The Bayes' factor: the coherent measure for hypothesis confirmation

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

Paolo Garbolino,
Department of Art and Architecture,
University IUAV of Venice,
Italy

Silvia Bozza,
Department of Economics,
Ca' Foscari University of Venice,
Italy

Colin Aitken,
School of Mathematics and Maxwell Institute,
University of Edinburgh,
United Kingdom

November 6, 2021

Abstract

What have been called 'Bayesian confirmation measures' or 'evidential support measures' offer a numerical expression for the impact of a piece of evidence on a judicial hypothesis of interest. The Bayes' factor, sometimes simply called the 'likelihood ratio', represents the best measure of the value of the evidence. It satisfies a number of necessary conditions on normative logical adequacy. It is shown that the same cannot be said for alternative expressions put forward by some legal and forensic quarters. A list of desiderata are given that support the choice of the Bayes' factor as the best measure for quantification of the value of evidence.

Keywords

Bayesian confirmation theory, Bayes' factor, Likelihood ratio, Logical properties, Philosophy of science, Probabilistic reasoning, Scientific evidence evaluation.

1 Introduction

Uncertainty is a complication that accompanies actors of the justice system who face inference and decision-making as core aspects of their activities. Inference and decision-making require logical assistance because unaided human reasoning is liable to bias. Bias represents a critical cause for concern because fallacious reasoning and erroneous conclusions in legal proceedings put defendants at risk and can lead to miscarriages of justice.

Across legal systems, numerous courts have repeatedly highlighted that practising forensic scientists are continually required to assess their domain of expertise (Murphy [2017]), to scrutinize both the rationale underlying the various domains and the methods that are pursued for the evaluation and presentation of scientific evidence. An acknowledgement that incomplete knowledge inevitably results in uncertainty (which is, unfortunately, the regular state of affairs) means that inferences must be approached within a probabilistic framework (e.g. Tillers [2011], Saks and Neufeld [2011]), and therefore that the value of evidence be expressed in numerical terms. This means that the verbal opinions of the expert witness on the strength of the evidence no longer represent a justifiable approach to

the presentation of evidence (see Giannelli et al. [2011])[1]. Today, therefore, values based on scrutinized statistical or probabilistic approaches are increasingly used. Scientific and judicial literature have clearly pointed out that such approaches should be justified on logical and scientific grounds (e.g. European Network of Forensic Science Institutes [2015], Thompson et al. [2018]).

It is customary to read papers published in scientific and legal journals that attack the role of Bayes' theorem for probabilistic reasoning in the evaluation of evidence and the use of the so-called 'likelihood ratio' for the assessment of the value of the evidence to which a scientist reports in front of a Court of Justice (some examples are presented in Kaye and Sensabaugh [2011][2].) These attacks are often accompanied by claims for the benefits of other expressions for evidence evaluation. These criticisms are weak and the alternatives introduced to express the value or the weight of evidence are neither coherent nor justified. Alternative proposals have major drawbacks and illogical conclusions, typically based on intuitive ideas that ignore basic probabilistic principles. Examples of these are the use of verbal expressions such as 'consistent with' and 'could have come from' that misrepresent the evidence, and the use of relative frequencies (used to quantify probabilities of interest) that generally are prejudicial because they might be transposed[3] (see Aitken et al. [2021] for a discussion on these topics and Kaye [2021] for a description of the use of statistics in the courtroom). The main question still remains - as presented in Kaye and Sensabaugh [2011] - to what extent will 'the presentation assist the jury in understanding the meaning of a match so that the jury can give the evidence the weight that it deserves?' (at p. 167).

Recently, Buckleton et al. [2020], listed a series of characteristics supporting the use of the likelihood ratio as the measure of the value of the evidence; the aim was to respond to criticisms of Stiffelman [2019]. They reported that (a) the likelihood ratio does not infringe on the ultimate issue (i.e. it does not express an opinion on the hypothesis of judicial interest), (b) the Bayesian approach clearly separates the role of the scientist from that of the decision makers (i.e. the Court) so that the scientist is distanced from comment on the hypotheses put forward by parties at trial, (c) the likelihood ratio does not affect the reasonable doubt standard and it does not infringe on the presumption of innocence, (d) hypotheses need just be exhaustive within the context of the case, and finally (e) the likelihood ratio can be easily deduced from the ratio between posterior odds and prior odds, so that posterior odds are obtained through the multiplication of the prior odds by the likelihood ratio.

The justification of the use of Bayes' theorem in a forensic context has been provided in the past by Finkelstein and Fairley [1970] and Lempert [1977][4].

---

[1] The authors wrote about 'inconsistent statements' that place limitations on testimonies (at p. 121).

[2] The authors reported criticisms expressed by others by affirming that it appears that 'The major objection to likelihoods is not statistical but psychological' (at p. 173).

[3] One of the common mistakes (if an approach using the likelihood ratio is not followed) is to transpose the probabilities for evidence and the proposition. It may be that it is very unlikely that the evidence will be found in association with an innocent person. Consider evidence that mineral traces found at a crime scene correspond in chemical profile to mineral traces found on clothing of the defendant. This chemical profile is shown to be rare. This may be thought to be strong evidence that the defendant (or, at least, their clothing) was present at the crime scene. An example to show that the evidence may not be strong was given by Darroch [1987]. Consider a town in which a rape has been committed. There are 10,000 men of suitable age in the town of whom 200 work underground at a mine. Evidence is found at the crime scene from which it is determined that the criminal is one of the 200 mineworkers. Such evidence may be traces of minerals which could only have come from the mine. A person of interest is identified and traces of minerals, similar to those found at the crime scene, are found on some of his clothing. The evidence to be assessed is that 'mineral traces have been found on clothing of the person of interest which is similar to mineral traces found at the crime scene'. The prosecution proposition is that the person of interest is guilty. The defence proposition is that he is innocent. Assume that all people working underground at the mine will have mineral traces similar to those found at the crime scene on some of their clothing. This assumption is open to question but the point about conditional probabilities will still be valid. The probability of finding the evidence on an innocent person may then be determined as follows. There are 9,999 innocent men in the town of whom 199 work underground at the mine. These 199 men will, as a result of their work, have this evidence on their clothing, under the above assumption. Thus, the probability of finding the evidence of mineral traces on an innocent person is $199/9,999 \simeq 200/10,000 = 0.02$, a small number. However, this does not imply that the probability of innocence of a man who is found to have the evidence on him is 0.02. There are 200 men in the town who can be expected to have the evidence (mineral traces) on them. Of these 200 men, 199 are innocent. Thus, the probability a person on whom the evidence is found is innocent is $199/200 = 0.995$. The fallacious equation of the probability of finding evidence on an innocent person with the probability of innocence for a person on whom the evidence is found is known as the fallacy of the transposed conditional (Diaconis and Freedman [1981], Thompson and Schumann [1987], Evett [1995]).

[4] Arguments for the use of the likelihood ratio as a measure for the value of evidence can also be found in Evett and Weir [1998], Gittelson et al. [2018] and in the Guidelines principles developed under the programme 'Probability and Statistics in Forensic Science' (2017) (Report available at: www.newton.ac.uk/files/preprints/ni16061.pdf.)

It is unlikely the position of Buckleton et al. [2020] will be the end of this scientific discussion. Our aim is to support their initiative with the introduction of a list of *normative logical desiderata* that formally justify the adequacy of the Bayes' factor as the measure for the value of evidence. The arguments are extended to consideration of the whole case and the entirety of the evidence with no suggestion that the forensic scientist consider the whole case and the entirety of the evidence, an improper extension of their role. The extension is justifiable, however, when applied to the finders of fact, the judge and jury. For their consideration of the whole case and the entirety of the evidence, it is shown that the best way for this to be done is for there to be consideration of the relative values of the probabilities of the evidence conditional on the propositions of the prosecution and the defence. In forensic science, the 'Bayes' factor' is more commonly called 'likelihood ratio', even when it is known that a Bayes' factor does not always simplify to a likelihood ratio[5]. In the rest of the paper, and without loss of generality, the two terms will be treated as synonymous.

Concerning the *normative logical desiderata*, and using words from Crupi et al. [2013], it can be said that in

> [s]eeking theoretical clarification, a natural goal is to axiomatize [families of measures] i.e. to identify conditions that are necessary and sufficient to single out each of them as capturing a target notion. (at p. 191)

To do this, it is fundamental to recognize that the connection between evidence and a series of competing propositions is characterized by uncertainty that cannot be eliminated but can be measured by probabilities. The theory of inductive reasoning provides a strong foundation for the relationship between evidence and hypotheses. The impact of an item of evidence on the credibility of a hypothesis can be studied through what is known as *probabilistic confirmation theory* [Maher, 1996]. This paper introduces this theory to a forensic and judicial audience and develops it in order to provide a series of logical requirements that justify consideration of Bayes' factor (and functions of it) as a coherent measure of confirmation. As noted in Crupi and Tentori [2016]

> [...] confirmation has to do with how evidence affects the credibility of a hypothesis, an issue that is crucial to human reasoning in a variety of domains, from scientific inquiry to medical diagnoses, legal argumentation, and beyond. (at p. 650)

This quotation could open a good discussion on the foundation of forensic science and on how current research in forensic science is often restricted to purely technical aspects, both analytical and statistical, but this is a discussion that is beyond the scope of this paper. Researchers should however take advantage of all disciplines that can facilitate the understanding of a problem and hence indicate solutions and lines of reasoning. The philosophy of science should be of great interest for forensic scientists. Unfortunately, reference to it in the relevant literature is rare. It is to be hoped that, in addition to a rational justification of the use of the Bayes' factor as a measure of the value of evidence, this article will explain how the domain of the philosophy of science can provide key elements with reference to the evaluation and interpretation of evidence. The progressive weakening of the relationship between scientific and sociological disciplines might lessen the chance of valuable arguments advancing toward a deeper understanding of this subject. This article aims to shed light on the fundamental role that can be played by arguments from the domain of the philosophy of science with reference to the problems of evaluation and interpretation of the value of evidence. The suggested multidisciplinary approach is not new. A well-known example is that of the important articles written by the probabilist Bruno de Finetti (see, e.g. de Finetti [1930], de Finetti [1931], de Finetti [1968])[6] where statistical ideas were shared with philosophers in an open discussion and collaboration[7].

---

[5]For a comment on this point, please refer to Aitken et al. [2021].

[6]Note that de Finetti [1931] was reprinted in philosophical journal, see de Finetti [1931] and de Finetti [1989].

[7]In de Finetti [1970], the author quotes the relationship between statistics and philosophical interests: 'Various other questions [...] are currently objects of discussion in various places: for instance, the relationships between possibility and tautology seems to be attracting the attention of philosophers (the intervention of Hacking at the recent meeting, Chicago 1967); while the critical questions about the mathematical axioms of calculus of probability [...] are always a subject of debate.' (at p. 15), but he also strongly contrasted

The paper is structured as follows: Section 2 offers the general framework of hypothesis confirmation. Some basic logical requirements for confirmation, such as the requirements of compatibility, increase, formality and classificatory, are introduced in Section 3. The list of logical requirements to justify the use of the Bayes' factor for evidence evaluation is developed in Section 4. The paper ends with a conclusion in Section 5.

2 Bayesian hypothesis confirmation

This Section introduces, first, the motivation of a set of requirements on evidential support in an intuitively compelling way. Then, it introduces standard forensic and legal notation and defines - from a qualitative point of view - confirmation structure.

At the beginning of a criminal trial with a defendant, the trier of fact, judge or jury, starts with the philosophical view that the defendant is 'innocent until proven guilty' of some charge. At the end of the trial, the trier of fact either believes the case against the defendant has been 'proven beyond reasonable doubt' (or some other phrase with the same meaning) in which case the defendant is found 'guilty' of the charge or the case has not been proven beyond reasonable doubt in which case the defendant is found 'not guilty' (or, in Scotland, there is a third verdict that the case is 'not proven'). Between the beginning and the end of the trial, evidence is led by the prosecution and by the defence. For the purposes of the discussion here, the totality of the evidence is divided into what will be called 'items' each item corresponding to the testimony of a particular witness. The total number of items, the totality of the evidence, is thus taken to be equal to the total number of witnesses, prosecution and defence combined. The totality of the evidence is considered by the trier of fact. There are two questions the trier of fact should consider:

- How does a particular item of evidence affect their belief in the truth of the charge?

- How does a particular item of evidence interact with other items of evidence to affect their belief in the truth of the charge?

The first element of the construction of a coherent measure for hypothesis confirmation is the requirement for the existence of a hypothesis and then its definition. The hypothesis for which a measure for its confirmation is required is the charge against the defendant. For example, the defendant may have been charged with the murder of an individual, X say. The corresponding hypothesis would then be 'The defendant murdered X'. It is fundamental to the argument made in this paper that there is an alternative hypothesis about the case against the defendant and that it is associated with the defence to guarantee the balanced approach of justice. For example, the defence hypothesis could simply be 'The defendant did not murder X' or 'The defendant did kill X but the act was in self-defence'. The defence hypothesis may not be stated explicitly. In many cases it will simply be unstated. In such a situation, it is considered for the argument here that it is the complement of the prosecution's case.

some philosophical attitude; he wrote: 'Much more serious is the reluctance to abandon the inveterate tendency of savages to objectivize and mythologize everything; a tendency that, unfortunately, has been, and is, favoured by many more philosophers than have struggled to free us from it.' (at p. 22). De Finetti's works were honored by various philosophers focusing on probabilism: see, e.g. van Fraassen [1989] who wrote '[Pascal, Bayes, the Bernoullis, Jevons, De Morgan, Ramsey, de Finetti] [. . .] philosophy mainly ignored it as mathematical gamesmanship or materialistic technology of the mind. It is neither. Pascal and those who followed him showed us how to reconceive all the problems of epistemiology.' (at p. 153). Other examples are the praiseworthy references to de Finetti's works on the foundation of probability. See, e.g. Galavotti [1996] and von Plato [1989]. In order to highlight the link between statistics and philosophy, the words of H.E. Kyburg and H.E. Smokler in the preface of the second edition of Studies in Subjective Probability. R.E Krieger Publishing Company, Huntington, New York (1980) are very relevant: 'In the fifteen years since the first edition of Studies in Subjective Probability appeared, the point of view represented by de Finetti, Ramsey and Savage has become better known not only in philosophy and statistics where it originated [. . .].'. This link between philosophy and statistics is also put forward by Galavotti and Jeffrey [1989] in their Preface: '[. . .] in de Finetti's view technical and philosophical aspects of probability are strictly intertwined. De Finetti, the mathematical probabilist, is not to be separated from de Finetti, the philosopher of probability.' (at p. 165). Examples of the role played by de Finetti in the philosophical research programmes have been put forward in the prefaces written by philosophers of science to the Italian editions of his books, e.g. Giordano Bruno and Giulio Giorello wrote the preface to *L'invenzione della verit* and Marco Mondadori wrote that of *La logica dell'incerto*.

The two questions that have been listed above concern the effects of items of evidence on beliefs. There can only be said to be an effect on a person's belief if there is a change in the belief. Detection of a change requires a method of measurement for change.

The second element of the construction of a coherent measure for hypothesis confirmation is the requirement for a method of measurement for change. From the beginning to the end of a trial there is uncertainty. At the beginning there is initial uncertainty about the guilt or otherwise of the defendant. At the end there is final uncertainty, there is uncertainty if the defendant has been found guilty or if they have been found not guilty. At the beginning, there will be much uncertainty, at the end it is hoped there will be little uncertainty. There is a change in the degree of uncertainty. As with changes in effects, change in the degree of uncertainty requires a method of measurement.

The third element of the construction of a coherent measure for hypothesis confirmation is the requirement for a method of measurement for uncertainty. It will be shown that this is the last element necessary for the coherent measure sought.

Thus, there are three requirements for evidential support: (1) the existence of a hypothesis, with an alternative, (2) a method for the measurement of change in the degree of uncertainty of belief in a hypothesis, and (3) a method for the measurement of uncertainty.

Fortunately, there is a well-established measure for uncertainty - as stated in Section 1 - and that is *probability*. It is also well-established that this is the best measure of uncertainty (Lindley [1982]). Measures are represented by numbers and satisfy the laws of probability. Thus probability may be represented by a number and probabilities may be added, subtracted, multiplied and divided. Care has to be taken when doing so if one wishes the result of the mathematical operation to be a probability. One of the axioms of probability is that the values it may take lie between 0 and 1, inclusive. An event which is impossible has probability 0, an event which is certain has probability 1. Probability may also be used as a measure of belief. For example, the strength of one's belief in the victory of a certain football team in a local derby match may be represented by a number between 0 and 1, the closer it is to 1, the stronger the belief in victory.

Both the uncertainty associated with the truth of the proposition in a criminal trial and in the evidence presented in the trial may be represented by probability. There is then a need to show how these probabilities interact and provide a coherent measure for evidential support. Consider a betting game or sport (such as a horse race) with a set of exclusive outcomes (such as winners of a horse race with no unusual circumstances such as a tie for first place). A bookmaker offers a set of odds on each horse to be a winner[8]. A measure of support (e.g. support as measured by the probability of winning the race) is said to be *coherent* if the set of probabilities for all the possible outcomes (as represented by odds of winning the race) satisfy the rules of probability, e.g. the sum of all the probabilities should equal 1 (e.g., the sum of the probabilities of victory for each horse should add up to 1). If there is any other outcome for the sum of probabilities then the set of probabilities is said to be *incoherent*.

At the beginning of a trial, before any evidence has been led, the trier of fact has a belief (innocent until proven guilty) about the truth of the prosecution hypothesis, the strength of which may be represented by a probability. Hopefully, this is close to zero[9]. One possibility is to think the defendant is as likely (probable) to have committed the crime as anyone else. At the end of the trial, the trier's belief about the truth of the prosecution's proposition should have changed. The probability of the truth of the proposition is different from what it was at the beginning. If the defendant is found guilty, it is to be hoped this probability is very high. If it is not very high, then the defendant should be found 'not guilty'.

The best process by which the trier of fact moves from an initial belief to a final belief is the use of Bayes' factor or its logarithm. The use of the logarithm of the Bayes' factor is more intuitively satisfying than the use of Bayes' factor (see Section 3). The use of the logarithm is additive. The process of change of belief using logarithms from

---

[8]Odds and probability are interchangeable. For example, odds of 2 to 1 to win are equivalent to a probability of 2/3 of winning.
[9]This probability cannot logically be zero, otherwise, formally, no belief update will be possible.

before the leading of evidence to the delivery of the verdict is one of addition, with care in the consideration of any one item of evidence to allow for the effects of previous items of evidence. This is not an easy process.

The purpose of this paper is therefore threefold: (i) to explain the use of Bayes' factor, (ii) to explain the failings of other suggestions for the assessment of evidence, and (iii) to explain coherence and how Bayes' factor and functions of it satisfy coherence. The first step to achieve this purpose is to describe how an inference may be made that an item of evidence confirms (or supports), disconfirms (or undermines), or is neutral with respect to a given hypothesis.

Let E denote the evidence and let H denote a hypothesis (or proposition) of legal interest. The negation of E and H are denoted by $\bar{E}$ and $\bar{H}$, respectively.

It is well known to readers of judicial and forensic science journals that Bayesian reasoning proceeds as follows. Before an item of information - generically called finding, evidence or observation - is collected (or is known by the person in charge of the inferential reasoning), initial (prior) probabilities are assigned to each of a set of hypotheses of interest, given the knowledge, denoted I, collected until the probabilistic assignment is being made. The prior probability of the hypothesis of interest H can be formalized as $\Pr(H \mid I)$.

After acquiring a new item E of information, the prior probabilities assigned to the hypotheses are revised in the knowledge of E. This item of information could be scientific, such as features describing a recovered stain or mark, or non-scientific, such as eyewitness testimony. The probability of the hypothesis of interest H updated with the new information is called the posterior probability and it can be formalized as $\Pr(H \mid E, I)$. The transition from the prior to posterior probability is governed by Bayes' theorem, which enables the update of the probability quantifying the current state of uncertainty related to the hypothesis H of interest as new information become available.

Bayes' theorem is a natural consequence of the third law of probability. Mathematically it is not controversial (see, e.g. Salmon [1966]). Given the current notation, the posterior probability of a hypothesis H can be computed according to Bayes' theorem as:

$$\Pr(H \mid E, I) = \frac{\Pr(H \mid I)\Pr(E \mid H, I)}{\Pr(E \mid I)}. \tag{1}$$

Bayes' theorem can be expressed in an odds form by computing the ratio between $\Pr(H \mid E, I)$ and $\Pr(\bar{H} \mid E, I)$. The odds form of Bayes' theorem is as follows:

$$\frac{\Pr(H \mid E, I)}{\Pr(\bar{H} \mid E, I)} = \frac{\Pr(H \mid I)}{\Pr(\bar{H} \mid I)} \times \frac{\Pr(E \mid H, I)}{\Pr(E \mid \bar{H}, I)}, \tag{2}$$

where the posterior odds (the ratio between the posterior probabilities) equals prior odds (the ratio between the prior probabilities) times the likelihood ratio, a measure that quantifies the value of the evidence, $\Pr(E \mid H, I)/\Pr(E \mid \bar{H}, I)$ (Aitken and Taroni [2004]).

The reasoning of Bayes' theorem can be expressed, more generally, as follows: let $\Pr_0(\cdot)$ be the probability function that represents the opinion of a given individual, say X on a set $\mathbf{H}$ of hypotheses $H_i\,(i = 1, 2, ..., n)$, at a particular time t. The information acquired by individual X in the time interval $[t, t + k]\,(k > 0)$ is denoted E and the probability function that represents the opinion of X on $\mathbf{H}$ at the instant $t + k$ is denoted $\Pr_1(\cdot)$. The rational change of the opinion on $\mathbf{H}$ by X from the initial state $\Pr_0(\cdot)$ to the final state $\Pr_1(\cdot)$ is equivalent to the satisfaction in the passage from $\Pr_0(\cdot)$ to $\Pr_1(\cdot)$, in the light of the acquisition of information E, of the principles of what is known as *probability kinematics* providing a general updating rule that allows for uncertainty in the reported evidence (see Jeffrey [1983] and Taroni et al. [2020]).

Given a single hypothesis H, the prior probability $\Pr_0(H)$ is updated with information E to give a posterior probability $\Pr_1(H)$. There are three possibilities for the relative values of $\Pr_0(H)$ and $\Pr_1(H)$, and to which verbal descriptions are attached:

1a. E is said to *confirm* or *support* H if and only if $\mathrm{Pr}_1(H) > \mathrm{Pr}_0(H)$;

2a. E is said to be *neutral* with respect to H if and only if $\mathrm{Pr}_1(H) = \mathrm{Pr}_0(H)$;

3a. E is said to *disconfirm* or *undermine* H if and only if $\mathrm{Pr}_1(H) < \mathrm{Pr}_0(H)$.

If E is known for certain, then the posterior state of knowledge $\mathrm{Pr}_1(\cdot)$ of the individual X is given, based on the principle of conditionalization (Maher [1993], Eagle [2011], Taroni et al. [2020]), by a simple application of Bayes' theorem as $\mathrm{Pr}_1(H) = \mathrm{Pr}(H \mid E, I)$, so that it provides a qualitative response to the question whether a piece of evidence E confirms, disconfirms, or is neutral with respect to the hypothesis of interest H:

1b. E confirms or supports H if and only if $\mathrm{Pr}(H \mid E, I) > \mathrm{Pr}(H \mid I)$;

2b. E is neutral with respect to H if and only if $\mathrm{Pr}(H \mid E, I) = \mathrm{Pr}(H \mid I)$;

3b. E disconfirms or undermines H if and only if $\mathrm{Pr}(H \mid E, I) < \mathrm{Pr}(H \mid I)$.

Relationships 1b., 2b. and 3b. can also be formulated in terms of odds. Consider 1b. for the sake of illustration. If $\mathrm{Pr}(H \mid E, I) > \mathrm{Pr}(H \mid I)$, then it can be verified that $\mathrm{Pr}(H \mid E, I)/\mathrm{Pr}(\bar{H} \mid E, I) > \mathrm{Pr}(H \mid I)/\mathrm{Pr}(\bar{H} \mid I)$. Relationships 2b. and 3b. can be reformulated analogously.

The relationship of the posterior probability of the hypothesis H to its prior probability based on information E depends on the hypothesis H, the information E, the background knowledge I and the initial state of knowledge quantified by $\mathrm{Pr}(H \mid I)$.

Bayes' theorem constitutes a logical scheme to understand how an item of information E supports or undermines given hypotheses. Conceptually, Jeffrey (Jeffrey [1975]) has noted that

> Bayesianism does not take the task of scientific methodology to be that of establishing the truth of scientific hypotheses, but to be that of confirming or disconfirming them to degrees which reflect the overall effect of the available evidence, positive, negative, or neutral, as the case may be. (at p. 104)

Jeffrey's statement, through reference to 'degree', indirectly introduces the need for a quantitative part of the process.

3 Degree of confirmation and the basic *compatibility*, *formality* and *classificatory* requirements

An appropriate measure $c(E, H)$ of the degree of confirmation[10] that a hypothesis H receives from information E, is specified as one that quantifies the change in belief (or belief update) of H as introduced as motivation in Section 2. Such a measure c does not initially need to be either a probability or a function of a probability, and a question of interest is whether some appropriate function of probability can be such a measure of confirmation.

To answer this question, the approach taken by philosophers of science has been to formulate some intuitively reasonable requirements for a quantitative confirmation measure c and to show that these requirements are satisfied by probability.

First of all, it is desirable that the quantitative notion $c(E, H)$ be compatible with the qualitative notions of 'confirmation', 'neutrality' and 'disconfirmation' that have been expressed in probability terms in the previous definitions (see Section 2). A formulation of this requirement (call it the *compatibility* requirement) for $c(E, H)$ is as follows [Festa, 1996]: given sets of hypotheses H, H_0, H_{00} and information E, E_0, E_{00}

---

[10]The following language is that of confirmation theory. For those unfamiliar with that language, it may help if the term 'confirm' is substituted with the term 'support' and the term 'disconfirm' is substituted with the term 'refute'.

1. if E confirms H, E_0 is neutral with respect to H_0 and E_{00} disconfirms H_{00}, then $c(E, H) > c(E_0, H_0) > c(E_{00}, H_{00})$;

2. if E is neutral with respect to H, E_0 is neutral with respect to H_0 and E_{00} is neutral with respect to H_{00}, then $c(E, H) = c(E_0, H_0) = c(E_{00}, H_{00})$.

For example, consider evidence E that a DNA profile from a person of interest that matches, in some sense, that of a crime stain, and a source level proposition H that the person of interest is the source of the crime stain. Then E may be said to confirm H.

Alternatively, consider evidence F that a DNA profile from a person of interest does not match that of a crime stain, and the same source level proposition H as before that the person of interest is the source of the crime stain. Then F may be said to disconfirm H and $c(E, H) > c(F, H)$.

A reasonable assumption is that the confirmation measure depends solely on the degrees of beliefs about the two events of interest in the case. A confirmation measure $c(E, H)$ is said to be *formal* if it satisfies the *formality* requirement [Tentori et al., 2007b], which says that it depends only on the probability values concerning E and H: $\Pr(E, H), \Pr(E, \bar{H}), \Pr(\bar{E}, H)$ and $\Pr(\bar{E}, \bar{H})$. The dependence on background information I has been omitted from the notation for the sake of simplicity and without loss of generality.

A confirmation measure should also satisfy the following *classificatory* requirement expressed by Carnap [Carnap, 1962]:

$$c(E, H) \begin{cases} > 0 & \text{if} \quad \Pr(H \mid E) > \Pr(H) & \text{(confirmation)} \\ = 0 & \text{if} \quad \Pr(H \mid E) = \Pr(H) & \text{(neutrality)} \\ < 0 & \text{if} \quad \Pr(H \mid E) < \Pr(H) & \text{(disconfirmation)} \end{cases}$$

Consider, again, the DNA profile example above. Evidence E confirms H and F disconfirms H; $\Pr(H \mid E) > \Pr(H), \Pr(H \mid F) < \Pr(H)$ and $c(E, H) > 0$ and $c(F, H) < 0$.

This requirement is a general requirement as it does not require the specification of a predetermined value. Any positive or negative values characterize confirmation or disconfirmation. Several measures of confirmation have been proposed in the literature (Crupi and Tentori [2016]) that are in agreement with three logical requirements, those of compatibility and formality and a classificatory requirement. Two measures are - for simplicity - the common incremental measures of confirmation: first there is the difference (d) between the posterior and the prior probabilities,

$$c_d(E, H) = \Pr(H \mid E) - \Pr(H),$$

and, second, there is the ratio (r)

$$c_r(E, H) = \frac{\Pr(H \mid E)}{\Pr(H)} - 1,$$

sometimes reported as simply

$$c_r(E, H) = \frac{\Pr(H \mid E)}{\Pr(H)}.$$

A third measure, suggested by Good (Good [1950]), is the ratio between posterior and prior odds, instead of between posterior and prior probabilities. This measure defines the Bayes' factor. When one has a pair of simple, mutually exclusive, hypotheses, say H and $\bar{H}$, the Bayes' factor - the ratio between posterior and prior odds - reduces to the likelihood ratio

$$\text{LR} = \frac{\Pr(E \mid H)}{\Pr(E \mid \bar{H})}.$$

Notice the difference between the previous measures and the likelihood ratio. The previous measures refer to the probabilities of the hypotheses of interest (prior and posterior), the likelihood ratio refers to the probabilities of the evidence given the two hypotheses.

There is a conceptual distinction between posterior probability and a measure of confirmation. Notably, as specified by Tentori et al. [2013],

> [...] confirmation is a *relative notion* [*italics* added] in the following crucial sense: the credibility of a hypothesis can be changed by a given piece of evidence in either a *positive* (confirmation in a narrow sense) or *negative* way (disconfirmation). Confirmation (in the narrow sense) thus reflects an increase from prior to posterior probability, whereas disconfirmation reflects a decrease. As confirmation concerns the relationship between prior and posterior, there is simply no single probability value that can capture the notion. (at p. 240)

Any monotonic function of the Bayes' factor, such as its logarithm, satisfies the advocated requirements. Good (Good [1950]) denoted the logarithm of the Bayes' factor (BF) '*the weight of evidence*', log(BF). The Bayes' factor confirms a hypothesis if its value is greater than 1 and it disconfirms the hypothesis if the value is less than 1. Its logarithm has the interesting advantage of *additivity* (see Section 4.2).

An alternative function of the Bayes' factor is the one proposed in Kemeny and Oppenheim [1952]:

$$c_l(E, H) = \frac{\Pr(E \mid H) - \Pr(E \mid \bar{H})}{\Pr(E \mid H) + \Pr(E \mid \bar{H})} = \frac{\text{BF} - 1}{\text{BF} + 1}.$$

This expression also satisfies the previous requirements but is not considered further because of a difficulty of interpretation.

Focus here is on the Bayes' factor but a brief mention is made of the discussion by Schum (Schum [1994]) concerning problems with $c_d(E, H)$ and $c_r(E, H)$. He wrote:

> Unfortunately, there is trouble associated with grading the force of evidence in terms of either d or r, as defined above. The trouble is that changes in belief measured on a probability scale can be very misleading. What appears to be an insignificant belief change on a probability scale can in fact be a profound change in another scale directly related to probabilities; this scale involves the familiar term *odds*. (at p. 216)

Examples of such troubles are presented in Schum [1994] (see, also, Section 4.3). He noticed that - given that probability and odds scales are different[11] - a change in the probability scale when one goes closer to its maximum (value of 1) seems very slight as opposed to the same change (if measured in an odds scale) in other areas of the probability scale[12].

4 Rational justification of the use of a function of the Bayes' factor

Many measures have been proposed to quantify the value of evidence (Crupi and Tentori [2016])[13]. Many are problematic with respect to the previously mentioned basic requirements (see Section 3). In contrast, the Bayes' factor satisfies all these requirements. It is shown in this Section that the Bayes' factor satisfies other more detailed logical requirements.

---

[11] Probabilities have a scale form 0 to 1; the odds scale goes from 0 to infinity.

[12] Consider, as illustrated by Schum [1994], a first change in belief from $\Pr(H) = 0.1$ to $\Pr(H \mid E) = 0.526$ (a difference=0.426). Consider a second change in belief from $\Pr(H) = 0.9$ to $\Pr(H \mid E) = 0.989$ (a difference=0.089). If illustrated throughout the odds scale, the two situations are identical (from a prior odds of 9 and 1/9, and to a posterior odds of 89.9 and 1.11, respectively).

[13] In addition to the distance or the ratio measures, previously illustrated, it has been suggested to measure confirmation through $c_s(E, H) = \Pr(H \mid E) - \Pr(H \mid \bar{E})$ [Christensen, 1999], $c_m(E, H) = \Pr(E \mid H) - \Pr(E)$ [Mortimer, 1988], $c_g(E, H) = 1 - [\Pr(\bar{H} \mid E)/\Pr(\bar{H})]$ [Rips, 2001], $c_\tau(E, H) = \Pr(H, E) - Pr(H)Pr(E)$ [Eells and Fitelson, 2002] and many others.

4.1 The mathematical operators

The odds form of Bayes' theorem presents a compelling intuitive argument for the use of the likelihood ratio as a measure of the value of the evidence (see, e.g. Good [1985] reiterated in Buckleton et al. [2020]) or as a confirmation measure. A mathematical argument does exist to justify its use (see Good [1989b, 1991]). It is reproduced here to illustrate the purpose of the paper (see also Aitken et al. [2021]).

It is desired to measure the support of evidence E in favour of a hypothesis H. The likelihood ratio as a measure of the value of the evidence is in accordance with the formality requirement, according to which a confirmation measure must depend on the probability values concerning E and H, say $\Pr(E, H)$, $\Pr(\bar{E}, H)$, $\Pr(E, \bar{H})$ and $\Pr(\bar{E}, \bar{H})$ (see Section 3). A simple manipulation of the probabilities in the numerator and in the denominator of the likelihood ratio allows in fact to show:

$$\frac{\Pr(E \mid H)}{\Pr(E \mid \bar{H})} = \frac{\frac{\Pr(E,H)}{\Pr(E,H)+\Pr(\bar{E},H)}}{\frac{\Pr(E,\bar{H})}{\Pr(E,\bar{H})+\Pr(\bar{E},\bar{H})}}.$$

Let $x = \Pr(E \mid H)$ and $y = \Pr(E \mid \bar{H})$. The value V of the evidence can be expressed as $V = f(x, y)$ for some function f.

Consider another piece of evidence T which is irrelevant to E and H (and hence to $\bar{H}$) and which is such that $\Pr(T) = \theta$. Then

$$\begin{aligned} \Pr(E, T \mid H) &= \Pr(E \mid H) \Pr(T \mid H) \\ &= \Pr(E \mid H) \Pr(T) \\ &= \theta x. \end{aligned}$$

Similarly,

$$\Pr(E, T \mid \bar{H}) = \theta y.$$

The value of the combined evidence (E, T) is equal to the value of E, since T has been assumed irrelevant. The value of (E, T) is $V = f(\theta x, \theta y)$ and the value of E is $V = f(x, y)$. Thus

$$f(\theta x, \theta y) = f(x, y)$$

for all θ in the interval [0,1] of possible values of $\Pr(T)$.

The relationship between x and y within the function f may take one of four forms, depending on the four mathematical operators $+, \times, -$ and $/$. Take each in turn.

$+$ $f(x, y) = f(x + y)$

$f(\theta x, \theta y) = f(\theta x + \theta y) = f(\theta(x + y))$. This is not equal to $f(x + y)$ for all θ in the interval $[0, 1]$. For example, try

$$f(x, y) = (x + y)^2; f(\theta x, \theta y) = \theta^2 (x + y)^2 = \theta^2 f(x, y).$$

$\times$ $f(x, y) = f(x \times y)$;

$f(\theta x, \theta y) = f(\theta x \times \theta y) = f(\theta^2 (x \times y))$. This is not equal to $f(x \times y)$ for all θ in the interval [0,1].

$$f(x, y) = (x \times y)^2; f(\theta x, \theta y) = \theta^2 (x \times y)^2 = \theta^2 f(x, y).$$

$-$ $f(x, y) = f(x - y)$;

10

$f(\theta x, \theta y) = f(\theta x - \theta y) = f(\theta(x - y))$. This is not equal to $f(x - y)$ for all θ in the interval [0,1].

$$f(x, y) = (x - y)^2; f(\theta x, \theta y) = \theta^2(x - y)^2 = \theta^2 f(x, y).$$

/ $f(x, y) = f(x/y)$;
$f(\theta x, \theta y) = f(\theta x/\theta y) = f(x/y)$. This is equal to $f(x, y)$ for all θ in the interval [0,1].

$$f(x, y) = (x/y)^2; f(\theta x, \theta, y) = (\theta x/\theta y)^2 = f(x, y).$$

It follows that f is a function of x/y alone and hence that V is a function of

$$\Pr(E \mid H)/\Pr(E \mid \bar{H}),$$

namely, the likelihood ratio.

This argument is mathematical. It is abstract. The assumptions from which the result is derived are impeccable; it is very reasonable to assume that all is needed for the evaluation of evidence are the four probabilities given and this assumption is supported by the discussion in Section 3. As a mathematical result, it is general and applies to any form of evidence. The value of any form of evidence is a function of the likelihood ratio.

4.2 Additivity

As expressed in Section 3, the Bayes' factor is equivalent to the ratio of the posterior odds to the prior odds:

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

Expressed in logarithmic scale, the Bayes' factor is called 'the weight of evidence' and it can be re-written as follows:

$$\log\left[\frac{\Pr(H \mid E)}{\Pr(\bar{H} \mid E)}\right] - \log\left[\frac{\Pr(H)}{\Pr(\bar{H})}\right]. \tag{3}$$

The Bayes' factor and the logarithm of the Bayes' factor have 1 and 0, respectively, as the neutral value. This captures the idea - as expressed in the legal context - of the relevance of evidence as described by the Federal Rule of Evidence (FRE 401). The FRE 401 says that

'Relevant evidence' means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence. (Mueller and Kirkpatrick [1988] at p. 33)

Measured in terms of the logarithm of the Bayes' factor, the confirmation measure (or weight of evidence) has additivity as a desirable property as underlined by Lempert [1977] and Kaye [1986], respectively:

[E]vidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue. (at p. 1026)

and

Evidence is relevant as that term is used in Rule 401 if its log-likelihood ratio is not zero. (at p. 765)

Schum [1994] described the additivity property by illustrating the derivation of the weight for two items of evidence and he concluded by affirming that

In log likelihood ratio terms, the force of the evidence is always additive whether or not the evidence items are conditionally independent. This additivity property extends to any number of evidence items. (at p. 220)

In fact, consider two items of evidence, call them E and F and the Bayes' factor for those two items, $\Pr(E \mid H)/\Pr(E \mid \bar{H})$ and $\Pr(F \mid E, H)/\Pr(F \mid E, \bar{H})$, respectively, where $\Pr(E \mid H)/\Pr(E \mid \bar{H})$ can be defined as BF_E and $\Pr(F \mid E, H)/\Pr(F \mid E, \bar{H})$ as $\mathrm{BF}_{F|E}$.

Suppose first that the value of F does depend upon E. The posterior odds in favor of H given both items of evidence E and F can be written as

$$\frac{\Pr(H \mid E, F)}{\Pr(\bar{H} \mid E, F)} = \frac{\Pr(H)}{\Pr(\bar{H})} \times \mathrm{BF}_E \times \mathrm{BF}_{F|E}. \tag{4}$$

By dividing both sides by prior odds, we obtain

$$\frac{\frac{\Pr(H|E,F)}{\Pr(\bar{H}|E,F)}}{\frac{\Pr(H)}{\Pr(\bar{H})}} = \mathrm{BF}_{E,F} = \mathrm{BF}_E \times \mathrm{BF}_{F|E}. \tag{5}$$

In the logarithm form, we have $\log(\mathrm{BF}_{E,F}) = \log(\mathrm{BF}_E) + \log(\mathrm{BF}_{F|E})$. The total weight of evidence is indicated by the sum of the logarithms of the two Bayes' factors.

Suppose now that the two items of evidence, E and F, can be considered as independent conditional on H and $\bar{H}$, so that the Bayes' factor for items F, $\mathrm{BF}_{F|E}$, becomes $\Pr(F \mid H)/\Pr(F \mid \bar{H})$ and can be re-written as BF_F. The total weight of evidence is again indicated by the sum of the logarithms of the two Bayes' factors: $\log(\mathrm{BF}_E) + \log(\mathrm{BF}_F)$. The additivity property does not depend on the potential dependence between items of evidence.

It may be easier to understand the concept of additivity with an analogy with the scales of justice. The odds form of Bayes' theorem (2), repeated here for convenience with reference to the background information I omitted, is

$$\frac{\Pr(H \mid E)}{\Pr(\bar{H} \mid E)} = \frac{\Pr(E \mid H)}{\Pr(E \mid \bar{H})} \times \frac{\Pr(H)}{\Pr(\bar{H})}. \tag{6}$$

Its logarithmic form may be written as

$$\begin{aligned}
\log(\Pr(H \mid E)) - \log(Pr(\bar{H} \mid E)) &= \log(\Pr(E \mid H)) - \log(\Pr(E \mid \bar{H})) + \log(\Pr(H)) - \log(\Pr(\bar{H})) \\
&= [\log(\Pr(E \mid H)) + \log(\Pr(H))] - [\log(\Pr(E \mid \bar{H})) + \log(\Pr(\bar{H}))].
\end{aligned}$$

The left-hand expressions on both sides of the equality, $\log(\Pr(H \mid E)$ and $[\log(\Pr(E \mid H)) + \log(\Pr(H))]$, may be considered as the weight in one pan of the scales, denoted the H pan. The right-hand expressions $\log(\Pr(\bar{H} \mid E))$ and $[\log(\Pr(E \mid \bar{H})) + \log(\Pr(\bar{H}))]$ may be considered as the weight in the other pan of the scales, denoted the $\bar{H}$ pan. The pan with the greater weight indicates the proposition better supported by the evidence.

The additive property of the logarithmic form of Bayes' theorem then is illustrated when new evidence F is introduced. Add $\log(\Pr(F \mid E, H))$ to the H pan and $\log(\Pr(F \mid E, \bar{H}))$ to the $\bar{H}$ pan.

4.3 Adequacy and Logicality

Crupi et al. [2007] considered a function, say v, such that for any argument (E, H), v assigns it the same positive value (e.g. $+1$) if and only if E implies H (i.e. $\Pr(H \mid E) = 1$), an equivalent value with the opposite sign (i.e. -1) if and only if E refutes H (with 'refutes' H being used as a synonym for 'implies' $\bar{H}$) (i.e. $\Pr(H \mid E) = 0$), and value

0 otherwise (i.e. $0 < \Pr(H \mid E) < 1$). The authors wrote that 'The relationships between the logical implication or refutation of H by E and the conditional probability of H given E yield that any Bayesian confirmation measure c agrees with v in the minimal sense that if $v(E, H)$ is positive, the same is true of $c(E, H)$; and if $v(E, H)$ is negative, the same is true of $c(E, H)$' (at p. 231-232). As a consequence, this *adequacy constraint* is expressed in the following general terms implying two items of evidence (E_1 and E_2) and two sets of hypotheses (H_1 and H_2):

$$\text{If } v(E_1, H_1) > v(E_2, H_2), \text{ then } c(E_1, H_1) > c(E_2, H_2).$$

This expression guarantees that the value of $c(E, H)$ for a conclusive confirmatory argument (evidence E confirms H) is higher than that of an argument that is not conclusively confirmatory (e.g. E is correlated with H). Similar reasoning may be used for disconfirmatory arguments.

In this respect, Fitelson [2006] supported the idea that the Bayes' factor (and every equivalent measure, such as its logarithm) is characterized by a property he called *logicality* that the confirmation measure $c(E, H)$ is maximal when evidence E implies hypothesis H and minimal when evidence E implies hypothesis $\bar{H}$. Consider a Bayesian confirmation measure c. It is a function of E, H and a probability model Pr. The logicality requirement is that it takes a maximum value (is *maximal*) when $E \Rightarrow H$ (i.e. when $\Pr(H \mid E) = 1$) and a minimum value (is *minimal*) when $E \Rightarrow \bar{H}$ (i.e. when $\Pr(\bar{H} \mid E) = 1$). Both the maximum value and the minimum value are independent of E and H.

(a) If evidence E implies hypothesis H, $\Pr(H \mid E) = 1$ and $\Pr(H \mid E)/\Pr(\bar{H} \mid E) = \infty$; the Bayes' factor which is the ratio of posterior odds to prior odds is equal to ∞ also and is at its maximum. Consider, for sake of illustration, the following extreme situation in which every human being on planet Earth has been genetically typed and their result recorded in a DNA database. Consider also an error-free laboratory. A genetical correspondence between the recovered bloodstain stain profile and that of a defendant has been reported. All other individuals in the database but the defendant have been categorically excluded as potential source of the stain. The defendant is the donor of the strain, so the BF $= \infty$.

(b) If evidence E implies hypothesis $\bar{H}$, $\Pr(\bar{H} \mid E) = 1$ and $\Pr(H \mid E)/\Pr(\bar{H} \mid E) = 0$; the Bayes' factor takes its minimal value. Consider another extreme situation in which an error-free laboratory is able to categorically discriminate between the DNA profiles of the recovered biological stain and that of a defendant. The defendant cannot be the source of the stain; the BF $= 0$.

On the contrary, consider, for sake of illustration, the confirmation measure $c_d(E, H)$. This measure does not satisfy logicality.

(a) If evidence E implies hypothesis H, then $c_d(E, H) = \Pr(H \mid E) - \Pr(H) = 1 - \Pr(H) = \Pr(\bar{H})$ so that its maximum value depends on the prior probability of H.

(b) If evidence E implies hypothesis $\bar{H}$, then $c_d(E, H) = \Pr(H \mid E) - \Pr(H) = 0 - \Pr(H) = -\Pr(H)$ and again its minimum value depends on the prior probability of H.

Table 1 shows that for the confirmation measure c_r only the minimal value condition is satisfied; i.e. c_r is minimal when $E \Rightarrow \bar{H}$. The maximal value depends on the prior probability of H. Table 1 also shows that for the confirmation measure c_g (see footnote 12 above) only the maximal value condition is satisfied; i.e. c_g is maximal when $E \Rightarrow H$ while the minimal value depends on the prior probability of $\bar{H}$.

Some further desirable properties can be specified. Crupi et al. [2007] consider

1. if $\Pr(H \mid E_1) > \Pr(H \mid E_2)$, then $c(E_1, H) > c(E_2, H)$;

Table 1: Confirmation measure $c(E, H)$ under situations $E \Rightarrow H$ ($\Pr(H \mid E) = 1$, maximal) and $E \Rightarrow \bar{H}$ ($\Pr(\bar{H} \mid E) = 1$, minimal).

Measure	$E \Rightarrow H$	$E \Rightarrow \bar{H}$
$c_r(E, H) = \frac{\Pr(H\mid E)}{\Pr(H)} - 1$	$\frac{\Pr(\bar{H})}{\Pr(H)}$	-1
$c_g(E, H) = 1 - \frac{\Pr(\bar{H}\mid E)}{\Pr(\bar{H})}$	1	$\frac{-\Pr(H)}{\Pr(\bar{H})}$

2. if hypothesis H_1 implies evidence E, hypothesis H_2 implies evidence E and $\Pr(H_1) > \Pr(H_2)$, then $c(E, H_1) > c(E, H_2)$;

3. if $\Pr(E \mid H_1) > \Pr(E \mid H_2)$ and $\Pr(E \mid \bar{H}_1) < \Pr(E \mid \bar{H}_2)$, then $c(E, H_1) > c(E, H_2)$.

Consider, for sake of illustration, the case where proposition H_1 can be formalized as 'the person of interest is the source of the recovered stain' and that H_2 is 'the person of interest touched the object of interest'. So, property 3 specifies that if the probability of observing a correspondence between the genetic DNA profiles (E) given that the person of interest is the source of the recovered stain (H_1) is greater than the probability to observe such correspondence given that that person directly touched (primary transfer) the object (H_2), and that the probability of observing a correspondence between the genetic DNA profiles (E) given that the person of interest is not the source of the recovered stain ($\bar{H}_1$) is smaller than the probability to observe such correspondence given that that person did not directly touch (secondary transfer) the object ($\bar{H}_2$), then the likelihood ratio for the first pair of source hypotheses is greater than the likelihood ratio for the second pair of activity hypotheses.

4.4 The symmetries

To discriminate between confirmation measures, Eells and Fitelson [2002] suggested a critical analysis based on three questions that represented various aspects of the concept of symmetry. The authors supported the use of the Bayes' factor noticing that it gave correct answers to all three questions. The questions referred to Evidential symmetry (ES), Commutativity symmetry (CS) and Hypothesis symmetry (HS), respectively (at p. 129).

ES : Does a piece of evidence E support a hypothesis H equally well as E's negation ($\bar{E}$) undermines, or counter-supports, the same hypothesis H? In mathematical terms, is $c(E, H) = \bar{c}(\bar{E}, H)$[14]?

CS : Does a piece of evidence E support a hypothesis H as equally well as H supports E? In mathematical terms, is $c(E, H) = c(H, E)$?

HS : Does a piece of evidence E support a hypothesis H as equally well as E undermines, or counter-supports, the negation of H ($\bar{H}$)? In mathematical terms, is $c(E, H) = \bar{c}(E, \bar{H})$?

A coherent measure of confirmation does not need to provide positive answers to the first two questions as demonstrated by Eells [2000] throughout a series of illustrative examples. For example, CS does not hold in general. A piece of evidence E can confirm a hypothesis H to a much different degree than H confirms E. Eells and Fitelson [2002] illustrated this through an example: 'Consider for example whether the observation that a card

---

[14]Note that $\bar{c}$ is the negation of c. In particular, if c confirms the support of evidence E for hypothesis H then $\bar{c}$ disconfirms the support of $\bar{E}$ for H.

is the seven of spades confirms the proposition that the card is black equally well as the proposition that the card is black confirms the proposition that the card is the seven of spades. With initial uncertainty about the value of the card, we consider the seven of spades, as evidence, to be more highly informative and confirmatory of the blackness of the card, as a hypothesis, than the blackness of the card, as evidence, is for the card's being the seven of spades in particular' (at p. 133). Analogous examples in forensic science can easily be found; typically the well-known transposed conditional situation that emphasises the difference between $\Pr(E \mid H)$ and $\Pr(H \mid E)$ as illustrated by an example in Section 1, footnote 3.

The answer to the third question should be positive. Given that there are two mutually exclusive and exhaustive (within the context of the case) hypotheses, the evidential support of E for H should be of the opposite sign as the evidential support of the same evidence E for the alternative hypothesis $\bar{H}$. Demonstrations of the correct answers to these three questions can be found in Eells and Fitelson [2002]. A discussion is available in Crupi et al. [2007], Tentori et al. [2007a]. An illustrative comparison between $c_r(E, H) = \Pr(H \mid E)/\Pr(H)$ and $c_{bf}(E, H) = \Pr(E \mid H)/\Pr(E \mid \bar{H})$ in their logarithm forms[15] is presented in the Appendix (Section 6).

There are several measures that satisfy HS such as c_d, c_s, c_τ and c_{bf}. However, the very desirable property of additivity (Section 4.2) is only satisfied by c_{bf} as shown following (4) and (5). As mentioned by Edwards [1986]:

> It is symmetric. [...] Furthermore, the log-likelihood ratio has the lovely property of additivity. [...] log-likelihood ratio is the only measure available in all of probability theory, so far as I know, that has that attractive property. (at p. 626)

5   Conclusion

Faced with uncertainty, scientific evidence is increasingly presented in a numerical form related to probability as the measure for uncertainty. International guidelines (e.g. European Network of Forensic Science Institutes [2015]) refer to the use of the likelihood ratio (or Bayes' factor) as the operational standard measure for the value of evidence.

Any form of presentation for the evidence that is adopted by scientists must be logical. Logicality of the adopted form must be demonstrated. Criteria for the satisfaction of the requirements of logicality have been given. The Bayes' factor has been shown to satisfy the logicality requirements.

Compelling reasons have been shown for there to be a preference for the Bayes' factor and any function of it (such as its logarithm form) over other measurements of evidential value described in the scientific literature [Fitelson, 1999, 2011, Crupi and Tentori, 2014, 2016]. The satisfaction by the Bayes' factor of all the the reasonable logical requirements put forward in the philosophical literature justifies its use as a measure for the value of evidence and supports its use in forensic science. Therefore, there can be no controversy concerning the use of the Bayes' factor or its logarithm (the so-called 'weight of evidence') in a Court of Justice.

The analysis developed and results obtained may be considered as support for the use of the Bayes' factor; support that is additional to general criteria expressed earlier [Buckleton et al., 2020, e.g.]. The arguments above to support the use of the Bayes factor / likelihood ratio (which, it was stated, have been treated as synonyms) were introduced initially in the context of forensic science. The arguments were then extended implicitly to consideration of the whole case and the entirety of the evidence. This extension is not a suggestion that the forensic scientist consider the whole case and the entirety of the evidence. Such considerations are a clearly improper extension of their role. However, the extension is justifiable when applied to the finders of fact, the judge and jury. For their consideration of the whole case and the entirety of the evidence, it has been shown that the best way for this to be done is for there to be consideration of the relative values of the probabilities of the evidence conditional on the propositions of the prosecution and the defence.

---

[15]The use of the logarithm for $c_r(E, H)$ and $c_{bf}(E, H)$ ensures that the ratio measures are positive (+), negative (-) and neutral (equals 0) if and only if E confirms, disconfirms, is confirmationally irrelevant (neutral) to H.

Acknowledgements

6   Appendix

Consider the previous example about cards where one observed the seven of spades (E) that confirms that the card is black (H). Consider also a regular 52 deck of playing cards and the following probabilities:

- $\Pr(H) = 1/2$;

- $\Pr(E) = 1/52$.

The joint probabilities $\Pr(H, E)$, $\Pr(\bar{H}, E)$, $\Pr(H, \bar{E})$ and $\Pr(\bar{H}, \bar{E})$ can be obtained as follows:

- $\Pr(H, E) = \Pr(H \mid E) \times \Pr(E) = 1 \times 1/52 = 1/52$;

- $\Pr(\bar{H}, E) = \Pr(\bar{H} \mid E) \times \Pr(E) = 0 \times 1/52 = 0$;

- $\Pr(H, \bar{E}) = \Pr(H \mid \bar{E}) \times \Pr(\bar{E})$. $\Pr(H \mid \bar{E})$ is obtained via Bayes' theorem. $\Pr(H, \bar{E}) = \Pr(\bar{E} \mid H) \times \Pr(H) = 25/25 \times 1/2 = 25/52$;

- $\Pr(\bar{H}, \bar{E}) = \Pr(\bar{H} \mid \bar{E}) \times \Pr(\bar{E})$. $\Pr(\bar{H} \mid \bar{E})$ is obtained via Bayes' theorem. $\Pr(\bar{H}, \bar{E}) = \Pr(\bar{E} \mid \bar{H}) \times \Pr(\bar{H}) = 1 \times 1/2 = 1/2$.

Using the card example, measures c_r and c_{bf} are tested against symmetry considerations.

6.1   Symmetry considerations and the ratio measure $c_r(E, H)$

6.1.1   Evidence Symmetry, ES, $c_r(E, H) = -c_r(\bar{E}, H)$

$c_r(E, H) = \log[\Pr(H \mid E)/\Pr(H)] = \log[1/0.5] = \log[2]$ and $-c_r(\bar{E}, H) = -\log[\Pr(H \mid \bar{E})/\Pr(H)] = -\log[(25/51)/0.5] = -\log[50/51] = \log[51/50]$. Therefore, the measure c_r violates the Evidence Symmetry because $c_r(E, H) \neq -c_r(\bar{E}, H)$.

6.1.2   Commutativity Symmetry, CS, $c_r(E, H) = c_r(H, E)$

$c_r(E, H) = \log[\Pr(H \mid E)/\Pr(H)] = \log[1/0.5] = \log[2]$ and $c_r(H, E) = \log[\Pr(E \mid H)/\Pr(E)] = \log[(1/26)/(1/52)] = \log[2]$. Therefore, the measure c_r satisfies the Commutativity Symmetry because $c_r(E, H) = c_r(H, E)$.

6.1.3   Hypothesis Symmetry, HS, $c_r(E, H) = -c_r(E, \bar{H})$

$c_r(E, H) = \log[\Pr(H \mid E)/\Pr(H)] = \log[1/0.5] = \log[2]$ and $-c_r(E, \bar{H}) = -\log[\Pr(\bar{H} \mid E)/\Pr(\bar{H})] = \log[\Pr(\bar{H})/\Pr(\bar{H} \mid E)] = \log[0.5/0] = \infty$. Therefore, the measure c_r violates the Hypothesis Symmetry because $c_r(E, H) \neq -c_r(E, \bar{H})$.

The measure $c_r(E, H)$ has an undesirable property under Commutativity Symmetry and Hypothesis Symmetry.

6.2   Symmetry considerations and the ratio measure $c_{bf}(E, H)$

6.2.1   Evidence Symmetry, ES, $c_{bf}(E, H) = -c_{bf}(\bar{E}, H)$

$c_{bf}(E, H) = \log[\Pr(E \mid H)/\Pr(E \mid \bar{H})] = \log[(1/26)/(0)] = \infty$ and $-c_{bf}(\bar{E}, H) = -\log[\Pr(\bar{E} \mid H)/\Pr(\bar{E} \mid \bar{H})] = -\log[(25/26)/1] = \log[26/25]$. Therefore, the measure c_{bf} violates the Evidence Symmetry because $c_{bf}(E, H) \neq -c_{bf}(\bar{E}, H)$.

6.2.2 Commutativity Symmetry, CS, $c_{bf}(E,H) = c_{bf}(H,E)$

$c_{bf}(E,H) = \log[\Pr(E \mid H)/\Pr(E \mid \bar{H})] = \log[(1/26)/(0)] = \infty$ and $c_{bf}(H,E) = \log[\Pr(H \mid E)/\Pr(H \mid \bar{E})]$. $\Pr(H \mid \bar{E})$ is obtained via Bayes' theorem. $c_{bf}(H,E) = \log[1/(25/51)] = \log[51/25]$. Therefore, the measure c_{bf} violates the Commutativity Symmetry because $c_{bf}(E,H) \neq c_{bf}(H,E)$.

6.2.3 Hypothesis Symmetry, HS, $c_{bf}(E,H) = -c_{bf}(E,\bar{H})$

$c_{bf}(E,H) = \log[\Pr(E \mid H)/\Pr(E \mid \bar{H})] = \log[(1/26)/(0)] = \infty$ and $-c_{bf}(E,\bar{H}) = -\log[\Pr(E \mid \bar{H})/\Pr(E \mid H)] = \log[\Pr(E \mid H)/\Pr(E \mid \bar{H})$. Therefore, the measure c_{bf} satisfies the Hypothesis Symmetry because $c_{bf}(E,H) = -c_{bf}(E,\bar{H})$.

The measure $c_{bf}(E,H)$ respects all the three symmetry properties.

References

C G G Aitken and F Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 2nd edition, 2004.

C G G Aitken, F Taroni, and S Bozza. *Statistics and the Evaluation of Evidence for Forensic Scientists*. John Wiley & Sons, Chichester, 3rd edition, 2021.

J S Buckleton, B Robertson, J Curran, C Berger, D Taylor, J A Bright, T Hicks, S Gittelson, I W Evett, S Pugh, G Jackson, H Kelly, T Kalafut, and F R Bieber. A review of likelihood ratios in forensic science based on a critique of Stiffelman 'No longer the gold standard: probabilistic genotyping is changing the nature of DNA evidence in criminal trials'. *Forensic Science International*, 310, 2020. https://doi.org/10.1016/j.forsciint.2020.110251.

R Carnap. *Logical foundations of probability*. University of Chicago Press, Chicago, 2nd edition, 1962.

DS Christensen. Measuring confirmation. *Journal of Philosophy*, 96:437–461, 1999.

V Crupi and K Tentori. State of the field: measuring information and confirmation. *Studies in History and Philosophy of Science*, 47:81–90, 2014.

V Crupi and K Tentori. Confirmation theory. In A Hájek and C Hitchcock, editors, *Oxford Handbook of Philosophy and Probability*, pages 650–665. Oxford University Press, Oxford, 2016.

V Crupi, K Tentori, and M Gonzalez. On Bayesian measures of evidential support: theoretical and empirical issues. *Philosophy of Science*, 74:229–252, 2007.

V Crupi, N Chater, and K Tentori. New axioms for probability and likelihood ratio measures. *The British Journal for the Philosophy of Science*, 64:189–204, 2013.

J Darroch. Probability and criminal trials; some comments prompted by the Splatt trial and The Royal Commission. *Professional Statistician*, 6:3–7, 1987.

B de Finetti. Funzione caratteristica di un fenomeno aleatorio. *Memorie della Regia Accademia dei Lincei*, 4:86–133, 1930.

B de Finetti. Probabilismo. In A Aliotta, editor, *Biblioteca di Filosofia*, pages 1–57. Editrice F. Perrella, Napoli, 1931.

B de Finetti. Probabilismo. *Logos*, 14:163–219, 1931.

B de Finetti. Probability: the subjective approach. In R Klibansky, editor, *La philosophie contemporaine*, pages 45–53. La Nuova Firenze, Firenze, 1968.

B de Finetti. *Theory of Probability. A critical introductory treatment, Volume 1*. John Wiley & Sons, New York, 1970.

B de Finetti. Probabilism. *Erkenntnis*, 31:169–223, 1989.

P Diaconis and D Freedman. The persistence of cognitive illusions. *Behavioural and Brain Sciences*, 4:333–334, 1981.

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.

W Edwards. A diagrammatic approach to evidence - Comment. *Boston University Law Review*, 66:623–628, 1986.

E Eells. Review: The foundations of causal decision theory, by James M. Joyce. *The British Journal for the Philosophy of Science*, 51:893–900, 2000.

E Eells and B Fitelson. Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107:129–142, 2002.

European Network of Forensic Science Institutes. Guideline for evaluative reporting in forensic science. Bruxelles, 2015.

I W Evett. Avoiding the transposed conditional. *Science & Justice*, 35(2):127–131, 1995.

I W Evett and B S Weir. *Interpreting DNA Evidence*. Sinauer Associates Inc., Sunderland, 1998.

R Festa. *Cambiare opinione. Temi e problemi di epistemologia bayesiana*. Edizioni CLUEB, Bologna, 1996.

M O Finkelstein and W B Fairley. A Bayesian approach to identification evidence. *Harvard Law Review*, 83:489–517, 1970.

B Fitelson. The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66:S362–S378, 1999.

B Fitelson. Logical foundations of evidential support. *Philosophy of Science*, 73:500–512, 2006.

B Fitelson. Favoring, likelihoodism, and Bayesianism. *Philosophy and Phenomenological Research*, 83:666–671, 2011.

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

M C Galavotti and R Jeffrey. Preface to Bruno de Finetti's Philosophy of Probability. *Erkenntnis*, 31:165–167, 1989.

P Giannelli, E J Imwinkelried, and J L Peterson. Reference guide on forensic identification expertise. In Federal Judicial Center and National Research Council, editors, *Reference Manual on Scientific Evidence*, pages 55–127. The National Academies Press, Washington D.C., 3rd edition, 2011.

S Gittelson, , C E H Berger, G Jackson, I W Evett, C Champod, B Robertson, J M Curran, D Taylor, B S Weir, M D Coble, and J S Buckleton. A response to 'Likelihood ratio as weight of evidence: a closer look' by Lund and Iyer. *Forensic Science International*, 288:e15–e19, 2018.

I J Good. *Probability and the Weighing of Evidence*. Griffin, London, 1950.

I J Good. Weight of evidence: a brief survey (with discussion). In J M Bernardo, M H DeGroot, D V Lindley, and A F M Smith, editors, *Bayesian Statistics 2*, pages 249–270. North Holland, Amsterdam, 1985.

I J Good. C319: Weight of evidence and a compelling metaprinciple. *Journal of Statistical Computation and Simulation*, 31:121–123, 1989b.

I J Good. Weight of evidence and the Bayesian likelihood ratio. In C G G Aitken and D A Stoney, editors, *The Use of Statistics in Forensic Science*, pages 85–106. Ellis Horwood, Chichester, 1991.

R C Jeffrey. Probability and falsification: critique of the Popper program. *Synthese*, 30:95–117, 1975.

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

D H Kaye. Quantifying probative value. *Boston University Law Review*, 66:761–766, 1986.

D H Kaye. Forensic statistics in the courtroom. In D Banks, K Kafadar, DH Kaye, and M Tackett, editors, *Handbook of Forensic Statistics*, pages 225–248. CRC Press, Boca Raton, 2021.

D H Kaye and G Sensabaugh. Reference guide on DNA identification evidence. In Federal Judicial Center and National Research Council, editors, *Reference manual on scientific evidence*, pages 129–210. The National Academies Press, Washington D.C., 3nd edition, 2011.

J Kemeny and P Oppenheim. Degrees of factual support. *Philosophy of Science*, 19:307–324, 1952.

R O Lempert. Modeling relevance. *Michigan Law Review*, 75:1021–1057, 1977.

D V Lindley. Scoring rules and the inevitability of probability. *International Statistical Review*, 50:1–26, 1982.

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

P Maher. Probability captures the logic of scientific confirmation. In C R Hitchcock, editor, *Contemporary debates in philosophy of science*, pages 69–93. Blackwell, Oxford, 1996.

H Mortimer. *The Logic of Induction*. Prentice Hall, Paramus, 1988.

C Mueller and L Kirkpatrick. *Federal Rules of Evidence*. Little, Brown, Boston, 1988.

E Murphy. No room for error: clear-eyed justice in forensic science oversight. *Harvard Law Review Forum*, 130: 145–153, 2017.

L J Rips. Two kinds of reasoning. *Psychological Science*, 12:129–134, 2001.

M J Saks and S L Neufeld. Convergent evolution in law and science: the structure of decision-making under uncertainty. *Law, Probability and Risk*, 10:133–148, 2011.

W C Salmon. *The Foundations of Scientific Inference*. University of Pittsburgh Press, Pittsburgh, PA, 1966.

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

B Stiffelman. No longer the gold standard: probabilistic genotyping is changing the nature of DNA evidence in criminal trials. *Berkeley Journal of Criminal Law*, 24:110–146, 2019.

F Taroni, P Garbolino, and S Bozza. Coherently updating degrees of beliefs: Radical Probabilism, the generalization of Bayes' theorem and their consequences on evidence evaluations. *Law, Probability and Risk*, 19:293–316, 2020.

K Tentori, V Crupi, N Bonini, and D Osherson. Comparison of confirmation measures. *Cognition*, 103:107–119, 2007a.

K Tentori, V Crupi, and D Osherson. Determinants of confirmation. *Psychonomic Bulletin and Review*, 14:877–883, 2007b.

K Tentori, V Crupi, and S Russo. On the determinants of the conjuction fasllacy: probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142:235–255, 2013.

W C Thompson and E L Schumann. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behaviour*, 11:167–187, 1987.

W C Thompson, R Hofstein Grady, H Lai, and H S Stern. Do jurors give appropriate weight to forensic identification evidence? *Law, Probability and Risk*, 17:133–155, 2018.

P Tillers. Trial by mathematics - reconsidered. *Law, Probability and Risk*, 10:167–173, 2011.

B C van Fraassen. *Laws and Symmetry*. Clarendon Press, Oxford, 1989.

J von Plato. De Finetti's earliest works on the foundations of probability. *Erkenntnis*, 31:263–282, 1989.