations, say E, then their final opinions should be represented by the conditional probability function $\Pr(\cdot \mid E)$. The conditional

[1] As reported by [14] (at p. 253) the origin of such a view can be found in de Finetti's pragmatism: 'First of all, this involves a rejection of the notion of 'absolute truth'. Such a rejection is the origin of probabilism, taken as the attitude according to which we can only attain knowledge that is probable.' For a historical discussion on 'probabilism', see [34]

[2] As de Finetti wrote [7]: '[...] subjectivism does not mean ignoring or neglecting objective data, but rather using them as a sensible responsible way, instead of appealing to oversimplified and stereotyped schemes.' (at p. 97)

probability depends on the evidence E that has been acquired, and not on what could have been observed, but, in fact, it has not.

This principle, which will be described in Sections 3 and 4, can only be applied if the acquired information is made up of *certain* evidence. In many cases, in truth, very often, this information is made up of *uncertain* evidence, so we should speak of acquisition of *soft evidence*[3] compared to *hard evidence*[4]. The problem is well posed by Schum [38] who wrote:

> If we contemplate using Bayes' rule, how are we to revise our own prior belief about [hypotheses] H and $\bar{H}$, based upon evidence it is not a proposition but a probability distribution? (at p. 352)

This aspect is strictly related to a doctrine called *Radical Probabilism* that holds that no facts are known for certain.

In most forensic and judicial literature related to probabilistic approaches to evidence evaluation and interpretation, it can be noticed that a discussion about the quantification of degrees of belief in the presence of soft evidence is completely missing. It is therefore a question of exposing how a rational individual can govern the rational update of his state of mind when the acquired information is not constituted by certain evidence (also called *hard* evidence). This paper is devoted to considering how rational individuals ought to revise probability judgments in reaction to their experiences in presence of *soft* evidence.

The paper provides a comprehensive discussion on this issue and is structured as follows. Section 2 gives an overview of fundamental definitional aspects for the representation of beliefs. Section 3 and Section 4 cover the aspects related to the revision of beliefs in presence of hard and soft evidence, respectively, by introducing the concept of probability kinematics, with a simple example developed in Section 5. Section 6 highlights the connection between probability kinematics and Schum's cascaded inference. Section 7 and 8 extends further the discussion to evidence evaluation through the use of Bayes' factor for unequivocal testimony (Section 7) and a Bayes' factor adapted to take into account for uncertainty on the reported evidence: equivocal evidence (Section 8). Section 9, finally, concludes the paper.

2. Logic for reasoning under uncertainty: representation of beliefs and relevant propositions

Let us briefly recall some fundamental definitional aspects suitable for describing how it is possible to 'measure' degrees of belief that a certain fact occurred. It is first necessary to be precise about what a 'degree of belief' is. A degree of belief is *personal*, it is the judgement of a given person at a given time about the truth of a given event

[3]Soft evidence can be interpreted as evidence of uncertainty, sometimes called 'probable knowledge'. There is uncertainty about the specific state of a variable of interest but there is a probability assignment associated with in.

[4]Hard evidence is knowledge that some state of a variable definitely occurred, so that information arrives in the form of a proposition stating that event, say E, occurred.

or proposition. 'Evidence' bearing on that proposition is expressed by means of other propositions. Therefore, it shall be said, on first approximation, that a proposition, say B, is *relevant*, according to an opinion, for another proposition A if, and only if, knowing the truth (or the falsity) of B would change the degree of belief in the truth of A.

There is, however, an important distinction between defining the general concept of representation of beliefs and the evaluation of a particular case. It can be established that coherent, quantitative measures of uncertainty about events or propositions of interest must take the form of probabilities [2]. These probabilities are subjective, since they are based on personal judgements, they are a personal numerical representation of the uncertainty relation between events. Consider a *probability function*, denoted by the symbol $\Pr(\cdot)$ where the $(\cdot)$ contains the event or proposition, the probability of which is of interest. Numerical degrees of belief must satisfy, for any propositions A and B, the laws of probability theory. De Groot [9] proved that the acceptance of certain assumptions concerning the uncertainty relation between events (e.g. the complete comparability of events) implies that for any event (or proposition) A there exist a unique probability $\Pr(A)$ satisfying the laws of probability theory (axioms of probability). Briefly, it can be proved that (*i*) if it is known that a proposition is true (false), then the degree of belief should take the maximum (minimum) numerical value, that is $\Pr(A) = 1$ ($\Pr(A) = 0$); (*ii*) coherent quantitative degrees of belief have a finitely additive structure (i.e., if A and B are incompatible, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$); (*iii*) events which are practically possible but non certain should be assigned a probability in the interval $(0, 1)$, $0 < \Pr(A) < 1$. The degrees of belief in events not known to be true, are somewhere between the certainty that the event is true and the certainty that it is false. Some more results follow from these axioms. A straightforward consequence of the axiomatic formulation is that if the probability of an event A is assessed, the probability of its complement $\bar{A}$ (i.e. the event that takes place when A is not true) can be simply obtained as $\Pr(\bar{A}) = 1 - \Pr(A)$.

This is the simplest example of how probability calculus works as a *logic for reasoning under uncertainty*. The logic places constraints on the ways in which numerical degrees of belief may be combined. Notice that the laws of probability require the degrees of belief in any two mutually exclusive and exhaustive events A and in B to be such that they are non-negative and their sum is equal to one. Within these constraints, there is not an obligation for A to take any particular value. Any value between the minimum (0) and the maximum (1) is allowed by the probability axioms. The question is: how is a subjective probability to be determined? How is newly available knowledge to be incorporated? We will focus now on the first question, while the second one will be addressed later in Section 3 and in Section 4. The simplest way to measure a subjective probability is a direct measurement technique based on comparison with standard events with given probabilities [35]:

> A formal derivation of subjective probability based on this approach would need to assume that any two events can be compared to say which You regard as the more probable, and also that there exists a set of standard events with given probabilities against which any other event can be compared to determine its probability. (at p. 98)

A subject is entitled to its own measures of belief, but must be consistent with them. The appropriateness of a set of probabilities held by a particular subject needs to be checked [36]. Probability values need to be expressed in an operational way that will also make clear what coherence means and what coherent conditions are. de Finetti [6] framed the operational perspective as follows:

> However, it must be stated explicitly how these subjective probabilities are defined, i.e., in order to give an operative (and not an empty verbalistic) definition, it is necessary to indicate a procedure, albeit idealized but not distorted, an (effective or conceptual) experiment for its measurement. (at p. 212)

Therefore, one should keep in mind the distinction between the *definition* and the *assignment* of probability. A description of de Finetti's perspective has been published by [5].

Note that coherence on degrees of beliefs is the only condition that should be guaranteed. Assignments that violate the laws of probability are incoherent in the sense that they will lead to a sure loss, no matter which proposition turns out to be true (see, for example, [25]).

Another fundamental aspect for the definition of one's degree of belief is that of *relevance*. A proposition B is said to be relevant for another proposition A if and only if the answer to the following question is positive: if it is supposed that B is true, does that supposition change the degree of belief in the truth of A? A judgment of relevance is an exercise in hypothetical reasoning. There is a search for a certain kind of evidence because it is known in advance that it is relevant; if someone submits certain findings maintaining that they constitute relevant evidence, a hypothetical judgment has to be made as to whether or not to accept the claim. In doing that, a distinction has to be drawn, not only between the proposition A and the evidence B for the proposition A, but also between that particular evidence B and whatever else is known.

When a proposition's degree of belief is assigned, there is always exploitation of available *background information*, even though it is not explicit. An assessment of the degree of belief in the proposition 'this coin lands heads after it is tossed' is made on the basis of some background information that has been taken for granted: if the coin looks like a common coin from the mint, and there is no reason for doubting that, then it is usually assumed that it is well balanced. Should it be realized, after inspection, that the coin is not a fair coin, this additional information is 'evidence' that changes the degree of belief about that coin, even though it is still believed that coins from the mint are well balanced.

A *relevant proposition* is taken to mean a proposition which is not included in the background information. The distinction between 'evidence' and 'background information' is important, because sometimes it has to be decided that certain propositions are to be considered as evidence, while others are to be considered as part of the background information. For example, suppose a DNA test has been evaluated. Assume that all scientific theories which support the methodology of the analysis are true, that the analysis has been done correctly, and that the chain of custody has not been broken. These assumptions all form part of the background information. Relevant evidence is is made up of only those propositions which describe the result of the test, *plus* some other propositions reporting statistical data about the reliability of the evidence. Alternatively, propositions concerning how the analysis has been done, and/or the chain of custody can also be taken to be part of the evidence whilst scientific theories are to be considered background information. Therefore, as said before, it is useful to make a clear distinction between what is considered, in a particular context, to be 'evidence', and what is considered to be 'background'. The importance of this aspect will be discussed further in Sections 3 and 4. For this reason, background information is introduced explicitly in the notation.

Let $\Pr(A \mid I)$ denote 'the degree of belief that proposition A is true, given background information I', and let $\Pr(A \mid B, I)$ denote 'the degree of belief that proposition A is true, given that proposition B is assumed to be true, *and* given background information I'. Given that a probability - the measure of uncertainty - as a degree of belief, is conditional on the status of information of a given subject who assigns it[5], $\Pr(A \mid B, I)$ should be written as $\Pr_{s,t}(A \mid B, I)$ where s and t are the information available to subject s at time t, respectively. For ease of notation, in what follows, the subscripts s and t are omitted. In Section 3, it will be shown why time t plays an important role in the beliefs updating procedure.

The purpose here is only to emphasize the point that *all subjective probabilities are conditional on available knowledge*. It is obvious that personal beliefs depend upon the particular knowledge one has. If the choice is made to represent degrees of belief by means of probabilities, then it must be kept in mind that it will always be the case that probabilities are necessarily relative to the available knowledge *and* to the assumptions made. Note that the term I for background information is often omitted, and it is solely reported $\Pr(A)$. This probability expresses the personal degree of belief that the event A is true, given all available knowledge I. Note that, as expressed in [45]:

> It should be emphasized that logically speaking - and contrary to many textbook expositions - this conditional probability has nothing to do with learning or opinion change or 'updating' on new information. Literally, that my conditional probability $\Pr(A \mid I)$ equals $1/3$, for example [...] clearly expresses only my present opinion about two eventualities. (at p. 18)[6].

Once the occurrence of another event or proposition B (e.g., the scientific evidence) is observed, the degrees of belief can be updated by incorporating the newly available knowledge. This is usually indicated by $\Pr(A \mid B)$, though what is really meant is $\Pr(A \mid B, I)$. The letter I is suppressed for simplicity of notation, but $\Pr(A \mid B)$ must be read as the current degrees of belief about the truth of A, given the evidence B and all

[5]Galavotti [14], by quoting de Finetti, wrote: '[...] probability can only be taken as the expression of the feelings of the subjects who evaluate it. Being matter for subjective opinions, probability evaluations are always definite and known. Put differently, 'unknown probabilities' taken as objective 'true' probabilities pertaining to phenomena do not exist; in their place we have subjective evaluations which can always be formulated, insofar as they are the expression of the feelings of evaluating subjects.' (at p. 259).

[6]Note that in the original quote the background information was denoted by B. The letter B has been replaced here by I for homogeneity with the text where the background information is denoted by I, while the letter B is used to denote the evidence.

background knowledge.

3. Logic for reasoning under uncertainty: probabilities' updating with certain evidence

Section 2 served the purpose of discussing both definitional and operational aspects relative to the representation of beliefs. It has been clarified that degrees of belief must take the form of subjective probabilities, and that these must obey the laws of probability calculus in order to be considered a coherent representation of beliefs. But there is another question that must be answered: *how is a new item of evidence to be incorporated in one's knowledge?* The answer is in some cases given by the well known Bayes' Theorem, as it will be recalled in the current section, but there are instances for which there is not an immediate answer. This will be addressed in Section 4.

Consider a proposition A and another proposition B, relevant for A. As it was highlighted in Section 2, a relevant proposition B, with respect to another proposition A, is a proposition such that the supposition, if it were true, would change the degrees of belief in A. The new degrees of belief are given by $\Pr(A \mid B, I)$, that is the conditional measure of degree of belief (with the information regarding B), to the initial measure of degree of belief (without the information regarding B).

The conditional probability $\Pr(A \mid B, I)$ can be defined as follows. For any propositions A and B, the degree of belief that A is true, given that one assumes that B is true, is equal to the degree of belief that A and B are both true, given background information I, divided by the degree of belief that B is true, given background information I, provided that $\Pr(B \mid I) > 0$:

$$\Pr(A \mid B, I) = \frac{\Pr(A, B \mid I)}{\Pr(B \mid I)}. \tag{1}$$

Equation (1) can be rearranged as

$$\Pr(A \mid B, I) = \frac{\Pr(B \mid A, I)\Pr(A \mid I)}{\Pr(B \mid I)}. \tag{2}$$

Equation (2) is the simplest algebraic version of the formula first proved in the second half of the eighteenth century by the Reverend Thomas Bayes. The importance of Bayes' Theorem is due to the fact that it is a rule for *turning around* conditional probabilities and *updating degrees of belief on receiving new evidence*.

The process of evidence acquisition may be modeled as a two-step process in time. At time t_0 it is planned to look for evidence B because B is believed to be *relevant* for proposition A. At time t_0 the change in degree of belief in A, if it were to be discovered that B were true, may be calculated by use of Equation (2). Denote the degrees of belief about an event $\cdot$ at time t_0 by $\Pr_0(\cdot)$. The Bayes' rule in (2) has a more general form that applies to partitions of the sample space. Let us consider the partition given by proposition A and its complement $\bar{A}$. Bayes' Theorem may be rewritten as

$$\Pr_0(A \mid B, I) = \frac{\Pr_0(B \mid A, I)\Pr_0(A \mid I)}{[\Pr_0(B \mid A, I)\Pr_0(A \mid I)] + [\Pr_0(B \mid \bar{A}, I)\Pr_0(\bar{A} \mid I)]}. \tag{3}$$

The result in the denominator is an expression of the law of total probability also known as 'Extension of the conversation rule' [31]). The probability $\Pr_0(A \mid B, I)$ is also called the probability of A, *conditional on* B (at time t_0) or the *posterior* probability of A. $\Pr_0(A \mid I)$ and $\Pr_0(\bar{A} \mid I)$ are called *prior*, or *initial*, probabilities. $\Pr_0(B \mid A, I)$ is called the *likelihood* of A given B (at time t_0). Analogously, $\Pr_0(B \mid \bar{A}, I)$ is the *likelihood* of $\bar{A}$ given B (at time t_0).

At time t_1 it is discovered that B is true. Denote the degree of belief at time t_1 by $\Pr_1(\cdot)$: what is the degree of belief in the truth of A at time t_1, *i.e.*, what is $\Pr_1(A \mid I)$? A reasonable answer seems to be that, if it has been learned at time t_1 that B is true, and no further information is available, then knowledge that B is true has become part of the *background knowledge* at time t_1; therefore, the overall degree of belief in A at time t_1 is equal to the degree of belief in A, conditional on B, at time t_0:

$$\Pr_1(A \mid I) = \Pr_0(A \mid B, I). \tag{4}$$

This process is called *Simple conditioning principle* [33, 12]; it represents the process in which the prior probability of each proposition A is replaced by a posterior probability that coincides with the prior probability of A conditional on B at time t_1. Jeffrey [26] noticed that:

> In the unlikely event that your judgmental states today and tomorrow are both representable by definite probability distributions - say, P for today's distribution and Q for tomorrow's - it may be that tomorrow's unconditional probabilities $Q(H)$ for hypothesis H are simply today's conditional probabilities $P(H \mid A)$ for those hypotheses given the answer A to some question. This is the prized special case of *Conditioning* $Q(H) = P(H \mid A)$ (or 'conditionalization'). (at p. 215)

In order to clarify the general mechanism underlying probabilistic updating, it may be helpful to consider Equation (4) in another way.

Consider all the possible scenarios that can be derived from the combination of two logically compatible propositions A and B at time t_0: $\Pr_0(A, B \mid I), \Pr_0(A, \bar{B} \mid I), \Pr_0(\bar{A}, B \mid I)$, and $\Pr_0(\bar{A}, \bar{B} \mid I)$. These scenarios can be represented graphically by means of a *probability tree*, as shown in Figure 1. At time t_0, it is not known if proposition A is true or not; the same for proposition B. Four exclusive and exhaustive scenarios do exist but we do not know which one is true. We just know that the probabilities related to this scenario must sum up to 1, $\Pr_0(A, B \mid I) + \Pr_0(A, \bar{B} \mid I) + \Pr_0(\bar{A}, B \mid I) + \Pr_0(\bar{A}, \bar{B} \mid I) = 1$.

A probability tree is a type of graphical model which consists of a series of branches stemming from nodes, usually called random nodes, which represent uncertain events [40]. At every random node there are as many branches as the number of the possible outcomes of the uncertain event. In this context, outcomes of uncertain events are described by propositions and branches containing more than one node correspond to the logical conjunction of as many propositions as the number of nodes. Branches have associated with them probabilities of the corresponding conjunctions of propositions calculated *via* the multiplication rule, and the probability of each proposition is given

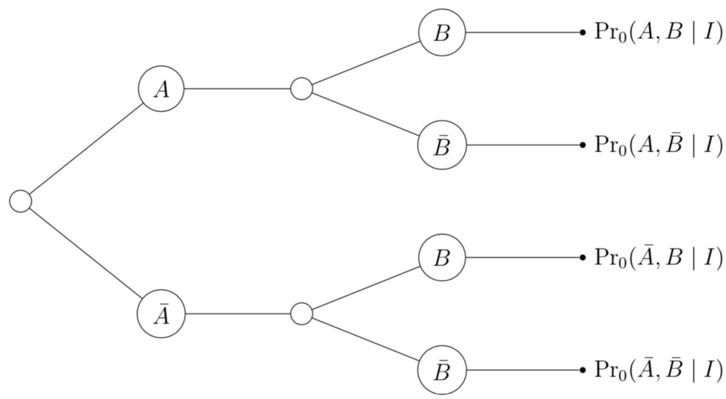

Figure 1: The probability tree for propositions A and B at time t_0, given background information I.

by the sum of the probabilities of the branches containing it (extension of the conversation). Following the axioms of probability, the sum of the probabilities of all the branches must add to one.

Consider now time t_1, where proposition B is known to be true. The probability tree at time t_1 is shown in Figure 2. Since B is known to be true, only two scenarios are possible in the new state of knowledge and thus two of the probability nodes in Figure 2, inconsistent with B, must be assigned a probability equal to zero: $\Pr_1(A, \bar{B} \mid I) = \Pr_1(\bar{A}, \bar{B} \mid I) = 0$.

Again, the sum of the probabilities of the branches in Figure 2 must add up to 1. The original probabilities of the branches must be amended in such a way that their sum turns out to be one: $\Pr_1(A, B \mid I) + \Pr_1(\bar{A}, B \mid I) = 1$.

How can we redistribute these new probabilities? Considering that the only change in the state of information that has occurred is the probability of B, and no new information has been given about the probability ratios of different branches, then it is reasonable to redistribute beliefs in such a way that the ratio between the new and the old probabilities of the branches containing B is the same as the ratio between the new and the old probabilities of B:

$$\frac{\Pr_1(A, B \mid I)}{\Pr_0(A, B \mid I)} = \frac{\Pr_1(B \mid I)}{\Pr_0(B \mid I)}. \tag{5}$$

Equation (5) can be called *Principle of symmetry* (see, for example, [15] for historical comments and [16] for applications). In this way, the new information B is distributed symmetrically and neither of the two scenarios is privileged. A simple manipulation of

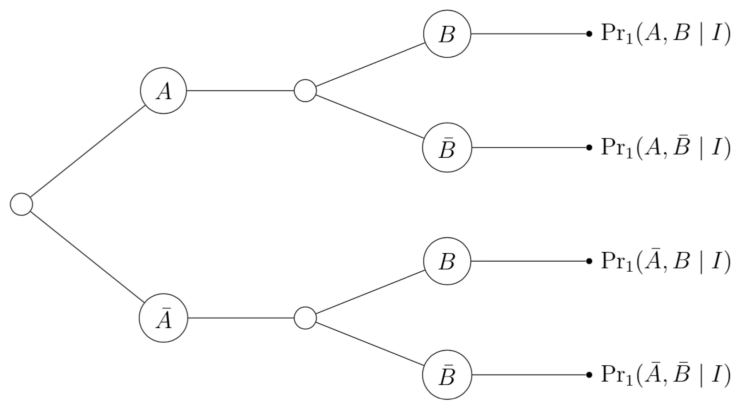

Figure 2: The probability tree for propositions A and B at time t_1, given background information I.

(5), as well as the fact that $\Pr_1(B \mid I) = 1$, allows us to express $\Pr_1(A, B \mid I)$ as:

$$\Pr_1(A, B \mid I) = \Pr_0(A, B \mid I) \times \frac{\Pr_1(B \mid I)}{\Pr_0(B \mid I)} = \frac{\Pr_0(A, B \mid I)}{\Pr_0(B \mid I)}. \tag{6}$$

Analogously, it can be obtained $\Pr_1(\bar{A}, B \mid I)$ as:

$$\Pr_1(\bar{A}, B \mid I) = \Pr_0(\bar{A}, B \mid I) \times \frac{\Pr_1(B \mid I)}{\Pr_0(B \mid I)} = \frac{\Pr_0(\bar{A}, B \mid I)}{\Pr_0(B \mid I)}. \tag{7}$$

Indeed, addition of the terms in Equation (6) and (7) gives the desired result, that is

$$\begin{aligned} \Pr_1(A, B \mid I) + \Pr_1(\bar{A}, B \mid I) &= \frac{\Pr_0(A, B \mid I)}{\Pr_0(B \mid I)} + \frac{\Pr_0(\bar{A}, B \mid I)}{\Pr_0(B \mid I)} \\ &= \frac{\Pr_0(B \mid I)}{\Pr_0(B \mid I)} = 1. \end{aligned}$$

The new probability of A is now equal to the probability of the branch containing both A and B. Since it is known that B is true, $\Pr_1(A, \bar{B} \mid I) = 0$ and hence $\Pr_1(A \mid I) = \Pr_1(A, B \mid I)$. Therefore, substitution of this result in Equation (6) and (7) gives Equation (4), in fact

$$\begin{aligned} \Pr_1(A \mid I) = \Pr_0(A \mid B, I) &= \frac{\Pr_0(A, B \mid I)}{\Pr_0(B \mid I)}; \\ \Pr_1(\bar{A} \mid I) = \Pr_0(\bar{A} \mid B, I) &= \frac{\Pr_0(\bar{A}, B \mid I)}{\Pr_0(B \mid I)}, \end{aligned} \tag{8}$$

where evidence B at time t_0 becomes *background knowledge* at time t_1 as previously discussed in Section 2.

Results in (6) and (7) allow us to show that the use of Bayes' Theorem for probability updating is equivalent to the redistribution of probabilities among possible scenarios in a symmetric way; given that the state of information has changed on learning only that B is true, $\Pr_1(B \mid I) = 1$, and nothing else, there is no reason to make a change biased for or against certain particular scenarios.

4. Logic for reasoning under uncertainty: probabilities updating with uncertain evidence

The *Simple conditioning principle* (Section 3) is not widely applicable. Suppose that a given report expresses some uncertainty and we cannot be certain of evidence B as truth. We wish to use this 'equivocal' report in a further inference about propositions A and $\bar{A}$ given that B is not known to be true but there is instead a probability distribution expressing uncertainty about it [38] (at p. 352). The problem is well posed by Jeffrey [23][7] and reiterated in [24]:

> H is the hypothesis that it's hot out. Smith and Jones have each testified as to H's truth or falsity, but you don't regard their testimony as completely reliable. Perhaps you're not even quite sure what either of them said. (You are deep in a tunnel, into which they have shouted their report.) Let E and F be the propositions that Smith and Jones, respectively, said that it is hot. How can you represent your judgment about E and F, and your consequent judgment about H? (at p. 391)

We are so often faced with uncertain evidence. We can more drastically say that examples of this type are the norm in real life situations as in forensic science and in judicial settings where things are partially perceived or remembered. An illustrative example in medicine is described in [26].

Consider a forensic situation in which our previous propositions A and $\bar{A}$ are substituted with H and $\bar{H}$ (for *Hypotheses*) as routinely used in legal literature. Notation B is also updated by using letter E (for *Evidence*) to take into account the scientific information delivered by a forensic scientist through his report.

Assume, for the sake of argument, that the scientist's degree of belief in the truth of proposition E at time t_1 is higher than his initial degree of belief at time t_0, but it falls short of certainty, that is $\Pr_0(E \mid I) < \Pr_1(E \mid I) < 1$. What is the effect of this *uncertain evidence* upon the hypotheses H and $\bar{H}$?

The problem is that the scientist *cannot* take the probability of H conditional on E as his new degree of belief, because *he does not know* E for certain. If there is any uncertainty left in the report, the best that the scientist can do is to assess directly the effect of such imperfect information on his degrees of belief, and, indeed, this is all that the scientist is reasonably entitled to do. A probabilistic rule can be formulated

[7] The first edition of [23] dated back to 1965.

that allows him to update directly his degrees of belief once dealing with uncertain evidence E.

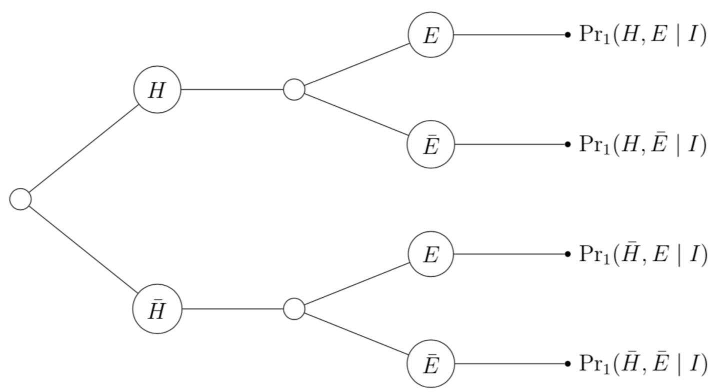

Figure 3: The probability tree for the hypothesis H and the evidence E at time t_1, given background information I.

Consider, for the sake of clarity, the probability tree of the scientist's problem given in Figure 3, *i.e.*, a probability tree containing the nodes H (*Hypothesis*) and E (*Evidence*). The difference with the probability tree in Figure 2 is that the probabilities of all four scenarios at time t_1 are now greater than zero. The problem is thus: *how can a decision-maker redistribute the probabilities of the four different scenarios in such a way that they add up to 1?*

Taking the same arguments illustrated in Section 3, it is entirely reasonable to redistribute probabilities so that the ratio between the new and the old probabilities of the scenarios is the same as the ratio between the new and the old probabilities of E as expressed in Section 3 through the *Principle of symmetry*:

$$\frac{\Pr_1(H, E \mid I)}{\Pr_0(H, E \mid I)} = \frac{\Pr_1(E \mid I)}{\Pr_0(E \mid I)}. \tag{9}$$

From Equation (9) the rule for calculating the new probability $\Pr_1(H, E \mid I)$ is then derived:

$$\Pr_1(H, E \mid I) = \Pr_0(H, E \mid I) \times \frac{\Pr_1(E \mid I)}{\Pr_0(E \mid I)}. \tag{10}$$

In this way, the change in the scientist's state of information between time t_0 and time t_1 is taken into account about event E. The other probabilities corresponding to all branches of the probability tree can be calculated analogously. It can be easily verified that they sum up to 1.

Equation (10) obeys the same principle as Equation (6). The constant factor now is $\Pr_1(E \mid I)/\Pr_0(E \mid I)$ instead of $1/\Pr_0(B \mid I)$, but the probabilities have been redistributed among the possible scenarios in a symmetric way as before: given that the state of information has changed on learning the new probability $\Pr_1(E \mid I)$, and nothing else happened, there are no reasons to make a change biased for or against certain particular scenarios.

The probability of hypothesis H at time t_1 can be obtained as:

$$\begin{aligned}
\Pr_1(H \mid I) &= \Pr_1(H, E \mid I) + \Pr_1(H, \bar{E} \mid I) \\
&= \Pr_0(H, E \mid I)\frac{\Pr_1(E \mid I)}{\Pr_0(E \mid I)} + \Pr_0(H, \bar{E} \mid I)\frac{\Pr_1(\bar{E} \mid I)}{\Pr_0(\bar{E} \mid I)} \\
&= \Pr_0(H \mid E, I)\Pr_0(E \mid I)\frac{\Pr_1(E \mid I)}{\Pr_0(E \mid I)} + \Pr_0(H \mid \bar{E}, I)\Pr_0(\bar{E} \mid I)\frac{\Pr_1(\bar{E} \mid I)}{\Pr_0(E \mid I)} \\
&= \Pr_0(H \mid E, I)\Pr_1(E \mid I) + \Pr_0(H \mid \bar{E}, I)\Pr_1(\bar{E} \mid I). \qquad (11)
\end{aligned}$$

Equation (11) is a straightforward *generalization of Bayes' Theorem*, known in the philosophical literature under the name of *Jeffrey's rule* because it was the philosopher of science Richard Jeffrey who first argued it was a reasonable general updating rule [23]. The approach is also known as *Generalized conditioning*, *Jeffrey conditioning* or *Probability Kinematics*. Jeffrey conditioning can be formally expressed in the following terms under a dichotomous situation:

> If a person with a prior such that $0 < \Pr_0(E \mid I) < 1$ has a learning experience whose sole immediate effect is to change their subjective probability for E to $\Pr_1(E \mid I)$, then their post-learning posterior for H should be $\Pr_1(H \mid I) = \Pr_1(E \mid I) \times \Pr_0(H \mid E, I) + [1 - \Pr_1(E \mid I)] \times \Pr_0(H \mid \bar{E}, I)$.

Note that Jeffrey conditioning reduces to *Simple conditioning rule* (Section 3) when $\Pr_1(E \mid I) = 1$; so, note that the Simple conditioning rule is a special case of Jeffrey's rule. So, as expressed by [21], Jeffrey conditionalization offers 'a consistent and certainly convenient release from the apparent dogmatism implicit in ordinary conditionalization. If we regard probabilities of 1 as in practice unattainable, we can view ordinary conditionalization merely as a convenient approximation of Jeffrey conditionalization for a proposition whose probability shifts to almost 1.' (at p. 204)

Schum ([38]) explains Jeffrey's rule of conditioning in the following terms:

> According to Jeffrey's rule, we determine $\Pr(H \mid W\text{'s probabilistic report})$ as follows: We first determine $\Pr(H \mid E)$, *as if* E did occur; this requires the priors $\Pr(H)$, $\Pr(\bar{H})$ and the likelihoods $\Pr(E \mid H), \Pr(E \mid \bar{H})$. We then determine $\Pr(H \mid \bar{E})$, *as if* E did not occur. This requires the above priors and the likelihoods $\Pr(\bar{E} \mid H), \Pr(\bar{E} \mid \bar{H})$. The final step involves taken the weighted average of $\Pr(H \mid E)$ and $\Pr(H \mid \bar{E})$, where the weights are W's assessments of the likeliness that he observed E or $\bar{E}$. (at p. 352)

In the discussion above, it has been considered to be a case where the evidence can take the form of a feature correspondance and its complement. There are occasions however where the evidence does not have such a dichotomic structure. For example, consider a car traffic accident and one's interest in paint flakes colors. Jeffrey's rule can be extended for any partition $E_1, \ldots, E_n$ (where n explicits, e.g., the number of paint flake colors that are taken into consideration). The degree of belief about the truth of H in presence of uncertain evidence $E_1, \ldots, E_n$ can be expressed as

$$\begin{aligned} \Pr_1(H_i \mid I) &= \sum_{j=1}^{n} \Pr_0(H_i \mid E_j, I)\Pr_1(E_j \mid I) \\ &= \Pr_0(H_i) \sum_{j=1}^{n} \frac{\Pr_0(E_j \mid H_i)}{\Pr_0(E_j)} \Pr_1(E_j \mid I). \end{aligned} \tag{12}$$

Comments and discussions on Formula (12) can be found in [18, 10, 41, 22, 25, 3] and a brief historical discussion is presented in Section 6.

5. A simple numerical example

Consider, just for sake of illustration, a simple case involving a medical diagnosis taken from [1]. A medical blood test detects certain symptoms of a disease, but unfortunately, the test may not always register the symptoms when they are present, or it may register them when they are absent. Therefore there is the need to evaluate the accuracy of the test; this can be done by assessing the sensitivity and the specificity of the performed test.

Letting H be the event that a person is affected by a given disease and E stand for the event that the test indicates a positive result, $\Pr_0(E \mid H, I)$ and $\Pr_0(\bar{E} \mid \bar{H}, I)$ stand for the sensitivity and the specificity of the test, respectively. Suppose now that the sensitivity of the test is equal to 0.95, while the specificity is equal to 0.99. This is equivalent to assess $\Pr_0(E \mid H) = 0.95$ and $\Pr_0(\bar{E} \mid \bar{H}) = 0.99$. Suppose also that the prevalence of the disease in the relevant population is known to be 0.1, that is assume $\Pr_0(H_i \mid I) = 1$.

The posterior probability, $\Pr_0(H \mid E, I)$, that a given member of the relevant population has the disease given the observation of the positive result and the background information I, becomes

$$\begin{aligned} \Pr_0(H \mid E, I) &= \frac{\Pr_0(E \mid H, I) \times \Pr_0(H \mid I)}{\Pr_0(E \mid H, I) \times \Pr_0(H \mid I) + \Pr_0(E \mid \bar{H}, I) \times \Pr_0(\bar{H} \mid I)} \\ &= \frac{0.95 \times 0.1}{0.95 \times 0.1 + (1 - 0.99) \times 0.9} \\ &= 0.913. \end{aligned} \tag{13}$$

If the scientist declares unambigously that the blood test is positive, then the probability that the individual is effectively affected by the disease increases from a prior probability of 0.1 to a posterior probability greater than 0.91.

Consider now the case where for some reasons the scientist is doubtful about the result of the test. Imagine there is just a 0.7 probability for the correctness of the reported result of the test. The scientist cannot present such result in an unequivocal way. By taking Jeffrey's rule, probabilities 0.7 and its complement 0.3 are used as weights $\Pr_1(E \mid I)$ and $\Pr_1(\bar{E} \mid I)$ for the corresponding conditional probabilities $\Pr_1(H \mid E, I)$ and $\Pr_1(H \mid \bar{E}, I)$. The posterior probability of H at time t_1 becomes:

$$\begin{aligned} \Pr_1(H \mid I) &= \Pr_1(E \mid I) \times \Pr_0(H \mid E, I) + [1 - \Pr_1(E \mid I)] \times \Pr_0(H \mid \bar{E}, I) \quad (14) \\ &= 0.7 \times 0.913 + 0.3 \times 0.006 \\ &= 0.641. \end{aligned}$$

Results obtained through Jeffrey's rule by changing prior probability on H are a linear function. Note that once there is no uncertainty about the evidence E, $\Pr_1(E \mid I)$ can be set equal to 1, and the posterior probability in (14) is equivalent to that obtained in (13) for unequivocal evidence.

6. Some remarks of interest for the forensic scientist

In the previous section, we approached the problem of handling either hard or soft evidence through the quantification of the posterior probability of the main hypothesis of interest, say H.

Situations involving equivocal evidence are not so rare in forensic science. See, for example, data obtained to support classification of individuals according to an age threshold, given sex and the third molars' dental maturity on a given scale as presented in [4] in which the forensic scientist has difficulties in classifying third molars into one definite stage, for example due to an unclear radiographic image. Other scenarios involve the case where there is uncertainty about the reported testimony related to the hair color of a person of interest, or there is uncertainty about animal hair classification by microscopy, or about the result of a test. An example involving how to deal with equivocal testimony about the result of a test is provided in Section 8.

Dodson [11] gave a rule for beliefs updating based upon probabilistic equivocation of an evidence. He called it *Modified Bayesian theorem* and his solution represented exactly Jeffrey's conditioning rule. It seems that the model was criticized on its axiomatic aspects but it has been developed further in [17] independently from Jeffrey (1965) [23]. Their aim was to propose

> An algorithm [...] that relaxes the requirement of Bayes' theorem that the true data state be known with certainty by postulating a true but unobservable elementary event, w [E in our notation], which gives rise to posterior probabilities which reflect the uncertainty of the data. (at p. 125)

This historical discussion was pointed out by Schum in his discussion paper [13].

It is of interest to take advantage of the works done by Schum on what he called 'cascade of inference' (see, for example [39] and [38]). Schum described it in the following form [13]:

> The point that initially stimulated our interest in cascaded inference was that the observation of evidence about an event is *not* diagnostically equivalent to observation of the event itself. (at p. 235)

In other words, Schum [38] pointed out that 'evidence about some event and the actual occurrence of this event are not the same' (at p. 18). So that an event, say R, represents evidence that an event E happened or is true.

> From evidence R, I must *infer* whether or not E actually happened or is true. Under a stipulation that all evidence is inconclusive to some degree, this inference can only be expressed in probabilistic terms. (at p. 18)

His work is somehow related to Jeffreys principle. Schum wrote [13]:

> The Dodson and Gettys-Willke-Jeffrey formulations concern instances in which a source of evidence gives *equivocal* testimony in the form of a probability estimate, and so we cannot be sure whether or not E occurred. But DuCharme and I were interested in the *very* many other situations in which a source give *unequivocal* testimony the E occurred but we are still uncertain about whether or not it actually did occur. This happens whenever the source is less than perfectly credible. So we began by distinguishing between testimonial evidence E^* [R in our notation] that event E occurred and event E itself. (at pp. 235-236)

The question that - from the Schum point of view - discriminates between the Gettys-Willke-Jeffrey's and his own work is 'who does the credibility assessment of witness W?' He wrote:

> In the Jeffrey situation, [a witness] W provides an assessment of his own credibility as far as his observation was concerned. He is uncertain about whether his observation was E or $\bar{E}$, and he expresses this uncertainty by means of $\Pr_1(E)$ and $\Pr_1(\bar{E})$. In [Schum's development], *we* make an assessment of the credibility of W's unequivocal testimony by means of [$\Pr(R \mid E)$ and $\Pr(\bar{R} \mid \bar{E})$]. (at p. 353)

7. Bayes' factor for unequivocal testimony

In forensic science, it is often of interest to assess the value of the evidence. The coherent metrics to assess the value of the evidence is the Bayes' factor [19], often simply called 'likelihood ratio' (though the two expressions are not, in general, equivalent and the likelihood ratio just represents a special case of Bayes' factor).

Our first interest is focused on highlighting the link between Jeffrey's solution for *equivocal* testimony which is focused on posterior probabilities, and Schum's works offering solutions for the value of *unequivocal* testimony through cascaded inference, where source inaccuracy is considered in two inferential steps (R to E, and E to H). The potential for a parallelism between the two approaches (Jeffrey's posterior probability and Schum's Bayes' factor) is discussed in [37].

Consider a reported testimony R of an event E. The current scenario allowing for unequivocal testimony, but where the credibility of the source is questioned is described in Figure 4, reporting a probability tree containing nodes H (*Hypothesis*), E (*Evidence*) and R (*Reported testimony*).

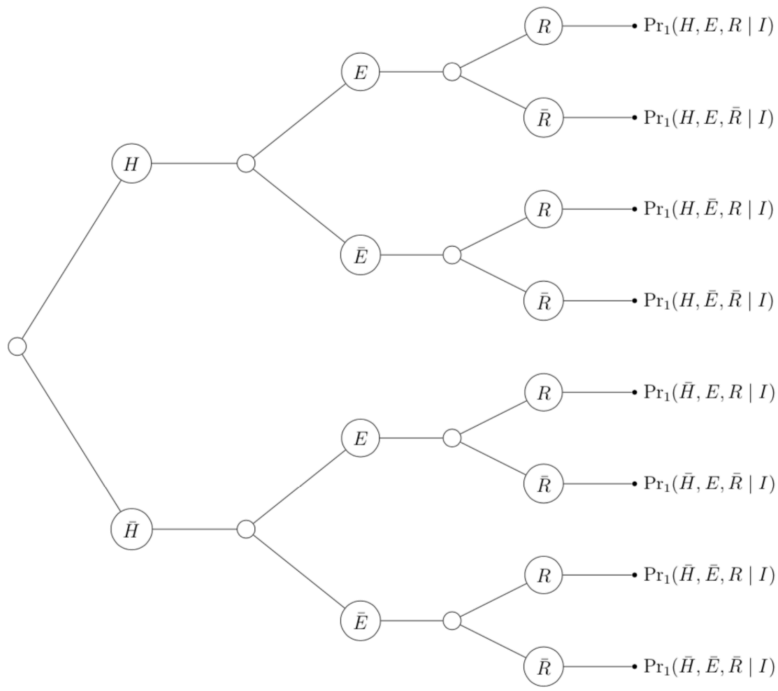

Figure 4: The probability tree for the hypothesis H, the evidence E and the reported testimony R at time t_1, given background information I.

Suppose that it is reported R, so that $\mathrm{Pr}_1(R) = 1$ (unequivocal). Taking the same arguments illustrated in Section 3, it is entirely reasonable to redistribute probabilities in such a way that the *Symmetry principle* is satisfied:

$$\mathrm{Pr}_1(H, E, R) = \frac{\mathrm{Pr}_0(H, E, R)}{\mathrm{Pr}_0(R)}.$$

So, the probability of proposition H can be obtained as

$$\begin{aligned} \mathrm{Pr}_1(H) &= \mathrm{Pr}_1(H, E, R) + \mathrm{Pr}_1(H, \bar{E}, R) \\ &= \frac{\mathrm{Pr}_0(H, E, R)}{\mathrm{Pr}_0(R)} + \frac{\mathrm{Pr}_0(H, \bar{E}, R)}{\mathrm{Pr}_0(R)} \\ &= \mathrm{Pr}_0(H \mid E, R)\mathrm{Pr}_0(E \mid R) + \mathrm{Pr}_0(H \mid \bar{E}, R)\mathrm{Pr}_0(\bar{E} \mid R) \\ &= \mathrm{Pr}_0(H \mid E)\mathrm{Pr}_0(E \mid R) + \mathrm{Pr}_0(H \mid \bar{E})\mathrm{Pr}_0(\bar{E} \mid R). \end{aligned} \tag{15}$$

Note that Equation (15) assumes that events R and H are considered as independent conditional upon E. This assumption should be carefully considered as mentioned by Schum [38] (see pp. 311-312). An example of the acceptance of such an assumption is presented in [44] dealing with errors in DNA evidence evaluation.

Note also that the background information I has been omitted for sake of simplicity, though its existence must not be forgotten. This can be generalized to any partition $E_1, \ldots, E_n$ as

$$\Pr_1(H) = \sum_{j=1}^{n} \Pr_0(H \mid E_j)\Pr_0(E_j \mid R).$$

Consider the posterior probability $\Pr_1(H)$ in (15). Recalling that $\Pr_0(H \mid E) = \Pr_0(E \mid H)\Pr_0(H)/\Pr_0(E)$, the posterior odds takes the following form

$$\frac{\Pr_1(H)}{\Pr_1(\bar{H})} = \frac{\frac{\Pr_0(H)\Pr_0(E|H)}{\Pr_0(E)}\Pr_0(E \mid R) + \frac{\Pr_0(H)\Pr_0(\bar{E}|H)}{\Pr_0(\bar{E})}\Pr_0(\bar{E} \mid R)}{\frac{\Pr_0(\bar{H})\Pr_0(E|\bar{H})}{\Pr_0(E)}\Pr_0(E \mid R) + \frac{\Pr_0(\bar{H})\Pr_0(\bar{E}|\bar{H})}{\Pr_0(\bar{E})}\Pr_0(\bar{E} \mid R)}. \tag{16}$$

Dividing (16) by the prior odds $\Pr_0(H)/\Pr_0(\bar{H})$, and after a simple manipulation, it is possible to express the Bayes' factor for the assessment of unequivocal testimony as

$$\mathrm{BF} = \frac{\Pr_0(E \mid H)\left[\Pr_0(E \mid R)\Pr_0(\bar{E}) - \Pr_0(\bar{E} \mid R)\Pr_0(E)\right] + \Pr_0(\bar{E} \mid R)\Pr_0(E)}{\Pr_0(E \mid \bar{H})\left[\Pr_0(E \mid R)\Pr_0(\bar{E}) - \Pr_0(\bar{E} \mid R)\Pr_0(E)\right] + \Pr_0(\bar{E} \mid R)\Pr_0(E)}. \tag{17}$$

The Bayes' factor in (17) quantifies the value of the acquired information that is the result of an observation reported by a scientist or witness. It is straightforward to show that it is equivalent to that proposed by Schum [38] for unequivocal evidence by adopting a cascaded inference that considers the difference between a report R, on a given event, and the event itself E. It is suffcient to reformulate $\Pr_0(E \mid R)$ as $\Pr_0(R \mid E)\Pr_0(E)/\Pr_0(R)$ and a bit of algebra to obtain

$$\mathrm{BF} = \frac{\Pr_0(E \mid H)[\Pr_0(R \mid E) - \Pr_0(R \mid \bar{E})] + \Pr_0(R \mid \bar{E})}{\Pr_0(E \mid \bar{H})[\Pr_0(R \mid E) - \Pr_0(R \mid \bar{E})] + \Pr_0(R \mid \bar{E})}. \tag{18}$$

Probabilities in Equation (18)[8] can be defined as follows:

- $\Pr_0(E \mid H)$ represents the sensibility of the analytical method;
- $\Pr_0(E \mid \bar{H})$ represents the complement of the specificity of the analytical method;
- $\Pr_0(R \mid E)$, represents the probability to detect correctly a feature when the feature does exist;
- $\Pr_0(R \mid \bar{E})$ represents the probability to detect a given feature when that feature does not exist. This can be defined as a 'false association'.

These probabilities specify the 'quality' of the performed test (points 1 and 2) and that of a given laboratory (points 3 and 4).

[8] Examples of application of such a cascaded inference can be found in, e.g. [32, 44].

Note that if one considers that (a) the laboratory is able to detect a given feature every time it is faced with such a feature, so that $\Pr_0(R \mid E) = 1$, and that (b) the laboratory is error-free, $\Pr_0(R \mid \bar{E}) = 0$, then the Bayes' factor in (18) reduces to

$$\text{BF} = \frac{\Pr_0(E \mid H)}{\Pr_0(E \mid \bar{H})}.$$

This is an extreme situation, often difficult to justify and defend in front of a Court of Justice.

Recall the medical example presented earlier in Section 5 and consider the sensibility of the test $\Pr_0(E \mid H) = 0.95$ and the specificity of the test, $\Pr_0(\bar{E} \mid \bar{H}) = 0.99$, so that $\Pr_0(E \mid \bar{H}) = 0.01$; the Bayes' factor becomes 95. If one does not accept that the laboratory is error free, one should assign values for $\Pr_0(R \mid E)$ and $\Pr_0(R \mid \bar{E})$. Suppose that it is believed that the laboratory always reports a feature when it is supposed to, i.e. $\Pr_0(R \mid E) = 1$, while a value equal to 0.04 is assigned to the probability to detect erroneously a feature, i.e. $\Pr_0(R \mid \bar{E}) = 0.04$. The Bayes' factor in (18) becomes

$$\text{BF} = \frac{0.95 \times [1 - 0.04] + 0.04}{0.01 \times [1 - 0.04] + 0.04} = 19.2.$$

Clearly, whenever the possibility of a laboratory error is accounted for (i.e., $\Pr_0(R \mid E) < 1$ and $\Pr_0(R \mid \bar{E}) > 0$), the value of the evidence will be smaller.

Information useful to assign values for $\Pr_0(R \mid E)$ and $\Pr_0(R \mid \bar{E})$ can be obtained through results of internal laboratory tests or collaborative tests (i.e. proficiency tests) regularly performed by forensic laboratories. A discussion on the quality and relevance of proficiency tests in forensic science is out of the scope of this paper. The interested reader can refer to [27, 28, 29, 30] for comments and critical discussions on the current state of affairs.

The fact that a scientist offers an unequivocal evidence, by saying that 'the test result is positive and he is 100% sure of that', does not mean that there is no uncertainty around the evidence. It is matter of fact that a Court (or a patient in a medical context) should have information on the laboratory performances, not just on the test itself to be able to assess the meaning of the expert's statement. This can be done by using Equation (18).

8. Bayes' factor for equivocal testimony

A slightly different situation is that of a scientist saying that he 'supposes that the test is positive; he is, e.g. 70% sure that the test is positive'. Jeffrey [23] expressed this possible scenario through the following examples:

> The agent inspects a piece of cloth by candlelight, and gets the impression that it is green, although he conceded that it might be blue or even (but very improbably) violet. If G, B and V are the propositions that the cloth is green, blue and violet, respectively, then the outcome of the observation might be that, whereas originally his degrees of belief in G, B and V were .30, .30, and .40, his degrees of belief in those same propositions after the observation are .70, .25, and .05. (at p. 165)

and

> It is easy enough to cite examples like 1 [the previous one] for the other senses. Transcribing a lecture in a noisy auditorium, the agent might think he had heard the word ‘red’, but still think it possible that the word was actually ‘led’. He might be unsure about whether the meat he is testing is pork or veal [...]. (at p. 166).

The idea here is to extend the Bayes’ factor for the assessment of *unequivocal* testimony to consider situations involving *equivocal* testimony.

Consider the probability tree in Figure 4, where now $\Pr_1(R) < 1$. It is possible to redistribute probabilities as

$$\begin{aligned}\Pr_1(H,E,R) &= \Pr_0(H,E,R)\times\frac{\Pr_1(R)}{\Pr_0(R)}\\ &= \Pr_0(H\mid E,R)\Pr_0(E\mid R)\Pr_0(R)\times\frac{\Pr_1(R)}{\Pr_0(R)}\\ &= \Pr_0(H\mid E)\Pr_0(E\mid R)\Pr_1(R).\end{aligned}$$

The probabilities of all other branches can be obtained analogously.

The probability of proposition H given equivocal testimony R becomes

$$\begin{aligned}\Pr_1(H) &= \Pr_1(H,E,R)+\Pr_1(H,\bar{E},R)+\Pr_1(H,E,\bar{R})+\Pr_1(H,\bar{E},\bar{R})\\ &= \Pr_1(R)\Big[\Pr_0(H\mid E)\Pr_0(E\mid R)+\Pr_0(H\mid\bar{E})\Pr_0(\bar{E}\mid R)\Big]\\ &\quad+\Pr_1(\bar{R})\Big[\Pr_0(H\mid E)\Pr_0(E\mid\bar{R})+\Pr_0(H\mid\bar{E})\Pr_0(\bar{E}\mid\bar{R})\Big]\\ &= \Pr_1(R)\Pr_0(H)\left[\frac{\Pr_0(E\mid H)}{\Pr_0(E)}\Pr_0(E\mid R)+\frac{\Pr_0(\bar{E}\mid H)}{\Pr_0(\bar{E})}\Pr_0(\bar{E}\mid R)\right]\\ &\quad+\Pr_1(\bar{R})\Pr_0(H)\left[\frac{\Pr_0(E\mid H)}{\Pr_0(E)}\Pr_0(E\mid\bar{R})+\frac{\Pr_0(\bar{E}\mid H)}{\Pr_0(\bar{E})}\Pr_0(\bar{E}\mid\bar{R})\right]\end{aligned}\tag{19}$$

The posterior probability $\Pr_1(\bar{H})$ of the alternative proposition can be obtained analogously. The posterior odds can therefore be obtained as a ratio between $\Pr_1(H)$ and $\Pr_1(\bar{H})$ (see Appendix), and the Bayes’ factor takes the following form:

$$\text{BF}=\frac{\frac{\Pr_1(R)}{\Pr_0(R)}\Big\{\Pr(E\mid H)\Big[\Pr(R\mid E)-\Pr(R\mid\bar{E})\Big]+\Pr(R\mid\bar{E})\Big\}+\frac{\Pr_1(\bar{R})}{\Pr_0(\bar{R})}\Big\{\Pr(E\mid H)\Big[\Pr(\bar{R}\mid E)-\Pr(\bar{R}\mid\bar{E})\Big]+\Pr(\bar{R}\mid\bar{E})\Big\}}{\frac{\Pr_1(R)}{\Pr_0(R)}\Big\{\Pr(E\mid\bar{H})\Big[\Pr(R\mid E)-\Pr(R\mid\bar{E})\Big]+\Pr(R\mid\bar{E})\Big\}+\frac{\Pr_1(\bar{R})}{\Pr_0(\bar{R})}\Big\{\Pr(E\mid\bar{H})\Big[\Pr(\bar{R}\mid E)-\Pr(\bar{R}\mid\bar{E})\Big]+\Pr(\bar{R}\mid\bar{E})\Big\}}\tag{20}$$

This expression simplifies to (17) when the testimony is unequivocal, i.e. $\Pr_1(R) = 1$.

Consider now the case described in Section 7, where

- $\Pr_0(E\mid H) = 0.95$,
- $\Pr_0(E\mid\bar{H}) = 0.01$,
- $\Pr_0(R\mid E) = 1$,

- $\Pr_0(R \mid \bar{E}) = 0.04$.

To compute the Bayes' factor in (20) it must be first obtained $\Pr_0(R)$. Consider, for the sake of illustration, $\Pr_0(H) = 0.8$, then

$$\begin{aligned}
\Pr_0(R) &= \Pr_0(R \mid E)\Pr_0(E) + \Pr_0(R \mid \bar{E})\Pr_0(\bar{E}) \\
&= \Pr_0(R \mid E)\left[\Pr_0(E \mid H)\Pr_0(H) + \Pr_0(E \mid \bar{H})\Pr_0(\bar{H})\right] + \Pr_0(R \mid \bar{E})\left[\Pr_0(\bar{E} \mid H)\Pr_0(H) + \Pr_0(\bar{E} \mid \bar{H})\Pr_0(\bar{H})\right] \\
&= 1 \times [(0.95 \times 0.8) + (0.01 \times 0.2)] + 0.04 \times ([0.05 \times 0.8) + (0.99 \times 0.02)] \\
&= 0.77
\end{aligned}$$

Suppose now the practitioner is uncertain about the reported testimony, and that this uncertainty is quantified in a degree of belief $\Pr_1(R) = 0.7$ as illustrated in the previous Jeffrey's quote. The Bayes' factor in (20) becomes:

$$\begin{aligned}
\text{BF} &= \frac{\frac{0.7}{0.77}[0.95 \times (1 - 0.04) + 0.04] + \frac{0.3}{0.23}[0.95 \times (0 - 0.96) + 0.96]}{\frac{0.7}{0.77}[0.01 \times (1 - 0.04) + 0.04] + \frac{0.3}{0.23}[0.01 \times (0 - 0.96) + 0.96]} \\
&= 0.717
\end{aligned}$$

Table 1: Bayes' factor BF and posterior probability $\Pr_1(H)$ for various values of $\Pr_0(H)$ and $\Pr_1(R)$, given $\Pr_0(E \mid H) = 0.95$, $\Pr_0(E \mid \bar{H}) = 0.01$, $\Pr_0(R \mid E) = 1$ and $\Pr_0(R \mid \bar{E}) = 0.04$.

$\Pr_0(H)$	$\Pr_1(R)$	BF	$\Pr_1(H)$
0.1	0.5	4.702	0.343
0.1	0.7	8.249	0.478
0.1	0.9	14.271	0.613
0.5	0.5	0.997	0.499
0.5	0.7	2.123	0.68
0.5	0.9	6.155	0.86
0.9	0.5	0.209	0.653
0.9	0.7	0.417	0.79
0.9	0.9	1.392	0.926
-	1	95	-

It must be pointed out that the Bayes' factor in (20) depends on the prior probability $\Pr_0(H)$. The Bayes' factor doesn't in fact simplify to a likelihood ratio: it represents a measure of change in support of competing propositions, rather than a measure of support.

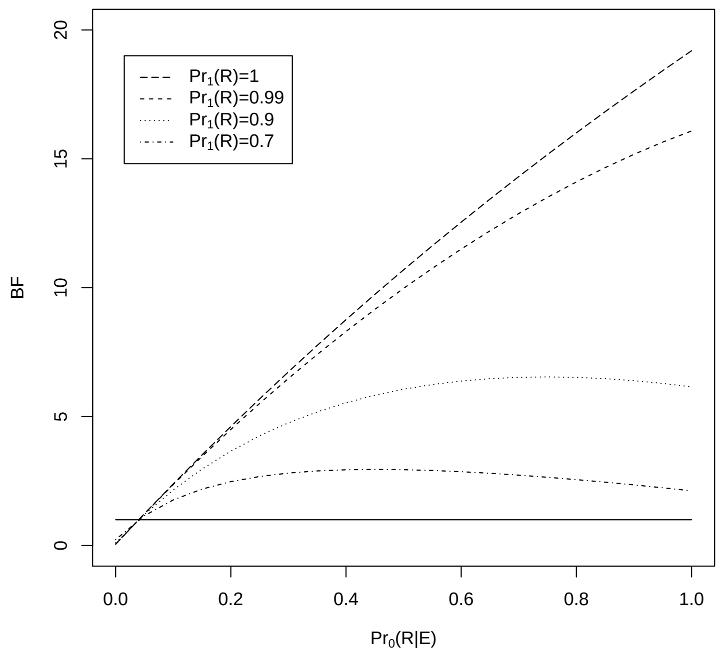

Figure 5: Bayes' factor in Equation (20) for $\Pr_0(H) = 0.5$, values of $\Pr_0(R \mid E)$ ranging from 0 until 1 in situations involving unequivocal ($\Pr_1(R) = 1$) and equivocal evidence ($\Pr_1(R) = 0.99$, $\Pr_1(R) = 0.9$ and $\Pr_1(R) = 0.7$). Note that the solid line indicates the neutral value of the Bayes' factor, BF = 1.

A small Bayes' factor doesn't mean that the probability of the hypothesis of interest H is low: it only means that available knowledge (e.g. the reported testimony) lowers the probability of the hypothesis of interest. Take the case where $\Pr_1(R) = 0.7$ and $\Pr_0(H) = 0.9$ in Table 1. The Bayes' factor in this case is 0.417, however the hypothesis H is more likely than hypothesis $\bar{H}$ ($\Pr_1(H) = 0.79$). In the same way, a large value of the Bayes' factor does not mean that the probability of the hypothesis of interest is elevated. Undoubtedly, equivocal evidence drastically reduces the value of the evidence as shown in Figure 5. In fact, the more uncertainty there is about the event R, i.e. the lower is $\Pr_1(R)$, the more the value of the Bayes' factor will decrease. Furthermore, the more uncertainty there is about the event R, the lower the impact there will be on the laboratory's ability to detect correctly a feature when that feature does exist (i.e. the value of $\Pr_0(R \mid E)$) for the calculation of the Bayes' factor.

9. Conclusion

The classical Bayesian perspective is based on conditioning related to evidence (laboratory experiments or observations in general) taken as certain. What is called 'Radical Probabilism' replaces such an extreme view by 'the conviction that probabilities need not be based on certainties.' [14] (at p. 254). Uncertainty related to the reporting does exist. An extended operational perspective can be adopted. Such an extension (the so-called Dodson-Gettys-Willke-Jeffrey formulation) can be view as a generalization of Bayes' Theorem and its use should be encouraged to deal with real life situations as supported in [45] by strongly affirming that:

> Our arguments cannot touch someone who insists that either changing opinion more or less by caprice, or only by rules which take into account factors other than those occurring explicitly in Jeffrey's problem, is no less rational than by such a rule as we have endeavored to discover. We can only say: if a general rule is to be followed for all cases falling into the scope of the problem as posed, then Jeffrey's rule as the required generality, and no other rule does. (at p. 23)

Forensic scientists deal with the value of the evidence through a measure called the Bayes' factor. A cascaded inference taking into account what has been called equivocal evidence has been introduced, theoretically justified and described through a series of examples.

In the authors' perspective, such a generalization for the the evidence evaluation should be applied.

Acknowledgment

The authors would like to thank the Swiss National Science Foundation for its support through grant n. IZSEZ0_19114.

References

[1] C G G Aitken, F Taroni, and S Bozza. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 3rd edition, 2020.

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

[3] R Bradley. Radical probabilism and Bayesian conditioning. *Philosophy of Science*, 72:342–364, 2005.

[4] F Corradi, W Pinchi, I Barsanti, and S Garatti. Probabilistic classification of age by third molar development: the use of soft evidence. *Journal of Forensic Sciences*, 58:51–59, 2013.

[5] A P Dawid and M C Galavotti. De Finetti's subjectivism, objective probability, and the empirical validation of probability assessment. In M C Galavotti, editor, *Bruno de Finetti, Radical probabilist*, pages 97–114. College Publications, London, 2009.

[6] B de Finetti. La probabilità: guardarsi dalle contraffazioni. Scientia 111 (1976) 225-281 (English translation). In H E Kyburg and H E Smokler, editors, *Studies in subjective probability*, pages 194–224. Dover Publications, Inc., New York, 2 edition, 1980.

[7] B de Finetti. Probabilità di una teoria e probabilità dei fati. In *Studi di probabilità, statistica e ricerca operativa in onore di Giuseppe Pompilj*, pages 86–101. Oderisi, Gubbio, 1989.

[8] B de Finetti. *Filosofia della probabilità*. Il Saggiatore, Milano, 1995.

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

[10] P Diaconis and S Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77:882–830, 1982.

[11] J Dodson. Simulation design for a TEAS simulation research facility. Technical Report AFCRL-1112, PRC R-194, Planning Research Corporation Report, Los Angeles, 1961.

[12] A Eagle. Updating degrees of belief - Conditionalisation and reflection: Introduction. In A Eagle, editor, *Philosophy of probability - Contemporary readings*, pages 115–131. Routledge, London, 2011.

[13] W Edwards, D A Schum, and R L Winkler. Murder and (of?) the Likelihood Principle: a trialogue. In J W Weiss and D J Weiss, editors, *A science of decision making: the legacy of Ward Edwards*, pages 234–245. Oxford University Press, Oxford, 2008.

[14] M C Galavotti. Probabilism and beyond. *Erkenntnis*, 45:253–265, 1996.

[15] P Garbolino. *I fatti e le opinioni. La moderna arte della congettura*. Editori Laterza, Roma-Bari, 1997.

[16] P Garbolino. *Probabilitá e logica della prova*. Giuffré Editore, Milano, 2014.

[17] C F Gettys and T A Willke. The application of Bayes' theorem when the true data state is uncertain. *Organizational Behavior and Human Performance*, 4:125–141, 1969.

[18] I J Good. A causal calculus. *Journal of Statistical Computation and Simulation*, 13:56–60, 1981.

[19] I J Good. Weight of evidence and a compelling metaprinciple. *Journal of Statistical Computation and Simulation*, 31:121–123, 1989.

[20] I Hacking. Slightly more realistic personal probability. *Philosophy of Science*, 34:311–325, 1967.

[21] C Howson. Bayesian rules of updating. *Erkenntnis*, 45:195–208, 1997.

[22] C Howson and P Urbach. *Scientific Reasoning: The Bayesian Approach*. Open Court, La Salle, 2 edition, 1993.

[23] R C Jeffrey. *The Logic of Decision*. University of Chicago Press, Chicago, 2 edition, 1983.

[24] R C Jeffrey. Alias Smith and Jones: the testimony of the senses. *Erkenntnis*, 26: 391–399, 1987.

[25] R C Jeffrey. *Subjective probability - The real thing*. Cambridge University Press, Cambridge, 2004.

[26] R C Jeffrey and M Hendrickson. Probabilizing pathology. *Proceedings of the Aristotelian Society*, 89:211–225, 1988-1989.

[27] J J Koehler. Fingerprint error rates and proficiency tests: what they are and why they matter. *Hastings Law Journal*, 59:1077–1099, 2008.

[28] J J Koehler. Proficiency tests to estimate error rates in the forensic sciences. *Law, Probability & Risk*, 12:89–98, 2013.

[29] J J Koehler. Forensics or Fauxrensics? ascertaining accuracy in the forensic sciences. *Arizona State Law Journal*, 49:1369–1416, 2017.

[30] J J Koehler. Intuitive error rate estimates for the forensic sciences. *Jurimetrics Journal*, 57:153–168, 2017.

[31] D V Lindley. Probability. In C G G Aitken and D A Stoney, editors, *The Use of Statistics in Forensic Science*, pages 27–50. Ellis Horwood, New York, 1991.

[32] D V Lindley and R Eggleston. The problem of missing evidence. *The Law Quaterly Review*, 99:86–99, 1983.

[33] P Maher. *Betting on theories*. Cambridge University Press, Cambridge, 1993.

[34] S Morini. *Probabilismo - Storia e teoria*. Bruno Mondadori Editori, Milano, 2003.

[35] A O'Hagan. *Kendall's Advanced Theory of Statistics Vol 2b: Bayesian inference*. Arnold, 1994.

[36] 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.

[37] R E Schaefer and K Borcherding. A note on the consistency between two approaches to incorporate data from unreliable sources and Bayesian analysis. *Organizational Behavior and Human Performance*, 9:504–508, 1973.

[38] D A Schum. *Evidential Foundations of Probabilistic Reasoning*. John Wiley & Sons, Inc., New York, 1994.

[39] D A Schum and W DuCharme. Comments on the relationship between the impact and the reliability of evidence. *Organizational Behavior and Human Performance*, 6:111–131, 1971.

[40] G Shafer. *The art of causal conjecture*. The MIT Press, Cambridge, Mass., 1996.

[41] B Skyrms. Dynamic coherence and probability kinematics. *Philosophy of Science*, 54:1–20, 1987.

[42] F Taroni, S Bozza, A Biedermann, and C G G Aitken. Dismissal of the illusion of uncertainty in the assessment of a likelihood ratio. *Law, Probability and Risk*, 15:1–16, 2016.

[43] F Taroni, P Garbolino, A Biedermann, C G G Aitken, and S Bozza. Reconciliation of subjective probabilities and frequencies in forensic science. *Law, Probability and Risk*, 17:243–262, 2018.

[44] W C Thompson, F Taroni, and C G G Aitken. How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Sciences*, 48: 47–54, 2003.

[45] B C van Fraassen. A demonstration of Jeffrey conditionalization rule. *Erkenntnis*, 24:17–24, 1986.

Appendix

Consider the posterior probability $\Pr_1(H)$ in (19). The posterior probability $\Pr_1(\bar{H})$ can be obtained analogously, and the posterior odds takes the following form:

$$\frac{\Pr_1(H)}{\Pr_1(\bar{H})} = \frac{\begin{array}{c}\Pr_1(R)\Pr_0(H)\left[\Pr(E \mid H)\frac{\Pr(R|E)\Pr(E)}{\Pr_0(R)}\Pr(\bar{E}) + \Pr(\bar{E} \mid H)\frac{\Pr(R|\bar{E})\Pr(\bar{E})}{\Pr_0(R)}\Pr(E)\right] + \\ \Pr_1(\bar{R})\Pr_0(H)\left[\Pr(E \mid H)\frac{\Pr(\bar{R}|E)\Pr(E)}{\Pr_0(\bar{R})}\Pr(\bar{E}) + \Pr(\bar{E} \mid H)\frac{\Pr(\bar{R}|\bar{E})\Pr(\bar{E})}{\Pr_0(\bar{R})}\Pr(E)\right]\end{array}}{\begin{array}{c}\Pr_1(R)\Pr_0(\bar{H})\left[\Pr(E \mid \bar{H})\frac{\Pr(R|E)\Pr(E)}{\Pr_0(R)}\Pr(\bar{E}) + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(R|\bar{E})\Pr(\bar{E})}{\Pr_0(R)}\Pr(E)\right] + \\ \Pr_1(\bar{R})\Pr_0(\bar{H})\left[\Pr(E \mid \bar{H})\frac{\Pr(\bar{R}|E)\Pr(E)}{\Pr_0(\bar{R})}\Pr(\bar{E}) + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(\bar{R}|\bar{E})\Pr(\bar{E})}{\Pr_0(\bar{R})}\Pr(E)\right].\end{array}}$$

After some algebra, the posterior odds can be rewritten as:

$$\frac{\Pr_1(H)}{\Pr_1(\bar{H})} = \frac{\Pr_1(R)\Pr_0(H)\left[\Pr(E \mid H)\frac{\Pr(R|E)}{\Pr_0(R)} + \Pr(\bar{E} \mid H)\frac{\Pr(R|\bar{E})}{\Pr_0(R)}\right] + \Pr_1(\bar{R})\Pr_0(H)\left[\Pr(E \mid H)\frac{\Pr(\bar{R}|E)}{\Pr_0(\bar{R})} + \Pr(\bar{E} \mid H)\frac{\Pr(\bar{R}|\bar{E})}{\Pr_0(\bar{R})}\right]}{\Pr_1(R)\Pr_0(\bar{H})\left[\Pr(E \mid \bar{H})\frac{\Pr(R|E)}{\Pr_0(R)} + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(R|\bar{E})}{\Pr_0(R)}\right] + \Pr_1(\bar{R})\Pr_0(\bar{H})\left[\Pr(E \mid \bar{H})\frac{\Pr(\bar{R}|E)}{\Pr_0(\bar{R})} + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(\bar{R}|\bar{E})}{\Pr_0(\bar{R})}\right]}.$$

2 The Bayes factor can therefore be obtained dividing the posterior odds by the prior odds:

$$\begin{aligned} \text{BF} &= \frac{\Pr_1(R)\left[\Pr(E \mid H)\frac{\Pr(R|E)}{\Pr_0(R)} + \Pr(\bar{E} \mid H)\frac{\Pr(R|\bar{E})}{\Pr_0(R)}\right] + \Pr_1(\bar{R})\left[\Pr(E \mid H)\frac{\Pr(\bar{R}|E)}{\Pr_0(\bar{R})} + \Pr(\bar{E} \mid H)\frac{\Pr(\bar{R}|\bar{E})}{\Pr_0(\bar{R})}\right]}{\Pr_1(R)\left[\Pr(E \mid \bar{H})\frac{\Pr(R|E)}{\Pr_0(R)} + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(R|\bar{E})}{\Pr_0(R)}\right] + \Pr_1(\bar{R})\left[\Pr(E \mid \bar{H})\frac{\Pr(\bar{R}|E)}{\Pr_0(\bar{R})} + \Pr(\bar{E} \mid \bar{H})\frac{\Pr(\bar{R}|\bar{E})}{\Pr_0(\bar{R})}\right]} \\ &= \frac{\frac{\Pr_1(R)}{\Pr_0(R)}\left\{\Pr(E \mid H)\left[\Pr(R \mid E) - \Pr(R \mid \bar{E})\right] + \Pr(R \mid \bar{E})\right\} + \frac{\Pr_1(\bar{R})}{\Pr_0(\bar{R})}\left\{\Pr(E \mid H)\left[\Pr(\bar{R} \mid E) - \Pr(\bar{R} \mid \bar{E})\right] + \Pr(R \mid \bar{E})\right\}}{\frac{\Pr_1(R)}{\Pr_0(R)}\left\{\Pr(E \mid \bar{H})\left[\Pr(R \mid E) - \Pr(R \mid \bar{E})\right] + \Pr(R \mid \bar{E})\right\} + \frac{\Pr_1(\bar{R})}{\Pr_0(\bar{R})}\left\{\Pr(E \mid \bar{H})\left[\Pr(\bar{R} \mid E) - \Pr(\bar{R} \mid \bar{E})\right] + \Pr(R \mid \bar{E})\right\}}. \end{aligned}$$