



Eyewitness evaluation through inference to the best explanation

Hylke Jellema¹

Received: 1 July 2021 / Accepted: 13 August 2022 / Published online: 28 September 2022
© The Author(s) 2022

Abstract

Eyewitness testimony is both an important and a notoriously unreliable type of criminal evidence. How should investigators, lawyers and decision-makers evaluate eyewitness reliability? In this article, I argue that Testimonial Inference to the Best Explanation (TIBE) is a promising, but underdeveloped prescriptive account of eyewitness evaluation. On this account, we assess the reliability of eyewitnesses by comparing different explanations of how their testimony came about. This account is compatible with, and complementary to both the Bayesian framework of rational eyewitness evaluation and with prescriptive methods for eyewitness assessment developed by psychologists. Compared to these frameworks, the distinctive value of thinking in terms of competing explanations is that it helps us select, interpret and draw conclusions from the available evidence about the witness's reliability.

1 Introduction

Eyewitness testimony is one of the most important kinds of criminal evidence. It is also notoriously unreliable and a major source of judicial errors (Cardozo, 2009). In criminal cases, decision-makers therefore regularly face the difficult but crucial task of evaluating eyewitness testimony. This sometimes means checking whether the witness's story fits with other established facts of the case. However, the veracity of such a story cannot always be verified or falsified directly. In such cases evaluators will have to look at whether the statement comes from a reliable source. How should they go about doing so? I argue that *Testimonial Inference to the Best Explanation* (TIBE) is a useful but underdeveloped account of eyewitness evaluation. On TIBE,

✉ Hylke Jellema
h.jellema@rug.nl

¹ Faculty of Law, University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, the Netherlands

we compare competing explanations of why the particular statement was offered (see Sect. 2 for further exposition).

Explanation-based approaches are increasingly popular in evidence scholarship (e.g., Amaya 2015; Allen & Pardo, 2019; Jellema, 2021). These approaches conceptualize legal proof as a competition between potential explanations of the available evidence – usually in the form of narratives told by the parties at trial. The main benefit of such approaches compared to competing accounts is that they fit with how people reason naturally and that they offer direction to decision-makers as to how to go about assessing the available evidence (Pennington & Hastie, 1993; Allen & Pardo, 2019; Nance, 2016, 84). So far, these approaches have focused mainly on the decision whether the proof standard has been met – for instance whether guilt has been proven beyond a reasonable doubt in a criminal case. However, as some have suggested – and as I show in this article – explanatory reasoning is similarly a plausible, useful way of thinking about the *reliability* of the available evidence, such as eyewitness testimony.

I am not the first to argue that we can use inference to the best explanation to assess eyewitness testimony. However, existing work on TIBE in the philosophy of testimony has mostly been limited to brief, descriptive accounts, intended to capture our intuitions about when we may trust the utterances of others in daily life. In this article I show that explanation-based reasoning is also a *rational* and *useful* approach to assessing eyewitness reliability. In particular, I develop a *prescriptive* account that offers guidance to eyewitness evaluators. Furthermore, I argue that, and explore under which conditions, this approach is rational. To this end, I connect my account to its main competitor, the Bayesian accounts of eyewitness reliability (Sect. 3). Contrary to what others have said, TIBE and Bayesianism are compatible.¹ I follow a line of thought from the philosophy of science, according to which inference to the best explanation is an efficient, but imperfect heuristic for optimal Bayesian reasoning. Because, and to the extent that, TIBE tracks the Bayesian ideal, it is rational. However, I do not suggest that TIBE is subsumed under Bayesianism. Bayesian probability theory offers a helpful way of making precise why and when TIBE is rational, but no more than that. I argue that the crucial question for an account of rational eyewitness evaluation is how to make sense of the available evidence concerning the witness's reliability. On this front, TIBE is doing all the hard, prescriptive work.

To illustrate both what is distinctive about TIBE, and how this approach may be further developed, I compare it with two popular types of prescriptive accounts: the capacity view and empirically-informed frameworks (Sect. 4). I argue that, though these accounts differ on crucial points, they can also complement one another. First, while these approaches are not inherently explanatory, once we incorporate an explanatory component, we end up with a more context-sensitive, flexible framework on which it is clearer how evaluators ought to come to their ultimate decision. Conversely, these frameworks can inform us on some of the details of explanatory

¹ By spelling out this relationship in detail, this article also contributes to the ongoing debate on the compatibility between Bayesian and explanation-based approaches to legal evidence in general (cf. Mackor et al., 2021). The approach suggested here is a plausible way of spelling out this relationship, though further work will need to be done to see how it could be applied to other aspects of legal evidential reasoning.

inference in practice. Finally, I further explore TIBE's distinctive value compared to both these prescriptive approaches and to Bayesianism by turning to two specific aspects of eyewitness evaluation where explanatory reasoning is especially helpful, namely the interpretation (Sect. 5) and selection (Sect. 6) of the available evidence about the witness's reliability.

2 Testimonial inference to the best explanation

Inference to the best explanation (IBE) is a well-known account of evidential reasoning from the philosophy of science (Harman, 1965; Lipton, 2003). On IBE we compare competing explanations of some datum and (depending on the particulars of the account), conclude that the best of these is true, probably true or, at the least, worthy of further investigation. TIBE applies this comparative approach to the evaluation of testimony. However, those who have written on TIBE have not offered an extensive characterization of what such explanation-based evaluation looks like in practice. Let's therefore fill in some of the details of this account.

2.1 Sketching the account

When we evaluate the reliability of a piece of testimony, we are not concerned with evaluating directly whether the statements in the report are true. Rather, what we evaluate is whether the report is a reliable source. The report in question can be the witness's utterances, a written report about those utterances, or even our own recollection of what the witness or the written report said. On TIBE we evaluate the reliability of this source by comparing a select number of explanations of why this particular statement is offered in this particular situation, by this particular source. For instance, suppose that the witness of a robbery was interviewed and the report from this interview states that:

The robber wore a red shirt.

This statement contains the *story* that the witness (or, more precisely, the report) tells. There can be various explanations *why* the report states that 'the robber wore a red shirt. For example:

- i. The witness accurately observed the robber up-close when the alleged events took place and is now sharing their observations.
- ii. The witness misremembered the colour of the shirt due to the stress of the situation and now reports on this flawed memory.
- iii. The witness is lying about the colour of the witness's shirt because they are trying to frame someone.
- iv. The person who wrote down the witness's remark misinterpreted what was said.²

² Note that, in this situation, the source of the report whose reliability we are assessing is the report of what the witness said, not the witness themselves.

We may generate relevant explanations by asking: “What could explain this statement in this context?” Which explanations we consider will therefore depend on the particulars of the case.³ Certain contexts will feature typical sources of error, such as memory-related biases in stressful situations. Or there may, for instance, be some indication that the witness is lying, which could make us consider this as a possibility.

On TIBE we compare two kinds of explanations: Explanation (i) would, if it were true, also imply the probable truth of the fact that the witness is reporting on. This is therefore a *truth-telling* explanation. Explanations (ii), (iii) and (iv) would, if they were true, not imply that the hypothesis in the report is reporting on is true. I call these *alternative* explanations.

On TIBE, if the best explanation implies the truth of what is reported on, we may assign the report a high degree of reliability with respect to this particular statement. Conversely, the greater the number of plausible alternative explanations, and the more plausible these are, the lower the degree of reliability that we should assign to this report.⁴ As I will argue later, the fact that it specifies a method of drawing conclusions about the source’s reliability is one of the main benefits of TIBE compared to alternative frameworks (see Sect. 4). Here are some examples of the kind of explanatory inferences that I have in mind:

This testimony comes from a credible source because it makes sense that the witness actually saw what happened and that they are willing to tell us. There are no alternative explanations for their utterance that make sense in this context.

This testimony comes from somewhat credible source, because while this statement is well-explained by the witness being truthful, it is also plausible that they are misremembering.

This testimony does not come from a credible source, there are various alternative explanations why the report might say this. For instance, it is plausible that the witness is lying.

The focal point of eyewitness evaluation on this account is therefore the generation and subsequent evaluation of potential explanations for the report. For instance, suppose that, at trial, the prosecution wants to argue for the reliability of a particular witness with respect to their statement that “the robber wore a red shirt”. They could do so by first putting forward a scenario in which the robber wore such a shirt and the witness accurately observed this fact and then arguing that this scenario is plausible in the given case. The defense might then respond by offering an alternative explanation of the witness’s utterance, for instance by arguing that it is plausible that they are lying. It is then up to the decision-maker (i.e., the judge or jury) to determine

³ As we will see in Sects. 4 and 6, TIBE’s open-endedness with respect to the explanations that we consider sets it apart from alternative prescriptive approaches and arguably allows for more flexibility.

⁴ Some might object that this is not inference to the best explanation (e.g., Laudan 2007). After all, we do not infer that the best explanation is true. Instead, we compare the disjunction of available truth-telling explanations to the disjunction of available alternative explanations. However, I believe that we should not be too nitpicky about our use of this term. What I propose may not be IBE in a strict sense, but it is nonetheless explanatory reasoning. To call such reasoning IBE is not uncommon (e.g., Pardo & Allen 2008).

whether the parties have succeeded in showing that these respective are plausible in the given case.

Much ink has been spilled by both philosophers of science and law on what makes an explanation good (or ‘plausible’, or ‘lovely’). In a general sense, an explanation is good to the extent that – if it were true – it would help us understand what caused the witness to offer this particular story. Explanatory goodness is usually spelled out further in terms of specific explanatory virtues. To give an example, Pardo & Allen (2008, 230) suggest that:

All other things being equal [an explanation of legal evidence] is better to the extent that it is consistent, simpler, explains more and different types of facts (consilience), better accords with background beliefs (coherence), is less ad hoc, and so on; and is worse to the extent that it betrays these criteria.

This is merely one among multiple possible explications of explanatory goodness.⁵ One issue that such characterizations tend to share is that they are quite vague. This vagueness about what makes an explanation ‘good’ is arguably the weakest point of most accounts of IBE, both in the philosophy of science and in the law. Whether it is even possible to be more precise about this is a thorny issue that I do not want to solve here.⁶ For my current purposes it is sufficient to note that an explanation’s goodness mostly depends on the extent to which it fits well with the available evidence and our background beliefs about the world. To begin with the latter, our background beliefs concern “the way things normally happen” (Walton, 2007, 128). An example relating to testimony is that ‘humans generally cannot hear soft sounds from a great distance.’ These folk-psychological generalizations can (and should) be partially based upon the extensive research on eyewitness reliability.⁷ The relevant evidence may include facts about the witness (e.g., was their eyesight good?) and about the situation (e.g., did it rain?). Furthermore, the content of the witness’s testimony may be relevant. For instance, if someone offers a detailed description of an event, but we also know that they were watching this event from a large distance, then the explanation that they are telling the truth becomes less plausible. In general, a witness who makes a surprising claim should – all else being equal – be regarded with more suspicion than one who

⁵ For instance, in the story model of legal proof, how good an explanation is depends on how much of the evidence the explanation accounts for, whether it contains internal contradictions and problematic gaps, whether it has all its parts and how well it fits with our knowledge of the world (Pennington & Hastie, 1992, 190–199). Josephson (2000, 1627) mentions the following desirable features of explanatory hypotheses: consistency, simplicity, explanatory power, predictive power, precision, specificity, and theoretical promise. Others, such as Thagard (2005) propose still other criteria.

⁶ Much more could be said about the difficulty of spelling out explanatory goodness. For instance, some authors such as Bovens & Hartmann (2003) give ‘impossibility results’ for explanatory virtues like coherence. These are mathematical proofs that suggest that more virtuous hypotheses are not necessarily more probable than less virtuous hypotheses. There are also philosophers of science who point out that historically, less virtuous (e.g., more complex) scientific theories regularly prevailed over more simpler ones – and that we should therefore be skeptical about the project of spelling out virtues that reliably indicate true hypotheses (Cabrera, 2017). Given such obstacles, I am hesitant to attempt to be precise about what makes an explanation ‘good’.

⁷ For reviews, see, for example, Dubelaar (2014); Ross et al., (2014); Wise et al., (2014).

reports on run-of-the-mill facts. So, even though TIBE is about the reliability of the witness, rather than the credibility of their claim, judgments about the two may go hand-in-hand.

Not every instance of testimony calls for inference to the best explanation. Thagard (2005) suggests that, in daily life, people do and should accept the testimony of others by default, *unless* the content of this testimony conflicts with their beliefs. If that is the case, they enter a ‘reflective pathway’ where they evaluate the witness’s claim through inference to the best explanation. As Thagard (2005, 299) points out, such a strategy is epistemically useful, because while we do not want to uncritically accept everything that anyone tells us, if we’d carefully reflect on everything that we’re told, we would be greatly restricted in acquiring new beliefs. I want to suggest something similar for criminal cases. In such cases there is arguably also a tendency by jurors and judges to believe witnesses by default (e.g., Brigham & Bothwell 1983). In contrast with daily life, this default trust is regularly unwarranted, leading to judicial errors. So, there are many situations in which critical evaluation is needed and where we should engage in TIBE. Nonetheless, this does not mean that eyewitness testimony should *always* be subjected to scrutiny. Witnesses may report on mundane facts that neither the defense nor the prosecution seek to challenge. In such instances, careful explanatory reasoning about the witness’s reliability would only obstruct the efficiency of the trial. Decision-makers may instead reserve their scrutiny for contentious, surprising statements, and for types of statements where eyewitness errors are common (e.g., an identification of the perpetrator of a violent crimes by a bystander).

Neither Thagard, nor anyone else has offered arguments why TIBE is the preferred approach once we enter this ‘reflective pathway’. Yet explanatory reasoning has numerous benefits. For instance, a well-known advantage of thinking in terms of competing explanations is that it may counteract confirmation bias by preventing decision-makers from becoming overly focused on a single possibility (O’Brien, 2009). Confirmation bias also poses a danger when interpreting eyewitness testimony. As mentioned, decision-makers will often trust eyewitness testimony too easily, ignoring any evidence for unreliability. At other times, they may overly focus on evidence contra reliability, leading to a deflated assessment of the witness’s reliability (Puddifoot, 2020). Rassin (2001, 2014) suggests that eyewitness evaluators may avoid confirmation bias by considering competing explanations for an eyewitness report. This forces evaluators to explicitly think about both the possibility that the witness is reliable and the possibility that they are not.⁸ He also notes that such explanatory comparison may lessen the degree to which people tend to draw extreme, unwarranted conclusions (e.g., ‘this witness is completely (un)reliable’). According to Rassin, explanatory comparison instead leads to more cautious conclusions such as ‘some alternative explanations need to be further investigated’ or ‘there are no plausible alternatives’. I discuss various other benefits of explanatory thinking throughout this article, including in the following section.

⁸ Though my account was developed without knowledge of Rassin’s work, his proposal is similar to mine as it also asks evaluators to compare competing explanations of a report. His account is aimed at the Dutch legal context and intended to be used by expert eyewitness evaluators. Rassin (2014) writes that his approach and similar approaches have successfully been used by experts in practice.

2.2 Storytelling about testimony

When we think about why an eyewitness offers their testimony, the explanations that we take into account may be general. However, such general explanations (e.g., ‘the witness is lying’) will often be difficult to evaluate as they lack detail. So, ideally the explanations that we consider should be more specific stories. For instance, Walton (2007, 109–10) has us imagine a situation in which the testimony of two police officers confirms a suspect’s alibi. He writes: “[T]hrough it may be conceded that generally police officers in the line of duty do not lie, there may be evidence in this specific case showing that in fact these two police officers did lie. This could be shown by means of a (...) narrative showing the goals of the police officers and the other physical and psychological conditions of the case.”

I am using the word story in a broad sense here. In the legal context, words like ‘narrative’, ‘story’ or ‘scenario’ are typically used in the context of the *story model* of evidential reasoning. According to this model, decision-makers come to factual conclusions by comparing competing stories that explain the evidence (Pennington & Hastie, 1993). Such stories have a specific structure: they contain a central action (e.g., a person killing someone else) and describe a context that make this action understandable. Other elements of a complete story are a description of the scene, a motive, a central actor and resulting consequences. When we describe why a person offers a piece of testimony, our explanation may have such a story structure. However, it may also take other forms. But even if a detailed account of some testimony does not have the aforementioned structure, we are - colloquially - telling a narrative about why this person offers this piece of testimony. For example, suppose that we consider the explanation that the witness misremembered. We could then further specify when and why the distortion in their memory likely occurred by considering the particulars of this specific witness and the context in which they observed the reported events. This leads to a story of why it is plausible that this witness misremembered these events.

As has often been noted, by constructing, critiquing and comparing competing narratives, decision-makers in criminal cases can structure their reasoning about the available evidence (Walton, 2007). Similarly, thinking in terms of competing narratives that explain a witness’s testimony can be of enormous help to evaluators. In particular, the more specific our explanations are, the more specific our expectations become about the kinds of evidence that we would, or would not expect to encounter. For example, it may be difficult to determine how we might falsify or verify the explanation that ‘the witness is lying’. In contrast, suppose that our explanation is that ‘the witness is lying about being at home when the crime took place, because they were actually at the crime scene and are trying to cover up this fact.’ Such a specific explanation will generate predictions that we could test. For example, that we would likely find trace evidence left behind by the witness at the crime scene or that certain other witnesses would likely contradict the witness’s alibi.⁹

⁹ That it helps us select, structure and interpret the available evidence is a key benefit of TIBE compared to existing frameworks. See Sects. 4–6.

The process of making our explanations more detailed will typically begin by considering what a given explanation would plausibly entail in the given circumstances if it were true. For instance, suppose that these two police officers lied, what would be the most likely reason why they did so? How might they try and hide the fact that they are lying? As Walton (2007, 128) points out, this process of filling in the missing bits is itself an instance of inference to the best explanation. Though there are countless ways of filling in the gaps in a story, we choose the version that is the most plausible in the given context.

Truth-telling explanations can also be made more specific. We may start with the general explanation that ‘the witness is accurately reporting on the events that took place’. This can then be turned into a full-fledged story by filling in the details of where the witness was standing when the event took place, how much time they had to observe the event, and so on. During a trial, one of the parties may argue that the witness is reliable by putting forward such a story. The opposing party may then cast doubt on the witness’s reliability by attacking this story. For instance, they can ask critical questions; or, argue that the story does not fit with the facts of the case or with certain background beliefs. They may also sow doubt about the witness’s reliability by presenting a plausible alternative story in which the witness does not speak the truth. It is then up to the decision-maker to decide which story best explains the witness’s utterance. To give an example of what such an exchange may look like in practice, consider the following example:

A defendant is accused of smuggling drugs by two witnesses. In exchange for getting their sentences for another crime reduced, these witnesses testify that they smuggled drugs across the border together with the defendant many times. The prosecution argues that the witnesses should be believed, as their testimonies cohere with one another and with the other facts of the case. The defense, in contrast, argues that the witnesses are not credible. They sketch an alternative scenario in which the witnesses lied. In this scenario, the agreement between the testimonies is explained as being the result of the witnesses fitting their stories to each other and to the case file. The defense points out that the testimonies agree with one another to a suspicious degree and are very detailed given that the events took place years ago – indicating that they were fabricated. The prosecution tries to show that this alternative scenario is not plausible, by arguing that the witnesses also reported on verifiable details that were not in the case file and that their stories also differed in places. The defense, in turn, counters this argument by showing that these verifiable details were easy to guess, even if the witnesses were lying.¹⁰

We can imagine such an exchange continuing for quite a while, with the prosecution trying to show that the alternative scenario is implausible, while the defense tries to argue that it is plausible. As the exchange goes on, the scenario can become increas-

¹⁰ This example is a highly simplified version of the arguments offered during a criminal trial which took place in the Netherlands in 2015. HR 29 september 2015, ECLI:NL:HR:2015:2842. I take this example from Stevens (2016).

ingly specific as the parties drill down into specific details that they believe make the story plausible or implausible. These can be facts from the case but, for instance, also psychological research on how well people tend to remember specific details of events after a given amount of time. Additionally, the interpretation of available facts can also change. For instance, in this example, the agreement between the witness's testimonies is used by the prosecution to argue for the reliability of the witnesses, while the defense uses it as an argument for their unreliability. As I will discuss in Sects. 5 and 6, TIBE explains not only how facts may become relevant, but also the phenomenon that facts can be interpreted in competing ways. After the exchange is completed, the decision-maker can decide how plausible each of the competing scenarios is, and assign a degree of reliability to the witnesses accordingly.

3 Being friends with Bayesians

Having laid out the basics of TIBE, I now want to turn to its rationality. My account is intended to be *prescriptive*. It offers evaluators guidance on how to reason in practice about the reliability of a particular report (e.g., what kind of questions they should ask and how they should answer those questions). We may contrast such an account with descriptive frameworks, which describe how people evaluate eyewitness reliability in practice. Existing accounts of TIBE are mostly of the descriptive kind: they are intended to capture our intuitions about why we usually, but not always, trust the utterances of others in daily life (Lipton, 2007; Malmgren, 2006; Fricker 2017). Another type of account is *normative*. Normative accounts of eyewitness reliability formulate a standard for epistemically rational evaluation. A good prescriptive account helps real-life decision-makers be (more) rational in their evaluations. Such an account therefore lies at the intersection of well-grounded descriptive and normative accounts: it should fit with how people normally reason¹¹, but there should also be good normative reasons to believe that the approach helps us be more accurate.

Little has been said about TIBE's normative basis. In this section I want to make some headway on this matter by connecting my account to Bayesianism. The latter has been the dominant account of epistemic rationality over the past decades (Talbot, 2008). Bayesianism's influence also extends to legal proof scholarship and to the study of witness reliability (Fenton et al., 2016; Merdes, von Sydow & Hahn, 2020). I focus on the work of Goldman (1999). Goldman is neither the first nor the last person to offer a Bayesian analysis of witness reliability. However, he is, to my knowledge, the author who most extensively discusses *why* Bayesianism provides the correct normative standard for thinking about this topic. I argue that Goldman's work also offers a basis for TIBE's rationality. In particular, I follow a line of thinking from the philosophy of science which casts IBE as a useful heuristic for optimal Bayesian reasoning. On my view, Bayesianism describes the goal at which we aim

¹¹ Though I will not defend TIBE's descriptive adequacy here, it is at least plausible that it is adequate in this sense. First, it is well-known that how people tend to make sense of criminal evidence through explanatory reasoning (e.g., Pennington & Hastie 1993). Second, the aforementioned descriptive accounts of TIBE have (to my knowledge) not been met with much resistance.

when we engage in rational eyewitness evaluation. However, I also argue that it is not by itself an adequate account of rational eyewitness evaluation. For this, we should turn to TIBE.

3.1 Bayesian rationality

Let us begin with the basics of the Bayesian framework. Bayesianism is first and foremost an account of how we should change our beliefs upon receiving new evidence. Suppose that we are interested in the status of some hypothesis H . Furthermore, suppose that we receive a piece of testimony that is relevant to this hypothesis. We then update our prior belief in H conditional on this report to arrive at a posterior probability of this hypothesis. How much this report heightens (or lowers) our prior belief in H depends on the *likelihood ratio* of the testimony:

$$\text{Probability}(\text{Witness report} \mid H)$$

$$\text{Probability}(\text{Witness report} \mid \neg H).$$

In other words, how much would we expect the witness to give this report, if H is true, and how much would we expect the witness to offer this testimony, if H were false? For instance, is it likely that the witness would say that the robber wore a red shirt if the robber wore such a shirt? What about if he did not wear such a shirt, how likely would it then be that he said this? The higher this ratio of true to false positives, the more reliable the witness is.¹²

The likelihood ratio is central to how the Bayesian account expresses eyewitness reliability. But there are different ways to interpret the probabilities in this ratio. Possible interpretations form a continuum from objective (i.e., ‘probability is in the world’) to subjective (i.e., ‘probability is in the head’) (Redmayne, 2003). Which interpretation we choose matters a great deal both for how we should assess this ratio and for Bayesianism’s normative status. For instance, if we choose an interpretation that is too subjective, what counts as a ‘rational’ eyewitness evaluation becomes too much a matter of personal belief. Such a view would allow a vast range of evaluations for any particular eyewitness, even ones that are blatantly irrational (e.g., those of a racist evaluator who considers all eyewitnesses of a certain race to be untrustworthy). A more objective probability interpretation would set further limits on what counts as a rational assessment of the likelihood ratio.

If we use an objective interpretation, then there is also a strong argument to be made for Bayesianism as the appropriate normative account of eyewitness evaluation. In particular, Goldman (1999, 116–33) offers a mathematical proof which shows that, if our subjective likelihoods of a report match its *objective* likelihoods, then Bayesian updating will always bring our posterior degree of belief in the hypothesis that the

¹² Note that I presume here that the hypothesis H under consideration is the same as the hypothesis that the witness reports on. This is not necessarily the case. For instance, we might consider the evidential impact of the witness’s testimony about the colour of the robber’s shirt on the hypothesis that ‘Edward is the robber’.

witness is reporting on closer to the truth compared to the prior. This result is part of a broader class of *veritistic* arguments for Bayesianism which purport to show that, if we update our beliefs using accurate likelihoods, our beliefs will (eventually) converge on the truth (Hawthorne, 1994). An account of rational eyewitness evaluation that can draw on such arguments has a strong claim to a normative status.

The central difficulty with this argument is finding the correct ‘objective’ interpretation. Various of the most promising proposals fail. Indeed, some have doubted that a workable objective interpretation exists for testimony, where clear-cut probabilities are often unavailable (Fallis, 2002). For instance, the most well-defined objective interpretation is the frequentist, where we measure probability as occurrence in a specified population of events. A typical example of this type of probability is the frequency with which a coin lands on heads out of n tosses. Some authors use this interpretation with respect to witness testimony. They suggest that a reliable eyewitness is one that has a tendency to speak the truth on the relevant topic. This tendency could be determined by looking at the witness’s *track record* – i.e., the number of times the witness has spoken truths versus falsehoods about this topic. However, as Thagard (2005) argues, this view is deeply problematic. First, we almost never have reliable records of a person’s track record. Second, this view leads to the wrong conclusions. For instance, we should not trust a person completely, no matter what other evidence we have, just because they have so far been accurate about some topic.¹³ Finally, as Coady (1992, pp. 210–11) argues, this interpretation relies on the mistaken view that “[people] have quite general tendencies to lie, whatever the context or subject matter, [and] to make mistakes in abstraction from particular circumstances.”¹⁴

Goldman (1999, 117) also rejects such a frequentist interpretation. He goes on to argue that objective probabilities concerning eyewitness reliability nonetheless exist (Goldman, 1999, 118):

[I]n testimony cases it looks as if jurors, for example, work hard at trying to get accurate estimates of such probabilities, which seems to presume objective facts concerning such probabilities. If the witness in fact has very strong incentives to lie about X, this seems to make it objectively quite probable that she would testify to X even if it were false. If the witness has no such incentives,

¹³ We also cannot save frequentism by interpreting the relevant probabilities as the tendency of witnesses in general to speak the truth under similar circumstances. Though the research on eyewitness reliability is extensive, we lack the required frequency data on eyewitness reliability in such specific sets of circumstances. Additionally, this suggestion faces the well-known *reference class problem*. For any individual witness there are different ways to define what counts as a witness under ‘similar circumstances’, without any particular reference class having a claim to being the ‘correct’ one.

¹⁴ Thagard (2005, 308), who defends TIBE, appears to fall into this trap. He suggests that assessing a witness’s credibility “requires a judgment about a person’s disposition to tell the truth on a particular topic.” He illustrates this point using the example of a police officer in the infamous O.J. Simpson case who was accused of being racist and whose testimony was not believed by the jury. According to Thagard, the jury inferred that the police officer had a disposition not to tell the truth in cases where a black person was accused of a crime and decided not to trust him on that basis. But even if we grant that this inference about the witness’s disposition best explains the jury’s decision in that particular instance, this point does not generalize to all instances of eyewitness evaluation. There may also be cases in which a generally reliable witness, who usually tells the truth, has to observe something in poor conditions. It would be strange to say that this person has a *disposition* to be mistaken.

nor any disabilities of perception, memory, or articulation [...] then the objective probability of her testifying to X even if it were false seems to be much lower.

What Goldman seems to be talking about is an *evidential* probability interpretation, where the probability of a hypothesis is the degree to which our evidence justifies us in believing that hypothesis. Hacking (2001, 131-3) suggests that the best way to interpret this type of probability is as how much a reasonable person would believe in H given the evidence. Alternatively, we might think of them as probabilities that are based on good grounds (Tang, 2016). What is important is that, though an evidential probability is subjective in the sense that it is a person's degree of belief, it is objective in the sense that it is not purely a matter of personal opinion. Various authors consider the evidential interpretation to be the most suitable 'objective' interpretation for the context of legal evidence (Nance, 2016, 47–8; Wittlin 2019; Spottswood, 2019; Hedden & Colyvan, 2019).

The most important objection to the evidential interpretation is that it is vague. Goldman does not spell out what it means for evidence to make something objectively probable and it may actually be impossible to offer a clear definition. For instance, Redmayne (2003) surveys various ideas that Goldman could draw upon for more precision, but concludes that they would all lead to interpretations that are either too limited in scope to capture the evidential richness that we want, or so vague that they are not much more informative than Goldman's own brief description above. However, this vagueness does not have to be lethal. For instance, Williamson (2002, 212), one of the most prolific proponents of the evidentialist interpretation, defends his own refusal to make this notion more precise by noting that to ask for too precise a definition may lead to never getting ahead with the matters at hand. He writes: "Sometimes the best policy is to go ahead and theorize with a vague but powerful notion" (212).

A similar response is possible for the evidential interpretation in the context of eyewitness evaluation. We do not need spell out precisely what it means for the evidence to objectively support a proposition if this is sufficiently clear in practice. As Goldman points out, we can often intelligibly ask: Does the available evidence make it probable that this witness would testify that X, if X were true? And does the available evidence make it probable that this witness would testify that X, if X were false? Furthermore, it is easy to imagine situations where a specific evaluator's assessment of an eyewitness's reliability will clearly not be reasonable given the evidence. For instance, consider two witnesses. The first is well-known to be honest. They have good vision and an excellent memory. When the crime occurred, they had ample time to observe the perpetrator. Anyone who assigns such a witness a low degree of reliability (without offering further, convincing reasons) is patently irrational, because their belief goes against the evidence. In contrast, consider a witness who is a compulsive liar, who has a motive to lie in this particular instance and whose testimony conflicts with much of what is known. It would be equally irrational for anyone to ascribe a *high* likelihood ratio to this person. In this way, Bayesianism constrains our evaluations of eyewitness reliability.

The above suggestion does, however, presume that evaluators have a way of determining what conclusions the evidence supports with respect to the two conditional probabilities in the likelihood ratio. However, this raises a problem for the Bayesian. In the examples that Goldman gives, determining how the evidence influences the eyewitness's reliability is a matter of applying commonsense generalization. An example of such a generalization is that having an incentive to lie clearly makes one more prone to report falsehoods. It is only common sense that discovering that someone has such an incentive should lower our assessment of their credibility. However, as I will set out in detail in Sects. 5 and 6, things are not always so simple. Evaluators often face the difficult tasks of deciding both which evidence to consider and which generalization to apply to each item of evidence. Bayesianism does not tell us how to go about these tasks. If we adopt the evidential probability interpretation of the likelihood ratio then this is a glaring gap. The slogan 'look at the evidence' is not a very informative statement for eyewitness evaluators when it is unclear what the evidence is or in what light we ought to see it. So, the Bayesian account is not a very informative theory of rational eyewitness evaluation by itself.

This then finally brings us back to TIBE. I propose that the Bayesian account clarifies the *aim* of rational eyewitness evaluators. However, it does not give a method for how to achieve this aim. What is needed is a prescriptive account that tells evaluators how to make sense of the available facts. I argue that TIBE succeeds as such an account on two fronts. First, it helps with the two aforementioned tasks: selecting the relevant evidence and interpreting this evidence. Second, as I will now argue, TIBE's conclusions quite straightforwardly track the Bayesian likelihood ratio, meaning that TIBE leads to rational conclusions from a Bayesian viewpoint. It therefore has a plausible claim to being rational (or, at the very least to being no less rational than Bayesianism).

However, before I move on to how TIBE tracks the likelihood ratio, I want to briefly comment on how my account relates to the general debate on whether Bayesianism and IBE are compatible in legal settings. The account that I defend shares some similarities with that of Hedden & Colyvan (2019) who suggest that we can fill in the Bayesian notion of evidential probability using explanationist means. However, for them this means that explanationism is *subsumed* under Bayesianism – with the latter being the proper framework of rational proof. Allen (2020) criticizes this suggestion. He argues that it is explanationism that is doing the hard work of processing and deliberating on evidence, with Bayesian probability only being used to attach numbers to the result of this deliberation. According to Allen (2020, 122) it is therefore Bayesianism that is turned into a species of explanationism rather than the other way around. My own position is in between these two suggestions. On the one hand I agree with Allen that if explanationism is doing all the hard work, then it is far more than merely a tool for spelling out Bayesianism, and arguably has a better claim to being the proper account of rational proof. However, I disagree with him that Bayesianism is of no use for the explanationist. Rather, on my view, Bayesianism is a useful method for making precise what aim we seek to achieve on TIBE and why and when this account achieves this aim.

3.2 Heuristic compatibilism

Though not much has been said about how explanation-based reasoning and Bayesianism relate in the context of eyewitness evaluation, various suggestions about this relationship have been offered in other contexts, especially with respect to scientific theory choice. One well-known strand of thought casts inference to the best explanation as a *heuristic* to approximate correct Bayesian reasoning (Okasha, 2000; Lipton 2003; McGrew, 2003; Dellsén, 2018). A heuristic is a method of reasoning that is efficient and tends to lead us to the approximately right answer. On this view, IBE complements Bayesianism by offering a rule of inference that is appropriate for non-ideal agents, yet enables these agents to approximate the probabilities that Bayesian reasoning would have them assign to hypotheses (Dellsén, 2018). I suggest that, similarly, TIBE is a heuristic for approximating the Bayesian likelihood ratio. In what follows, I give an argument for why the approach usually tracks this ratio, but makes sacrifices with respect to accuracy, for the sake of efficiency and respecting human cognitive limitations. In particular, rather than consider the entire probability space, this approach tells us to evaluate and compare a limited number of salient, well-specified explanations, and to ignore other possibilities.¹⁵

To make it more precise under which conditions TIBE tracks the likelihood ratio and when it diverges from this ratio, it will help to look at recent, Bayesian network models of eyewitness reliability.¹⁶ These approaches start from the idea that the likelihood ratio leaves much information about the reliability of the witness implicit, whereas we might want to represent this information explicitly. For instance, we may want to model the impact of receiving evidence concerning the report's reliability. In these models, the reliability of the report is therefore represented as a distinct variable, rather than being encapsulated in the likelihood ratio. This variable is called a 'reliability node', usually denoted as REL, which expresses the hypothesis (HYP) that the witness's report (REP) is accurate (Bovens & Hartmann, 2003, 57; Lagnado et al., 2013; Merdes et al., 2020). The idiom looks as follows:

Once we include this variable, we end up with the following formula for calculating the likelihood ratio:

$$\frac{P(\text{REP}|\text{H}) P(\text{REP} | \text{H} \ \& \ \text{REL})P(\text{REL})+P(\text{REP} | \text{H} \ \& \ \neg\text{REL})(\neg\text{REL})}{P(\text{REP}|\neg\text{H}) P(\text{REP} | \neg\text{H} \ \& \ \text{REL})P(\text{REL})+P(\text{REP} | \neg\text{H} \ \& \ \neg\text{REL})P(\neg\text{REL})}$$

According to Lagnado et al., (2013, 52), by including REL, we make it explicit that there can be "alternative possible causes of a [statement other than it being truthful]".¹⁷

¹⁵ This contrast between explanation-based and Bayesian approaches has also been noted by Allen & Pardo (2019).

¹⁶ Such approaches comprise the bulk of contemporary Bayesian work on witness reliability. These networks are used to explore the rationality of what sorts of changes in evidence increase or decrease the credibility of a witness and their statement and to what extent (see Merdes et al., 2020 for a review).

¹⁷ They divide the reliability node up into more specific causes (see Sect. 4.1 for more on this).

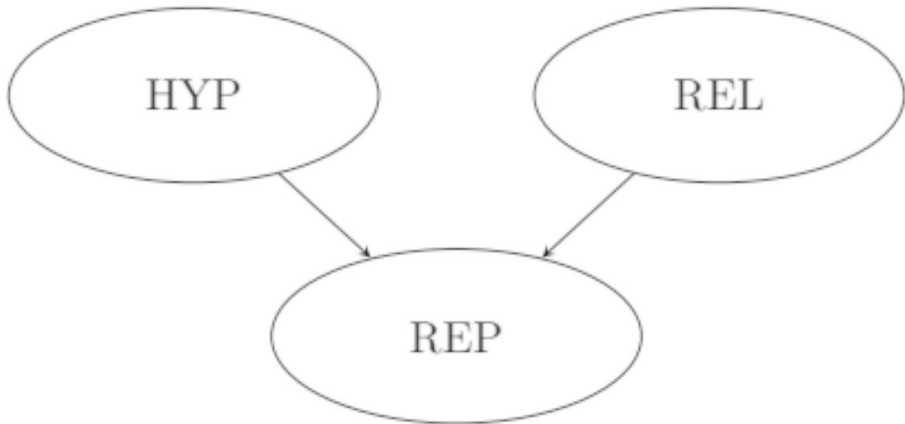


Fig. 1 The evidence reliability idiom

In this article, I also emphasize the usefulness of explicitly reasoning about competing possible causes of a testimony. Remarks like these therefore highlight the natural fit between Bayesianism and explanation-based reasoning.

As said, TIBE does not always perfectly track the likelihood ratio. It may deviate from this Bayesian formula due to the limited number of explanations that we consider. The above formula helps make it more precise when such deviation occurs. The first limitation on the explanations that we consider follows from the fact that, on TIBE, we want to draw a conclusion about whether the witness is reliable with respect to the facts that they report on. This means that hypothesis H is assumed to be the same as (part of) the content of the witness's report. For instance, if the witness reports that "the robber wore a red shirt", then hypothesis H is 'the robber wore a red shirt'. Though the formula above does not dictate this, it is an assumption that is made by many of the Bayesian authors as well. It is also a useful assumption, because it leads to a simplification of the formula. First, it means that we can set the value of $P(\text{REP}|\text{REL} \ \& \ \neg H)$ to 0. On the accounts of Bovens & Hartmann (2003) and Lagnado et al., (2013) a reliable witness is a truth-teller. It is therefore, for instance, impossible that (a) the witness reports that the robber wore a red shirt, (b) that this statement is false, and (c) that the witness is reliable. In addition to this, I propose that, on TIBE, we may *also* ignore $P(\text{REP} \mid \neg\text{REL} \ \& \ H)$. Admittedly, it is possible for an unreliable witness to accurately report on the true state of the hypothesis. For instance, Bovens & Hartmann (2003) as well as Lagnado et al., (2013) presume that a fully unreliable witness is a randomizer. Such a randomizer could accidentally report the truth. For example, the witness could be a liar who makes a random statement about the colour of the robber's shirt and happens to pick the right one. However, such epistemic luck

will often (though, admittedly, not always) be highly unlikely.¹⁸ A lucky randomizer hypothesis will therefore usually not be part of the most salient explanations.¹⁹

If we ignore the aforementioned terms, we can simplify the formula for calculating the likelihood ratio as follows:

$$\frac{P(\text{REP} | H) P(\text{REP} | H \ \& \ \text{REL}) P(\text{REL})}{P(\text{REP} | \neg H) P(\text{REP} | \neg H \ \& \ \neg \text{REL}) P(\neg \text{REL})}$$

My proposal is that TIBE maps onto this simplified formula. To be precise, the disjunction of our truth-telling explanations maps onto the numerator, whereas the disjunction of the alternative explanations maps onto the denominator. Let us focus on the truth-telling explanations first. A truth-telling explanation is a narrative in which the relevant hypothesis is true and the witness accurately reports on it. In other words, this is a situation where both H and REL are true and the witness testifies that ‘H’.

To offer a plausible truth-telling explanation means to offer a *ceteris paribus* reason to assume that both $P(\text{REP} | H \ \& \ \text{REL})$ and $P(\text{REL})$ have a high value. First, as various philosophers of science have noted, all other things being equal, when one hypothesis explains a fact better than a competitor, then the fact is also more likely to occur given this hypothesis than given its alternative (McGrew, 2003; Henderson, 2014). To give a good explanation of a fact is – at least in part – to show that this fact is expected if the explanation is true. Similarly, a narrative that explains the witness’s utterance well will also make this utterance likely. In other words, if we were to presume that a plausible truth-telling explanation is true, then we may assign a high value to $P(\text{REP} | H \ \& \ \text{REL})$. However, this is not all there is to plausibility. Consider the truth-telling explanation that ‘a powerful, all-knowing extraterrestrial mind-controlled the witness so that this witness offered a perfectly accurate statement.’ If we presume that this explanation is true, then obviously this makes it likely that the witness would be accurate in their report. However, such an explanation is also highly implausible, as it conflicts with background beliefs that most of us hold about the world. To offer a plausible explanation therefore also means to offer a reason to believe that this witness was indeed reliable in this way. For instance, we may sketch a plausible situation in which this particular witness could have accurately observed these particular events, which gives us a *ceteris paribus* reason to presume that this *is* the correct explanation of the witness’s testimony. Hence, to show that a truth-telling explanation is true also means to show that the value of $P(\text{REL})$ is *ceteris paribus* high.

The more plausible our truth-telling explanations are, and the more of these explanations we have, the higher the value is that we may therefore assign to $P(\text{REP} | H \ \& \ \text{REL})P(\text{REL})$. Similarly, the disjunction of alternative explanations tracks the denomi-

¹⁸ Additionally, randomizers may be easy to spot in practice, as we would expect them to make many false statements in their reports. This is another reason why we need not include this hypothesis in our assessments.

¹⁹ Friedman (1986, 688), a proponent of the Bayesian account of eyewitness evaluation has similarly suggested that we may ignore such unlikely possibilities. He gives the example of a witness who misperceives an event, then misremembers what they misperceived, and ends up telling the truth.

nator of the formula. An alternative explanation is a narrative in which the relevant hypothesis is false, but the witness nonetheless reports that it is true because they are unreliable. They might, for example, misremember. To offer plausible alternative explanations is therefore to offer reasons to assign a high value to both terms in the denominator, $P(\text{REP} \mid \neg H \ \& \ \neg \text{REL})$ and $P(\neg \text{REL})$. The argument for this is the same as for the numerator. Conversely, to argue that there are no plausible alternative explanations is to argue that we may assign a low value to this part of the formula.

However, our explanatory reasoning may diverge from optimal Bayesian reasoning in a second way. The key feature of TIBE is that the truth-telling and alternative explanations that we consider are specific and small in number. For instance, the prosecution may offer one version of why this witness is telling us what they are and the defense another. However, this means that TIBE diverges from optimal Bayesian reasoning when we overlook other plausible explanations. In other words, TIBE faces its own version of the well-known *bad lot problem* for IBE. Briefly put, the bad lot problem is that we are not justified in concluding that the best explanation is (probably) true if we are insufficiently certain that the true explanation is among those that we've considered. As I have argued elsewhere, both IBE and Bayesian inference face this problem (Jellema, 2022). Nonetheless, this does not take away that the fewer explanations that we consider and the more specific these explanations are, the bigger the risk that we overlook plausible alternatives. This is a strength of TIBE, because such specific explanations are often easier to evaluate than general ones and because it is unrealistic to presume that evaluators can consider every single possibility. However, it is also a weakness, as it limits the scope of inquiry, thereby making it more likely that we overlook other plausible explanations. When we miss such alternatives, we misjudge the relevant likelihoods. The evaluator must therefore engage in a balancing act between efficiency and accuracy.²⁰

To sum up the above, I have sketched a novel, feasible way in which testimonial inference is compatible with Bayesian accounts of eyewitness evaluation. On my proposal, we aim for a Bayesian norm, but we do so through the efficient but imperfect heuristic of explanation-based reasoning. The upshot of this proposal is that we may use arguments for the normativity of Bayes as a foundation for the rationality of TIBE. I discussed Goldman's (1999) argument above, which I view as the most prominent of these arguments. However, as argued earlier, this does not mean that TIBE is subsumed under Bayesianism. What we want to know from a theory of rational eyewitness evaluation is how we ought to determine what conclusions the evidence supports, which explanatory reasoning helps us do. It is this aspect of TIBE that I discuss in the remaining sections.

²⁰ There are specific aspects of criminal cases that arguably make it more likely that we overlook plausible alternatives (Jellema, 2022). In particular, the likelihood that we overlook such relevant explanations depends not only on the specificity of our conceived alternatives, but also on how good those alternatives are. Additionally, the quality of our investigation and our set of evidence, as well as the complexity of the case at hand also matter. To what extent TIBE helps us accurately assess the likelihood ratio will depend in part on these factors.

4 Comparing prescriptive accounts

Having spelled out TIBE's normative basis, I now turn to how this account compares to existing work.²¹ In particular, I want to juxtapose TIBE with two well-known prescriptive approaches to eyewitness evaluation: (i) the capacity approach, which has been defended by a number of epistemologists, and (ii) empirically-informed methods developed by psychologists. This discussion will help explain what is distinctive about TIBE, but also how it fits with this work. In particular, I want to suggest that TIBE can be a helpful addition to these existing accounts and, conversely, that ideas from these frameworks can help fill in some of the details of TIBE.

4.1 The capacity approach

The first account that I discuss is what I call the 'capacity approach'. This is arguably the best-known prescriptive account of eyewitness evaluation in the philosophical literature. On the capacity approach, we evaluate a witness's reliability by considering whether the witness's statement came about through an adequate exercise of certain capacities. A wide range of scholars adopt this approach, including a number of Bayesians, who use it to make reasoning about the evidential impact of testimony more tractable (Friedman, 1986; Schum, 1994; Goldman 1999; Lagnado et al., 2013). There is some difference in the capacities that various authors distinguish. Schum (1994) gives the most commonly used list, on which an eyewitness's reliability depends on whether they were (a) observationally sensitive, i.e., their senses functioned correctly, (b) objective, i.e., their memory aligns with what they perceived, and (c) veracious, i.e., they truthfully report what they believe. So, when we assess the likelihood ratio of the witness, we should ask critical questions such as:

1. 1. Is the witness sincere?
2. 2. Did the witness's memory function properly?
3. 3. Did the witness's senses function properly?

(Walton, Reed & Macagno, 2008).

The capacity approach can be seen as a kind of inference to the best explanation, where we consider multiple alternative explanations of the witness's utterance: lying, misremembering, misperceiving. This way of looking at the capacities fits with how various of its proponents present the approach. For instance, Lagnado et al., (2013) utilize the capacity approach in their Bayesian network-based modeling of eyewitness evidence. As said earlier, in their model, eyewitness reliability is expressed through a 'reliability node' that represents the "alternative possible causes of a [statement]" (52). By 'alternative', they mean other explanations than 'the statement is true'. They

²¹ Lipton (2007, 246-8) similarly contrasts TIBE with a number of alternative accounts from the philosophy of testimony, such as evaluating the witness's reliability based on their reputation, in order to show that it is a non-trivial approach. However, as far as I am aware, none of the approaches that he discusses are used in contemporary legal practice. In contrast, the accounts that I describe here *do* find such use.

then suggest splitting up this reliability node into separate sincerity, objectivity and sensitivity nodes, because “there are several different ways in which a source of evidence can be unreliable” (56). In other words, they have us consider several alternative explanations of a statement. Similarly, one of their predecessors, Friedman (1986, 668), uses a Bayesian network-like approach in which the likelihood ratio of the testimony is evaluated by considering “chains of circumstances that might have led to a given declaration.”

Despite the similarity between this account and my account, they differ in at least one important way. On the capacity approach, the aforementioned critical questions are the focal point of eyewitness evaluation. In contrast, on my account, the alternative explanations that we consider are not fixed in number and in generality. Which explanations we consider and how specific these explanations are, depends on context (see Sect. 2). Furthermore, these explanations do not have to pertain to the capacities of the witness. For instance, someone might consider the possibility that they misheard the witness’s statements.²² TIBE therefore fits with the idea behind the capacity-approach, but is a more flexible, context-sensitive way of spelling out this underlying idea.

This more flexible approach has both benefits and drawbacks. I will illustrate the benefits of this flexibility in depth in Sects. 5 and 6. But, briefly put, one downside of the capacity approach is that it is not very informative about how evaluators ought to interpret evidence about the reliability of the witness in a context-sensitive way. A second downside is that the approach can be both overly and underly inclusive in the kinds of evidence and explanations that it has evaluators consider. By interpreting the capacity approach as an instance of TIBE, we avoid these difficulties.

The capacity approach also has its benefits. An important one is that there is strength in simplicity. It gives evaluators a clear set of questions to ask, whereas TIBE makes them do more work in terms of generating and specifying the available explanations. The explanations that we consider on the capacity approach can therefore be a good starting point for TIBE, after which we then make them more specific and ask whether there are also alternative explanations.

4.2 Empirically informed evaluation methods

The second type of evaluation method with which I compare my account is one developed within psychology. Psychologists have extensively studied eyewitness reliability, distinguishing numerous aspects of events, environments and the witnesses themselves that can (positively or negatively) influence eyewitness reliability (Ross et al., 2014; Wise et al., 2014). This has led to a wide range of proposals on how to minimize eyewitness errors, such as better ways to design interviews. Psychologists have also tackled the question how these empirical insights can be used to improve eyewitness *evaluation* by developing prescriptive, empirically-informed methods that experts can use to assess eyewitness reliability (Griesel & Yuille, 2007). It would go beyond the scope of this article to review all of them. Let me just give two

²² In that case, the report whose reliability we are assessing is our own recollection of what the witness said.

examples of influential approaches to illustrate how these approaches differ from, but also fit with my proposal.

Arguably the best-known and most-used empirically-informed method is Criteria-Based Content Analysis (CBCA).²³ The bedrock of this approach is the hypothesis that testimony about statements derived from the memory of real-life experiences differs significantly in content and quality from fabricated or fictitious accounts (Steller, & Köhnken, 1989). The method consists of a checklist that scores testimony on nineteen criteria such as the quantity of details in a testimony, how self-deprecating a witness is about the reliability their statement and references to the witnesses' own mental state. For most of these criteria, research is available that supports a correlation with (un)reliable testimony (Vrij, 2005). The CBCA method was developed primarily to be used for the evaluation of testimony by children in sexual abuse cases, though some have argued that it could be used for a wider variety of cases.²⁴

The CBCA is sometimes interpreted as if it provides an algorithm for calculating the witness's reliability, where a high score on the list automatically corresponds to a high degree of witness reliability. However, such an algorithmic interpretation is problematic (Vrij, 2008, p. 241). For instance, while meta-analyses do show that a high score on the list correlates with truthful testimony, the correlation is weak and there are many false positives (Vrij, 2005, 2008; Amado et al., 2016; Oberlader et al., 2021). Steller & Köhnken (1989, 231) themselves also argued that the occurrence of criteria on the list does not only depend on whether the source is reliable, but also on personal and situational factors which may be unrelated to the reliability of a statement, such as the witness's age. When interpreting the results of the CBCA, such factors also have to be taken into account. That is why various authors suggest that the CBCA is part of a broader diagnostic process, known as Statement Validity Assessment (SVA). SVA begins with a case file analysis, which gives the expert insight into what may have happened and the issues that are disputed. The second part is a semi-structured interview on the event in question. Third, the CBCA is used to analyze the transcript of this interview. Finally, a 'Validity Checklist' is used to further interpret these results in context. During this analysis, the expert looks at eleven issues that are can affect CBCA scores, such as the quality of the interrogation and certain personality characteristics of the witness (e.g., suggestibility) (Vrij, 2008, p. 204).

The way that the above method structures eyewitness evaluation is not inherently explanatory. Nonetheless, it can straightforwardly be interpreted as such. For instance, as Oberlader (2019, 13) summarizes the method: "SVA examines various alternative hypotheses for the development of a statement". According to her, these hypotheses are derived during the case file analysis (Oberlader, 2019, 13). Explanatory reasoning is similarly present during the later steps of this method. For example, when we interpret the results of the CBCA, we must consider alternative reasons why the statement scores particularly high or low on the checklist. For instance, "a low-quality statement might be given if the event in question was so simple and short that many criteria could just not occur" (Oberlader, 2019, 14). In such a situation, there

²³ The CBCA is admissible evidence in some courts of North America and in several West European countries such as Austria, Germany, Sweden, Switzerland, and the Netherlands (Oberlader, 2019, 15).

²⁴ Admittedly, others have greeted this claim with skepticism (Vrij, 2008, p. 222).

is a plausible truth-telling explanation for why this statement was produced which is consistent with a low score on the CBCA. So, TIBE fits well with the underlying idea of the approach. Furthermore, viewing this approach in terms of explanatory comparison again has several benefits. Some of those benefits I mentioned earlier when I discussed the capacity approach. Another benefit relates to a well-known point of criticism of SVA, namely that it lacks a clear method for drawing conclusions from the resulting analyses (Steller and Köhnken, 1989, 231). To draw such conclusions, we may employ inference to the best explanation, where we assign a degree of reliability to the witness based on the plausibility of the truth-telling explanation compared to the alternative explanations (see Sect. 2.1).

Let's consider another, more recent method, which was developed by Wise, Fisher & Safer (2009). On this method, the evaluator asks whether the interview and identification procedures in the case were fair, unbiased and sufficiently thorough. Furthermore, the evaluators ask what eyewitness factors during the crime are likely to have increased or decreased the accuracy of the eyewitness testimony. The method consists of a list of questions that help evaluator assess these aspects, such as: "Did the interviewers contaminate the eyewitness's memory of the perpetrator of the crime?", "Was there reliable, valid corroborating evidence that establishes the veracity of the eyewitness testimony?" and "was the witness intoxicated?" According to its creators, this method offers a comprehensive analytical framework for "identifying and organizing the myriad of disparate factors that affect the accuracy of eyewitness testimony" (Wise, Fisher & Safer, 2009, 472). The authors argue that this method is useful for judges or jurors during the process of evaluation, providing them with a structured method and a way to incorporate the psychological findings on eyewitness reliability into their judgments. They also suggest that the method can similarly be useful for interviewers to optimize the reliability of their interview and for attorneys to develop ways to defend or attack an eyewitness's reliability.

Much like with SVA, the explanatory approach that I propose can complement this method. First, this method also lacks a clear way method for drawing conclusions from the resulting analysis. We may use inference to the best explanation as such a method. Furthermore, we may not always want to spend the same amount of time and attention on each aspect on the extensive list. Which aspects we focus on most and how we answer the questions from the list can be based on which explanations for the testimony are most salient. Conversely, both this approach and the SVA can inform how we go about inferring to the best explanation. For instance, the method of Wise et al. (2009) lists a number of questions that we might ask to check how plausible such a hypothesis is. So, (parts of) this method can be helpful for explanatory inference, by pointing to the relevant facts. Similarly, these checklists can also point to specific explanations that we could (or should) consider in a given case.

In summary, explanatory reasoning not only fits well with existing prescriptive approaches, it also complements them. In particular, TIBE offers a flexible, context-sensitive framework for drawing conclusions from the data that we gather using, for instance, psychological checklists. Such context-sensitivity is especially important given that which evidence deserves our attention and how we ought to interpret this evidence varies with the salient explanations for a particular piece of testimony. In the next two sections I consider this point in more depth. In particular, I distinguish

two challenges that evaluators face when drawing conclusions from the evidence about an eyewitness's reliability. The first is that of interpreting the evidence, the second is that of deciding which evidence is relevant. This discussion will further illustrate the benefits of explanatory reasoning about eyewitness reliability.

5 Interpretation and striking agreement

One difficulty that evaluators face is that they have to interpret the evidence about a witness's reliability. The problem is that the same fact can support contradicting conclusions about a report's reliability, depending on how we interpret it. Consider the work of Shapin (1994, 212–238) who discusses a list of familiar and intuitively plausible maxims for the assessment of testimony. His strategy is to find a 'counter-maxim' for each. For instance, a knowledgeable witness is often good, but can also be bad because they tend to over-interpret what they observe. Similarly, confidently delivered testimony may be a sign that the witness saw the events that they reported on clearly – as opposed to a witness who offers many caveats about their observations – but it may also be indicative of overconfidence or a liar. Shapin argues that we can do this for almost any common-sense generalization about how certain types of facts relate to the reliability of a witness. So, we cannot draw inferences such as: 'all else being equal, the fact that this witness is knowledgeable always supports them being reliable', because the opposite inference may sometimes also be warranted. Rather, we have to determine how a particular fact should be interpreted in a particular context. But how do we do so?

To further illustrate this problem, let's look at an example in more depth, namely the phenomenon of *striking agreement*. Striking agreement occurs when multiple witnesses agree on a specific, unlikely detail. For instance, imagine that a robbery took place and that there were several witnesses. Suppose that investigators take statements from each of the witnesses. Now compare the following situations:

- i. The witnesses all claim that the robber wore a t-shirt and jeans.
- ii. The witnesses all claim that the robber wore a clown suit.

People wearing t-shirts and jeans are much more common than clown suits. Therefore, the agreement between the witnesses in the second case is much more striking than in the first. But what conclusion should we draw from such strikingness? One common thought is that if witnesses *independently* agree on an implausible hypothesis, then this provides stronger confirmation for this fact than if the witnesses independently agree on a more likely fact. Bovens & Hartmann (2003) offer a Bayesian analysis of this idea. They argue that the likelihood ratio for the statement of the two witnesses in the second case is higher than in the first. The idea is that when witnesses misreport on what happened, some false stories are more likely than others. For instance, imagine that if the robber wore yoga pants and a hoodie, it would not be especially surprising for a witness to mistakenly report that he wore a t-shirt and jeans. People often wear t-shirts and jeans. So, as Bovens & Hartmann (2003, 113) put it:

[T]his is not simply one out of so many false stories. It is the sort of thing that someone is likely to say if she does not know, or does not wish to convey, any details about the person missing in action, but feels compelled to say something.

In contrast, it would be highly unlikely if several witnesses all independently, falsely claimed that the witness wore a clown suit when this was not the case. According to Bovens & Hartmann, such a story is ‘too odd not to be true’.

However, we can also view striking agreement between witnesses in a different light, namely as a sign that the witnesses are *unreliable*. Even relatively reliable witnesses will make small mistakes. Therefore, we would expect some incongruity between their reports. If their testimonies cohere to a surprising degree, this can lead to a suspicion of conscious or subconscious collusion – i.e., that the witnesses are *not* independent. We might then assign a lower likelihood ratio to their testimony. Gunn et al., (2016) offer a Bayesian formalization of this argument, where the agreement is considered ‘too good to be true.’ They use the example of a police line-up. If the line-up is sufficiently big and enough witnesses unanimously identify the same person as the perpetrator, then we can be virtually certain that the line-up was biased in some way. This may be “for example, because the suspect is somehow conspicuous, [or because] the staff running the parade direct the witnesses towards him” (Gunn et al., 2016, 6).

These conflicting ways to interpret this phenomenon pose a challenge for eye-witness evaluators. Suppose that we encounter an instance of striking agreement between multiple witnesses. Should we see this as a sign of truthfulness or of collusion? Is it too good to be true or too odd not to be true? My answer is that, on TIBE, which interpretation of some multi-interpretable fact we should choose depends on which explanation of that fact is the most plausible in context.²⁵ In the case of striking agreement we face multiple possible explanations of the witnesses’ utterances. One is the truth-telling explanation, emphasized by Bovens and Hartmann. They quote Lewis (1946, 346) who remarks that: “[O]n any other hypothesis than that of truth-telling, this agreement is very unlikely”. In contrast, Gunn et al. suggest several alternative explanations (e.g., the suspect being conspicuous in the line-up). Once we have these competing explanations, we can further investigate and evaluate them. Were the witnesses indeed independent? Are there signs of collusion? Was the suspect conspicuous in the line-up?

This answer may seem too obvious to mention. Nonetheless, we may contrast it with the two alternative prescriptive accounts that I discussed earlier. These frameworks do not specify how we should go about interpreting the evidence within these frameworks. On the capacity approach, we might ask whether these witnesses are lying or misremembering. But striking agreement can be both evidence for and against lying or misremembering. The problem of interpretation is also a thorny one for empirically informed methods, which presume that we can determine beforehand which factors will count for or against reliability. However, even if a fact *generally* correlates with, for instance, lower reliability, it may nonetheless be reasonable to take it to signal increased reliability in a specific context. So, it seems that these meth-

²⁵ Lipton (1998, p. 30) also makes this point with respect to the aforementioned work by Shapin (1994).

ods need some further mechanism to account for evidential interpretation. TIBE can be of service in that regard. The fact that explanatory reasoning offers such a seemingly obvious approach to interpretation only speaks to its naturalness.

The above discussion also shines further light on the relationship between TIBE and Bayesianism. The work of both Bovens & Hartmann and Gunn et al. is part of the extensive Bayesian modeling literature on witness reliability.²⁶ Thagard (2005), who argues that TIBE and Bayesianism are incompatible, further suggests that Bayesian modeling is generally unhelpful for studying eyewitness reliability. In particular, he notes that any Bayesian model will require many conditional probabilities and that it is unclear what numbers we should plug in for these probabilities (especially because it is unclear how these probabilities are to be interpreted).²⁷ But regardless of whether we grant Thagard's points about modeling *particular instances* of eyewitness testimony, we do not have to accept his blanket rejection of Bayesian modeling. Because TIBE is compatible with Bayesianism – and the latter provides a helpful way of understanding the former – such modeling is a helpful tool for elucidating the epistemic consequences of certain assumptions within our explanatory inferences. For instance, the work on striking agreement is an excellent example of how the underlying assumptions of arguments about the reliability of testimony based on competing explanations can be made explicit and precise.²⁸

6 Relevance

A second problem that evaluators have to deal with is deciding *which* evidence they should consider when evaluating an eyewitness's reliability. Bayesians sometimes assume that we should take into account our 'total evidence', i.e., everything relevant that we know (Goldman, 1999, 145). But this is often not practically feasible. For any particular eyewitness there are countless facts that could, conceivably, influence how reliable we ought to consider them. Among these are the hundreds of factors which psychological research suggests may influence the reliability of eyewitnesses, such as their stress level during the events, whether they were asked leading questions and how long ago the reported events took place (Wise et al., 2014). Evaluators cannot easily determine whether, and to what extent, each of these countless facts is present in a particular instance. Nor can they readily draw a conclusion about the witness's reliability if the facts that they know to be present are too numerous. Finally, they have to decide which further evidence to gather. As Lipton (2003, 116) puts it:

²⁶ See Merdes, von Sydow & Hahn, (2020) for a review of some of this literature.

²⁷ In contrast, Rassin (2014) notes that Bayesian quantification could help make explanation-based eyewitness evaluation more precise. I am skeptical of this suggestion because of the well-known difficulties that many judges and jurors have with interpreting numerically-expressed evidence, especially when this evidence lacks a clear statistical basis.

²⁸ See Bovens & Hartmann (2003), Chaps. 3 and 5, for how various assumptions about the reliability of witnesses and the probability of what they report lead to differences in the impact of their testimony on our beliefs.

In practice, investigators must think about which bits of what they know really bear on their question, and they need also decide which further observations would be particularly relevant.

The capacity approach and empirically informed methods offer valuable guidance on which factors to look at. However, I believe that thinking in terms of possible explanations has additional value here. For instance, the capacity view directs us to look at the evidence about the eyewitness's capacities (e.g., perception, memory, veracity). But there will typically be numerous facts that are potentially relevant for assessing whether the witness exercised these capacities correctly. On the other hand, some relevant facts may not be directly related to the capacities of the witness. The same goes for empirically-informed methods, where the facts that evaluators look at are determined beforehand. But this also means that if evaluators use, say, the CBCA, then this list will contain various facts that are not especially relevant in context and exclude others that are.

Explanatory thinking can help focus our attention on the most important facts. On TIBE, relevance of facts is dictated by the explanations that we consider. As Lipton (2003, 116) points out, we may discover relevant evidence for a hypothesis by asking what it would explain. I would add to this that we may also consider what the explanation would *not* explain. In other words, we select the relevant facts by considering what we would expect to see if this explanation were true and what we would not expect to see. What should be especially interesting during witness evaluation are *discriminating facts* – i.e., facts that confirm one of the available scenarios but falsify the other. The more specific and limited our set of explanations, the easier it will be to answer these questions and the more directed our search for evidence will be.

Let's consider an example that shows both the breadth of facts that can potentially be pertinent, and how explanatory reasoning may guide us in seeing their relevance. The following passage is taken from a lawyer's plea in a Dutch case, concerning a defendant who allegedly broke into his girlfriend's house and destroyed some of her belongings (van Oorschot, 2018, 209):

[W]hen we look at the [report] of the interrogation with my client, I see, typed down, in the middle of one of my client's questions, the phrase, "theft, unequal.", which arguably stands for "theft, unqualified". But my client would never express himself this way, nor would other defendants, presumably. So who is speaking here? The police officer or the defendant?

But why should it matter that the police officer phrased the report this way? The language of a report would normally not be especially relevant to the question of its reliability. Nonetheless, such a fact could *become* germane if we were (implicitly) considering the possibility that the officer did not accurately report what the defendant said. Indeed, that is how van Oorschot (2018, 209–210), from whom I borrow this example, interprets this passage:

This lawyer (...) challenges the neutrality of the written transcription (...). He suggests that the police officials have been so set on shaping and rewriting his

client's words that it has now become unclear who precisely is speaking (...) how can we be sure that the rest of the client's statements (...) are truly his - and not added by zealous police officials?

Van Oorschot (2018, 210) also remarks that "the lawyer does not merely offer an interpretation of the different stories present in the file; he also tells us a story about the file". In other words, the lawyer proposes an alternative explanation of this passage, which competes with the hypothesis that 'this report accurately reflects what the defendant said'. They draw our attention to a fact that we might not otherwise consider relevant, and shows that it *becomes* relevant, once seen in the light of this alternative explanation. In fact, this particular datum might have come to the lawyer's attention precisely *because* they were considering the alternative explanation that the police was putting words in his client's mouth.

This example also shows how we may generate novel explanations for a witness's testimony. As in the case of striking agreement, we encounter a surprising fact – in this instance an unusual phrase in the case file – and ask what might explain it. In other words, we employ abductive reasoning to generate novel explanations. Lipton (2003, 116) points out that this is one of the strengths of inference to the best explanation, compared to Bayesian epistemology. As Lipton and others have noted, Bayesianism has nothing to say about the act of creating hypotheses. Furthermore, the act of generating hypotheses is not an explicit part of either the capacity approach or empirically-informed methods. Nonetheless, hypothesis generation is an important aspect of both evidential reasoning in general and eyewitness evaluation specifically. The discovery of an alternative explanation for some fact will often lower our confidence in our initial explanation and can be a source of reasonable doubt (Jellema, 2022).

7 Conclusion

Testimonial inference to the best explanation begins with the question: 'what could explain this particular piece of testimony given by this source in this context?' Once we have one or more candidate explanations, it is helpful to make them more specific. We then compare these narratives of the witness's explanations and arrive at a conclusion about whether – and to what degree – the source of the testimony is reliable.

This work is part of a broader explanationist trend in legal evidence scholarship. Inference to the best explanation-based approaches are increasingly popular in the philosophy of legal evidence. Nance (2016, 84) observes the following about this trend:

One main motivating concern of those who press the explanatory approach is that [probabilistic accounts] focus on the end product of deliberation, rather than the process of arriving there, giving no direction to jurors as to how to go about assessing the evidence in the case.

My argument for *testimonial* inference to the best explanation rests on a similar idea. I embrace the Bayesian account of what we're trying to achieve when we evaluate witness reliability – namely to evaluate the likelihood ratio of the testimony in

the light of the available evidence. But this Bayesian story does not tell evaluators much about how to draw conclusions from the available evidence about the report's reliability. Yet this is precisely what we want from an account of rational eyewitness evaluation. I have outlined how explanation-based reasoning (i) tracks the Bayesian likelihood ratio, and (ii) helps structure, interpret and select the available evidence. These are the two pillars of TIBE's rationality.

I further explored TIBE's prescriptive ambitions by comparing it to existing approaches to eyewitness evaluation. I focused on two influential types here: the capacity approach and empirically-informed methods. As I argued, these approaches and TIBE may complement one another. On the one hand, these approaches may inform us of the kinds of considerations that will also be relevant in explanation-based inference. On the other hand, though these approaches are not inherently explanatory, once we incorporate an explanatory component, we end up with a more context-sensitive, flexible framework on which it is clearer how evaluators ought to come to their ultimate decision. I further explored this idea – that TIBE can complement other frameworks by telling us how to make sense of the evidence – by focusing on two specific of eyewitness evaluation where explanatory reasoning is especially helpful, namely the interpretation and selection of the evidence about the report's reliability.

Despite the increasing popularity of explanation-based accounts in legal evidence scholarship, explanationists have not given much attention to the question how we should assess whether our evidence is reliable. In this article, I argued that explanation-based thinking can be rational and helpful in assessing the reliability of one important type of evidence, namely eyewitness testimony. Whether similar accounts can be given for other kinds of evidence evaluation will have to be seen, though I have a hunch that explanationism has much to contribute on this front.

Acknowledgements I would like to thank Henry Prakken, Anne Ruth Mackor, Pepa Mellema, Marieke Dubelaar, Stefan Sleeuw, the audience at the Criminal Law Theory Workshop (especially commentator Michael Pardo) as well as the three anonymous reviewers for their helpful comments on past versions of this article.

Funding This work is part of the research programme with project number 160.280.142, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

Declarations

competing interests The author has no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen, R. J. (2020). Legal Probabilism - A Qualified Rejection: A Response to Hedden and Colyvan. *Journal of Political Philosophy*, 28(1): 117–128.
- Allen, R. J., & Pardo, M. S. (2019). Relative plausibility and its critics. *The International Journal of Evidence & Proof*, 23(1–2), 5–59
- Amado, B. G., Arce, R., Farina, F., & Vilarino, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16(2), 201–210
- Amaya, A. (2015). *The tapestry of reason: An inquiry into the nature of coherence and its role in legal argument*. Bloomsbury Publishing
- Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford University Press on Demand
- Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and human behavior*, 7(1), 19
- Cabrera, F. (2017). Can there be a Bayesian explanationism? On the prospects of a productive partnership. *Synthese*, 194(4), 1245–1272
- Cardozo, B. N. (2009). *Reevaluating Lineups: Why Witnesses Make Mistakes and How to Reduce the Chance of a Misidentification: An Innocence Project Report*. New York, NY: Innocence Project, Inc
- Dellsén, F. (2018). The heuristic conception of inference to the best explanation. *Philosophical Studies*, 175(7), 1745–1766
- Dubelaar, M. J. (2014). *Betrouwbaar getuigenbewijs: Totstandkoming en waardering van strafrechtelijke getuigenverklaringen in perspectief*. Kluwer. [Reliable witness evidence: The origin and evaluation of criminal witness statements in perspective]. Kluwer
- Fallis, D. (2002). Goldman on probabilistic inference. *Philosophical Studies*, 109(3), 223–240
- Fenton, N., Neil, M., & Berger, D. (2016). Bayes and the law. *Annual review of statistics and its application*, 3, 51–77
- Fricker, E. (2017). Inference to the Best Explanation and the Receipt of Testimony: Testimonial Reductionism Vindicated. In T. Poston, & K. McCain (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation* (pp. 262–294). Oxford University Press
- Friedman, R. D. (1986). Route analysis of credibility and hearsay. *Yale LJ*, 96, 667
- Goldman, A. I. (1999). *Knowledge in a social world* (281 vol.). Clarendon Press Oxford
- Griesel, D., & Yuille, J. C. (2007). *Credibility assessment in eyewitness memory*.
- Gunn, L. J., Chapeau-Blondeau, F., McDonnell, M. D., Davis, B. R., Allison, A., & Abbott, D. (2016). Too good to be true: When overwhelming evidence fails to convince. *Proc. R. Soc. A*, 472
- Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge university press
- Harman, G. (1965). The inference to the best explanation. *Journal of Philosophy*, 74, 88–95
- Hawthorne, J. (1994). On the nature of Bayesian convergence. In: Proceedings of the Biennial Meeting of the Philosophy of Science Association, vol. 1994, no. 1, 241–249
- Hedden, B., & Colyvan, M. (2019). Legal probabilism: A qualified defence. *Journal of Political Philosophy*, 27(4), 448–468
- Henderson, L. (2014). Bayesianism and inference to the best explanation. *The British Journal for the Philosophy of Science*, 65(4), 687–715
- Jellema, H. (2021). The reasonable doubt standard as inference to the best explanation. *Synthese*, 199(1), 949–973
- Jellema, H. (2022). Reasonable Doubt from Unconceived Alternatives. *Erkenntnis*
- Josephson, J. R. (2000). On the proof dynamics of inference to the best explanation. *Cardozo L Rev*, 22, 1621–1643
- Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, 4(1), 46–63
- Laudan, L. (2007). Strange bedfellows: Inference to the best explanation and the criminal standard of proof. *The International Journal of Evidence & Proof*, 11(4), 292–306
- Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. Open Court
- Lipton, P. (1998). The epistemology of testimony. *Studies in History and Philosophy of Science Part A*, 29(1), 1–31
- Lipton, P. (2003). *Inference to the best explanation*. Routledge
- Lipton, P. (2007). Alien abduction: Inference to the best explanation and the management of testimony. *Episteme*, 4(3), 238–251

- Malmgren, A. S. (2006). Is there a priori knowledge by testimony? *The Philosophical Review*, 115(2), 199–241
- Mackor, A. R., Jellema, H., & van Koppen, P. L. (2021). Explanation-Based Approaches to Reasoning about Evidence and Proof in Criminal Trials. *Law and Mind: A Survey of Law and the Cognitive Sciences* (pp. 431–470). Cambridge University Press
- McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *The British Journal for the Philosophy of Science*, 54(4), 553–567
- Merdes, C., Von Sydow, M., & Hahn, U. (2020). Formal models of source reliability. *Synthese*
- Nance, D. A. (2016). *The burdens of proof*. Cambridge University Press
- Oberlander, V. A. (2019). Meta-Analyses on the Validity of Verbal Tools for Credibility Assessment
- Oberlander, V. A., Laura, Q., Banse, R., Volbert, R., Schmidt, A. F., & Schönbrodt, F. D. (2021). Validity of Content-Based Techniques for Credibility Assessment-How Telling is an Extended Meta-Analysis Taking Research Bias into Account?. *Applied Cognitive Psychology*
- O'Brien, B. (2009). Prime suspect: An examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychology Public Policy and Law*, 15(4), 315
- Pardo, M. S., & Allen, R. J. (2008). Juridical Proof and the Best Explanation. *Law and Philosophy*, 27(3), 223–268
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, 62(2), 189
- Pennington, N., & Hastie, R. (1993). *The story model for juror decision making*. Cambridge University Press
- Puddifoot, K. (2020). Re-evaluating the credibility of eyewitness testimony: the misinformation effect and the overcritical juror. *Episteme*, 17(2), 255–279
- Rassin, E. (2001). Het bepalen van geloofwaardigheid: De methode van de alternatieve scenario's [Evaluating reliability: the method of alternative scenarios]. *De Psycholoog*, 36, 348–355
- Rassin, E. (2014). Het beoordelen van de validiteit van (getuigen)verklaringen met de methode van de alternatieve scenario's; een update [Judging the validity of (eyewitness) statements with the method of alternative scenarios; an update]. *Expertise en Recht*, 4, 119–123
- Redmayne, M. (2003). Objective probability and the assessment of evidence. *Law probability and risk*, 2(4), 275–294
- Ross, S. J., Tredoux, C. G., & Malpass, R. S. (2014). Evaluating eyewitness testimony of adults. In I. B. Weiner, & R. K. Otto (Eds.), *The handbook of forensic psychology* (4th ed., pp. 513–559). Hoboken, NJ: John Wiley & Sons
- Spottswood, M. (2019). On the Limits of a Unitary Model of the Proof Process. *The International Journal of Evidence & Proof*, 23(1–2), 75–81
- Stevens, L. (2016). Case note: Hoge Raad (Het toetsen van de betrouwbaarheid van de betrouwbaarheid van getuigenverklaringen en de bewijsconstructie: wat vermag de rechter?). *Ars aequi*, 2016 (4)
- Talbot, W. (2008). Bayesian Epistemology'. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Ed.). <http://plato.stanford.edu/archives/fall2008/entries/epistemology-bayesian/>
- Tang, W. H. (2016). Reliability theories of justified credence. *Mind*, 125(497), 63–94
- Thagard, P. (2005). Testimony, Credibility, and Explanatory Coherence. *Erkenntnis*, 63(3), 295–316
- Van Oorschot, I. (2018). *Ways of Case-Making*. Dissertation, Erasmus University Rotterdam
- Vrij, A. (2005). Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology Public Policy and Law*, 11(1), 3
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons
- Walton, D. (2007). *Witness Testimony Evidence: Argumentation and the Law*. Cambridge University Press
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge University Press
- Williamson, T. (2002). *Knowledge and its Limits*. Oxford University Press on Demand
- Wise, R. A., Sartori, G., Magnussen, S., & Safer, M. A. (2014). An examination of the causes and solutions to eyewitness error. *Frontiers in psychiatry*, 5, 102
- Wittlin, M. (2019). Common problems of plausibility and probabilism. *The International Journal of Evidence & Proof*, 23(1–2), 184–190