Forensic Science International: Digital Investigation 47 (2023) 301657

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

Adding transparency to uncertainty: An argument-based method for evaluative opinions

Nina Sunde [a], Virginia N.L. Franqueira [b,*]

[a] The Norwegian Police University College, Oslo, Norway
[b] Institute of Cyber Security for Society (iCSS), School of Computing, University of Kent, Canterbury, UK

ARTICLE INFO

Keywords:
Criminal investigations
Evaluative opinion
Digital evidence
Argumentation theory
Practitioners

ABSTRACT

Over the past 15 years, digital evidence has been identified as a leading cause, or contributing factor, in wrongful convictions in England and Wales. To prevent legal decision-makers from being misled about the relevance and credibility of digital evidence and to ensure a fair administration of justice, adopting a balanced, systematic and transparent approach to evaluating digital evidence and disseminating results is crucial. This paper draws on general concepts from argumentation theory, combined with key principles and concepts from probabilistic and narrative/scenario approaches to develop arguments and analyse evidence. We present the "Argument-Based Method for Evaluative Opinions", which is a novel method for producing argument-based evaluative opinions in the context of criminal investigation. The method may be used stand-alone or in combination with other qualitative or quantitative/statistical methods to produce evaluative opinions, highlighting the logical relationships between the components making up the argument supporting a hypothesis. To facilitate a structured assessment of the credibility and relevance of the individual argument components, we introduce an Argument Evaluation Scale and, ultimately, an Argument Matrix for a holistic determination of the probative value of the evidence.

1. Introduction

Digital evidence is a crucial component of modern criminal investigations. Due to the complex nature of such evidence, expertise is often required not only for securing and analysing the evidence but also for evaluating and communicating its probative value to legal decision-makers. In contrast to several other forensic science domains, there is no fixed or standardised knowledge base for digital evidence, partially due to digital technology's rapidly changing and complex nature. Hence, providing an evaluative opinion about digital evidence often requires domain expertise combined with extensive and rigorous experimentation and analysis. Such a process is characterised by many considerations, assessments and decisions made by the digital forensic practitioners – and thus, a high degree of subjectivity is involved.

The Case Assessment and Interpretation (CAI) framework (Cook et al., 1998a,b), developed in the 1990s by the Forensic Science Service in England and Wales, aimed to improve the form and reliability of scientific opinions. The principles were incorporated in authoritative guidelines such by ENFSI (European Network of Forensic Science Institutes, 2015) and the UK Forensic Science Regulator (Forensic Science Regulator, 2021). The ENFSI guidelines promote that the value of scientific findings should conform to the requirements of balance, logic, robustness and transparency and that evaluation and reporting should be performed in light of competing hypotheses (i.e., prosecution and defence); it recommends a probabilistic approach for measuring uncertainty. The process aims to prevent bias and subjectivity, although no method involving human reasoning can guarantee an error-free or unbiased opinion.

Research on miscarriages of justice in England and Wales during the last 15 years suggests a shift from more traditional forensic evidence (including medical, biological, chemical, and feature comparison) driving the majority of identified wrongful convictions before 2007 to digital evidence as the primary cause or contributing factor after 2007 (Helm, 2022). Smit et al. (2018) studied case transcripts of rulings between 2010 and 2016 regarded as unsafe by the Court of Appeal of England and Wales, and found that misleading evidence was the reason for the unsafe ruling in 235 cases. Their analysis showed that relevance, probative value, and validity of evidence were often misunderstood and

* Corresponding author.
E-mail address: v.franqueira@kent.ac.uk (V.N.L. Franqueira).

https://doi.org/10.1016/j.fsidi.2023.301657
Received 25 July 2023; Received in revised form 25 September 2023; Accepted 6 October 2023

miscommunicated within a criminal trial when expressing beliefs in (competing) hypotheses. Transparency of the reasoning associated with an opinion is crucial to enable an effective internal audit through, e.g., peer review, external scrutiny and cross-examination in court.

An empirical study of digital forensic opinions (Sunde, 2021) showed that practitioners tended to use categorical expressions of credibility to support conclusions in their reports. The reports were characterised by *one-sidedness* since the conclusions were all related to a single explanation or hypothesis and were not presented in a balanced manner. Although all participants had analysed the same evidence file, their conclusions - which were quite convincing - varied greatly and pointed in various directions. Some concluded that criminal activity did happen and that the suspect was responsible for it, while others suggested that the suspect did not perform any criminal activities. If the participants had assessed and presented the evidence in light of opposing hypotheses (as suggested by ENFSI), some would have seen that the evidence was strong and relevant for both hypotheses (supporting the suspect's guilt and innocence) and, therefore, that the evidence would help discriminate between them. Argumentation theory provides a framework and terminology for argument-based analysis of evidence under a hypothesis. It has been introduced in the context of digital forensics to increase rigour, standardisation and scrutiny in practices and reporting (Franqueira and Horsman, 2020). Therefore, it facilitates increased transparency of the reasoning behind practitioners' opinions and provides insights into how the components are linked. Argumentation theory helps display uncertainties not only to the credibility of evidence but also to the logical relationships between the components involved due to its attention towards the inferential leaps from the evidence to a hypothesis.

The overall objectives of the proposed method were to (1) ensure fair administration of justice and safeguard the rule of law by promoting a balanced approach, (2) enable scrutiny by increasing transparency in the reasoning of expert opinions, and (3) offer an accessible and flexible way to provide evaluative opinions, which can be used as a stand-alone method or in combination with other methods.

1.1. Contribution and organisation of the paper

This paper draws on argumentation theory combined with the ENFSI requirements of balance, logic, robustness and transparency and presents the *Argument-Based Method for Evaluative Opinions* as a method for structured evaluation of the probative value of evidence. It expands the practitioners' toolbox and offers an alternative, non-numerical, method to elaborate expert opinions.

The proposed method is novel in the sense that it empowers practitioners to apply an argumentation-based framework to equally consider opposing hypotheses when assessing evidence. It allows systematic assessment of both the credibility and relevance of the evidence, as well as the implications of missing evidence, enabling them to elaborate evaluative opinions in a clear and transparent qualitative manner.

The method emphasises the logic, which acts as the *glue* between the evidence and the hypothesis, therefore adding *logical reconstruction* to the existing strategies within digital forensics such as the temporal, relational and functional reconstruction strategies (Casey, 2011; King, 2006). The method is probabilistic and commits the evaluator to assess the evidence and other components of the argument in light of at least two opposing hypotheses. This prevents misinterpretations of probative value since a high probative value of an argument does not necessarily mean that the opposing argument would have a low probative value, a phenomenon referred to as the false dichotomy fallacy (Cederblom and Paulsen, 2012, Page 145). The *Argument-Based Method for Evaluative Opinions* may be used stand-alone to assess the relative probative value between the hypotheses or in conjunction with other qualitative or quantitative/statistical methods for determining a more accurate value, e.g., a likelihood ratio. It is flexible and may be applied for targeted analysis of one or a few pieces of evidence from a single forensic domain, such as digital forensics, as demonstrated in this paper. However, it may also be feasible for a holistic analysis of different types of evidence from different forensic domains.

The paper is organised as follows. First, the related applications of argumentation theory in digital forensics are discussed, followed by an elaboration on the key theoretical concepts constituting an argument and the concepts used for argument evaluation (Section 2). Then, the *Argument-Based Method for Evaluative Opinions* is outlined in Section 3, and elaborated with two examples for the application of the method and its associated *Argument Evaluation Scale* and *Argument Matrix* in Section 4. Finally, in Section 5, the utility and limitations of the method are discussed before conclusions are drawn in Section 6.

2. Using arguments for reasoning about evidence

There are three normative approaches to reasoning with legal evidence: probabilistic, narrative or scenario and argumentation based approaches (Verheij, 2012).

In probabilistic approaches to evidence evaluation, the uncertainty is handled within probabilistic terms and methods, and the Bayes' theorem is a fundamental formula. It applies an atomic perspective to evidence evaluation and allows for precision through a mathematical expression of evidential value in likelihood ratio. This approach includes a prior and an updated probability after the evidence is considered (Kolflaath, 2023). The main critique is that a knowledge base from which calculations may be obtained rarely exists in digital forensics, and the atomic approach does not consider dependencies between items of evidence (Kolflaath, 2023). Another limitation pointed out is that it does not consider the completeness of evidence (Tecuci et al., 2016).

Narrative (i.e., storyline) approaches to evidence evaluation allow for exploring whether there is an internal coherence and correlation between the events (e.g., a timeline) and whether they are backed up (anchored) with common sense generalisations and evidence generalisations (Pennington and Hastie, 1993; Wagenaar et al., 1993). It applies a holistic perspective to evidence evaluation, focusing on totality and coherence (Kolflaath, 2023). Evaluation criteria are coverage, uniqueness, coherence and completeness (Mackor and Van Koppen, 2021).

Scenario approaches offer further explication of stories, including explanations and causal relationships. Here, the evidence is evaluated in light of various scenarios and investigates how well the scenarios explain the evidence and related circumstances (Kolflaath, 2023), often referred to as *inference to the best explanation*. The process has two stages. The first involves generating potential explanations based on available evidence, and here, the general background knowledge plays a central part. The second stage involves assessing which scenario is the best based on evidence evaluation in light of each one. The criteria for testing the quality of causal explanations are uniqueness, consistency and plausibility (Mackor and Van Koppen, 2021). Robustness is yet another essential criterion for assessing the thoroughness of the search for evidence (and counter-evidence) and search for explanations (Mackor and Van Koppen, 2021). Falsification is the primary methodology for putting scenarios to the test. A limitation is that the incompleteness and uncertainties related to criminal evidence do not necessarily allow for scenarios to be straightforwardly falsified.

Argument approaches in reasoning with evidence are systematic approaches to assess the credibility and relevance of evidence by application of argumentation theoretical concepts. According to Bex (2021, Page 183), "Argumentation is the construction of arguments by performing consecutive reasoning steps, starting with an item of evidence and reasoning towards some conclusion using general rules of inference or generalizations where not just arguments for a conclusion but also counterarguments against a conclusion have to be considered". Argument approaches can span from simple evidential arguments from evidential data to premises and further to a conclusion – to complex arguments derived from multiple pieces of evidence chained together by multiple inferences, with intermediate and ultimate conclusions. Descriptive argumentation is assessed by considering the credibility of the grounds/premises and the relevance of the inferential leaps/warrants, which, and ultimately, the strength or probative force in light of the hypothesis or claim (Kolflaath, 2023; Bex, 2021). The approaches represent different styles of evidential reasoning and formal development, and all have their strengths and limitations. Researchers have explored various ways of combining probabilistic, scenario and argument based approaches for evidential reasoning (see, e.g., Verheij, 2012; Verheij et al., 2016). The main critiques of the argument-based approach have been the need for a systematic account of degrees of uncertainty in argumentation (Bex, 2021) and that argumentative reasoning does not explicitly allow for the construction and comparison of alternative scenarios to maintain an overview of a mass of evidence (Prakken et al., 2020).

The academic discourse around evidence evaluation in the broader forensic science domain, including the digital forensics discipline, mainly draws on probabilistic approaches (see, e.g., Kwan et al., 2011; Overill et al., 2013; European Network of Forensic Science Institutes, 2015; Casey, 2020; Casey et al., 2020; Tart, 2020). Although narrative and scenarios are well-known strategies for reconstruction during the investigative stage, such approaches to evidence evaluation have not influenced the digital forensics domain notably.

While argumentation approaches have played a significant role in reasoning about evidence in court contexts (see, e.g., Bex, 2021; Walton, 2002; Anderson et al., 2005; Kolflaath, 2023), there are few examples of such approaches to evidence evaluation in the digital forensics domain, e.g., applied to the validation of evidence (Boddington, 2012, 2016). Franqueira and Horsman (2020) proposed a broader use of Toulmin's argumentation structure (Toulmin, 1958, 2003) with recursiveness (Newman and Marshall, 1991) – therefore creating "threads of arguments" – applied to digital forensics. The dialectic structure allows exposure and review of key evidence and assumptions useful to support conclusions, logical reconstruction, and communication between stakeholders involved.

This paper draws on concepts from general argumentation theory mainly influenced by the Baconian probability view (Cohen, 1977, 1989; Schum, 2001, 2002) and the application in intelligence and investigative contexts by, e.g., Tecuci et al. (2016). This is combined with the key principle from the probabilistic approach of evaluating the evidence under opposing hypotheses (European Network of Forensic Science Institutes, 2015). Such a flexible approach allows for evaluating evidence under narrow or broad hypotheses and, similar to the narrative/scenario approach, may involve sequences and causal relationships. In essence, the proposed *Argument-Based Method for Evaluative Opinions* integrates foundational principles from argumentation theory with theoretical concepts from both probabilistic and narrative/scenario approaches.

2.1. Argument components

Fig. 1 illustrates the generic layout of an argument and, for clarity, the components are explained next.

A **hypothesis** is what is under evaluation, i.e., what one wants to establish the probability of being true or false. The hypothesis can be broad, involving relations to multiple and different types of grounds obtained by expertise from several domains, e.g. witness statement, DNA, fingerprint, digital evidence, or narrow, involving relation to one or a few grounds originating from a single domain.

An **evidence** is a piece of data or information obtained through an investigation.

A **ground** is an evidence-based assumption (often referred to as a fact) that links one or more pieces of evidence to the hypothesis. The value of a ground is assessed in context with the associated evidence and is understood to have two facets that constitute its credibility – the source and the information – which is further elaborated in Section 2.2.

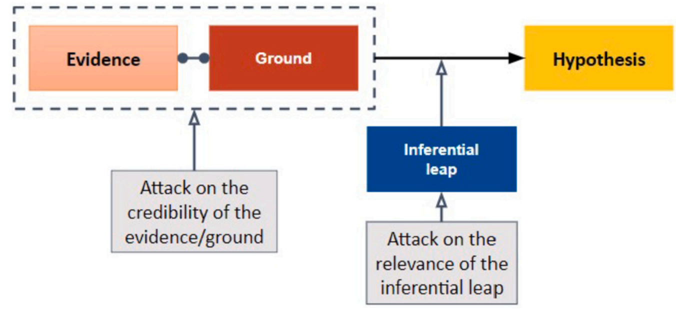

Fig. 1. Generic layout of an argument; meaning of different arrows explained in the text.

An **inferential leap** is a bridge statement or glue that connects the evidence and ground to the hypothesis. The inferential leap can be explanations of different types, e.g., a cause/effect relationship, an empirical generalisation, and any form of "common sense" statement universally regarded as true. The inferential leap is assessed according to its relevance: its ability to make the hypothesis more or less likely. An **attack** challenges any part of the argument, encompassing the assessment of both the credibility of the grounds and evidence presented, as well as the relevance of the inferential leaps made. Furthermore, apart from its role in critically appraising the existing evidence, an attack plays a significant function in evaluating the robustness of the foundational evidence and highlighting potential gaps or inadequacies. Moreover, it serves to identify any deficiencies in the reasoning process leading from the available evidence to the formulated hypothesis.

2.2. Evaluating the probative value of arguments

The assessment of an argument seeks to establish its probative value, as mentioned and is made up of two independent criteria (Kolflaath, 2023; Reitan, 2016; Tecuci et al., 2016): credibility and relevance.

- **Credibility** is the extent to which the ground and the associated evidence may be believed to be true, and is sometimes referred to as their "believability" (Tecuci et al., 2016). The credibility is assessed without considering the hypothesis in question, and by asking: *What is the probability that the ground is true?*. We underline that this is an expression of subjective and empirically founded non-numerical probability, i.e., an expression of the evaluator's belief based on the empirical findings, as opposed to aleatory or statistically determined numerical probability. The credibility of the ground would encompass two distinct components – the *source* and the *information* of the associated evidence. Sources may be humans or machines (devices or systems). The credibility of the source would, in a digital forensics context, relate to whether the source itself or handling of the source during the investigation process may have introduced evidence dynamics, which is described as "any influence that changes, relocates, obscures, or obliterates evidence, regardless of intent between the time evidence is transferred and the time the case is resolved" (Casey, 2011, Page 27). The information concerns interpretation at the semantic level and relates to factors such as authenticity, accuracy, clarity, completeness, and resolution of the information; for example, the accuracy of location data based on a cell tower position, the completeness of a chat log, or the resolution of a CCTV recording.

  Table 1 suggests criteria for assessing the credibility of the grounds and associated evidence. A conservative approach should be applied, meaning that if the ground meets the criteria of high credibility (value C4) on *source* and good credibility (value C3) on *information*, the lowest value (value C3) should be assigned.
- **Relevance** refers to the inferential leap between evidence/ground and hypothesis, and not the ground itself. The relevance of an item

Table 1
Argument Evaluation Scale: credibility of evidence & ground.

Credibility of evidence & ground	Source	Information
(C1) Very weak	The evidence is derived from a source which is regarded as highly questionable or lacks credibility. There is reason to believe that evidence dynamics have led to substantial changes affecting the meaning of the data.	The meaning is of high elasticity and very inconclusive due to vagueness, inaccuracy, low resolution or incompleteness. The ground is not supported by any evidence.
(C2) Weak	The evidence is derived from a source of unknown credibility. It is likely that evidence dynamics have led to substantial changes affecting the meaning of the data.	The meaning is elastic to some degree, and may be interpreted in more than one way. The ground has some evidential support.
(C3) Good	The evidence is derived from a credible source. Although there are some limitations and uncertainties, there is little reason to believe that the evidence dynamics have led to substantial changes affecting the meaning of the data.	The meaning is quite clear and unambiguous, with little elasticity. The ground has good evidential support.
(C4) High	The evidence is derived from a highly credible source. It is unlikely that the evidence dynamics have led to substantial changes affecting the meaning of the data.	The meaning is very clear and unambiguous, with very little elasticity. The ground has strong evidential support.
(C5) Very high	The evidence is derived from a source of unquestioned credibility. It is highly unlikely that evidence dynamics have led to substantial changes affecting the meaning of the data.	The information is of high precision, level of detail, and completeness. The meaning is very clear and unambiguous. The ground has very strong evidential support.

Table 2
Argument Evaluation Scale: relevance of inferential leaps.

Relevance of inferential leap	Inferential leap
(R1) Very weak	The inferential leap makes the hypothesis barely likely. It is theoretically possible, however, highly questionable or highly implausible. It lacks substantial or credible support.
(R2) Weak	The inferential leap makes the hypothesis likely, however, questionable or implausible. The inferential leap has some support, but it is weak or of questionable credibility.
(R3) Good	The inferential leap makes the hypothesis more likely than unlikely. It is plausible and justifiable. It is supported by credible knowledge from at least one source.
(R4) High	The inferential leap makes the hypothesis very likely. It is supported by credible and highly consistent knowledge from multiple sources.
(R5) Very high	The inferential leap makes the hypothesis certain or almost certain. It is supported by universally accepted knowledge (e.g., nature laws) or credible and fully consistent knowledge from multiple sources.

Table 3
Argument Matrix: assessment of the probative value of an argument.

		Credibility				
		(C1) Very weak	(C2) Weak	(C3) Good	(C4) High	(C5) Very high
Relevance	**(R1) Very weak**	Very weak	Very weak	Very weak	Very weak	Very weak
	(R2) Weak	Very weak	Weak	Weak	Weak	Weak
	(R3) Good	Very weak	Weak	Good	Good	Good
	(R4) High	Very weak	Weak	Good	High	High
	(R5) Very high	Very weak	Weak	Good	High	Very high

of evidence indicates how strongly does it support the hypothesis in question, and is assessed by asking: *Assuming that the evidence is true, what is the probability that the hypothesis is true?* (Tecuci et al., 2016). The criteria for assessing relevance are presented in Table 2.

When assessing relevance, the grounds and associated evidence may be considered individually. However, if the relevance of the ground under consideration rests on the relationship to another ground, their relevance should be considered together.

A holistic evaluation of credibility and relevance is necessary to determine the probative value of the argument (Anderson et al., 2005). To facilitate a structured assessment, we introduced argument evaluation scales (see Tables 1 and 2). We draw on a matrix introduced by Reitan (2016, Page 128), and present the *Argument Matrix* for a holistic determination of the probative value of the argument based on the assigned relevance and credibility, as shown in Table 3. Reitan's original matrix is expanded to enable a more detailed assessment and communication of an argument's probative value. As shown, the probative value can only be high if both the credibility and relevance are high.

3. The proposed method

Fig. 2 illustrates the steps of the proposed method, which are further elaborated in the remaining of the section.

The *Argument-Based Method for Evaluative Opinions* complies with the ENFSI principles for evaluative opinions (Section 1) mentioned above. The suggested method involves the assessment of at least two opposing hypotheses (propositions). It provides a detailed framework for articulating the reasoning concerning the credibility and relevance of the components of an argument. It provides a supplement to other methods due to the focus on the evidence itself, the inferential leaps necessary or sufficient for the hypothesis to be true, and the ability to consider any missing evidence.

The starting point of the proposed method is to define the investigative question from which the opposing hypotheses are derived. In the context of a criminal investigation, regardless of whether performed on pre-trial or trial stage – the hypotheses should represent both the prosecution and defence perspectives. If a suspect or defendant has not provided a hypothesis, one involving innocence should be constructed on the defendant's behalf to operationalise the presumption of inno-

Table 4
Expressions of relative probative value for elaboration of evaluative opinion.

Comparison of levels of probative value	Qualitative expression
No difference between levels (e.g., very weak vs. very weak)	The probative value of Hypothesis(a) is *equal* to probative value of Hypothesis(b)
1 level of difference (e.g., e.g. very weak vs. weak)	Hypothesis(a) has a *slightly higher* probative value than Hypothesis(b)
2 levels of difference (e.g., very weak vs. good)	Hypothesis(a) has a *significantly higher* probative value than Hypothesis(b)
3 levels of difference (e.g., very weak vs. high)	Hypothesis(a) has a *much higher* probative value than Hypothesis(b)
4 levels of difference (e.g., very weak vs. very high)	Hypothesis(a) has a *very much higher* probative value than Hypothesis(b)

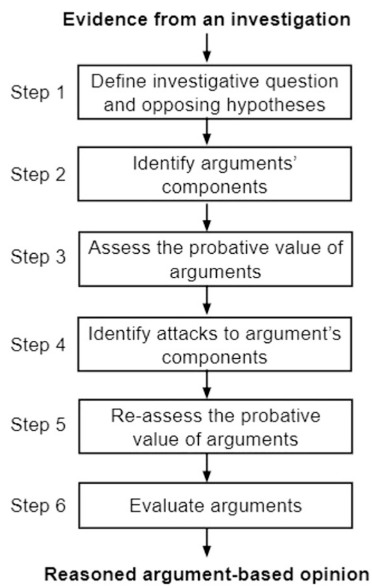

Fig. 2. Proposed Method for Argument-based Evaluative Opinion.

cence (European Convention on Human Rights (ECHR), 1950, Article 6(2)).

The method applies a rule that diverges from how argumentation theory typically is utilized. In general argumentation theory, one would construct an argument based on grounds that favour the hypothesis. However, to evaluate evidence, we apply the rule of including all evidence relevant to the investigative question under each hypothesis. Different grounds may be derived from all evidence under each hypothesis; nevertheless, the evidence should be included. By introducing this rule, the evaluator is forced to articulate the explanations that would allow the evidence to be linked to the hypothesis as well as an evaluation of their probative value.

Step 1: Define the investigative question and the opposing hypotheses

- Define the investigative question.
- Define Hypothesis(a) from the prosecutor's perspective.
- Define Hypothesis(b) from the defence's perspective.

Step 2: Identify arguments' components

- Identify all pieces of evidence (E) relevant to the investigative question. The same pieces of evidence should be considered for Hypothesis(a) and Hypothesis(b).
- Identify grounds (G) based on the evidence under Hypothesis(a) and Hypothesis(b).
- Determine inferential leaps (I) and supporting information, which enable inference from grounds (G) to Hypothesis(a) and construct the argument under it.
- Repeat the procedure for Hypothesis(b).

Step 3: Assess the probative value of arguments

- Assess the credibility of the evidence/grounds and the relevance of the inferential leaps by using the argument evaluation scales (Tables 1 and 2).
- Assess the probative value of relevance and credibility in context by using the argument matrix (Table 3).

Step 4: Identify attacks to arguments' components

- Identify any relevant attacks on credibility (Ac) of the existing grounds and associated evidence. Consider the robustness of the evidence base and point out any missing evidence.
- Identify any attacks against relevance (Ar) of the inferential leaps from ground to Hypothesis(a) and the associated backing. Consider the inferential leap from evidence to ground, and identify whether there are any gaps or logical breaches.
- Repeat the procedure for Hypothesis(b).

Step 5: Re-assess the probative value of arguments

- Consider the attacks in the former step and re-assess the probative value of the arguments.

Step 6: Evaluate arguments

- Evaluate the probative value of the evidence-based argument under Hypothesis(a) versus Hypothesis(b).
- Elaborate an opinion exposing the argument and the associated reasoning – including any attacks – underpinning Hypothesis(a) and Hypothesis(b), respectively. For a standardised expression of the relative probative value, refer to Table 4. The first column in the table provides options for the outcome when comparing the relative probative value of Hypothesis(a) vs. Hypothesis(b), illustrated by an example. The second column provides a qualitative expression to guide the formulation of the evaluation opinion.

4. Illustration of the proposed method

The proposed method is illustrated with two cases, elaborated in Sections 4.1 and 4.2.

4.1. Case 1: Apple Health app data and arson

Case description:

This fictitious case is adopted from Vink et al. (2022). During the questioned night, a man (Suspect Y) was suspected of setting a house on fire (Location II). He lived in the same neighbourhood (Location I). Eyewitness X stated that the fire started around 1:30 a.m. The suspect owned an iPhone 7 with the Apple Health app, which registers step count and travelled distances in time. The iPhone registered movement between 1:21 a.m. and 1:40 a.m. and had registered 1450 steps and 1250 meters. An eyewitness (Eyewitness Z) stated that the suspect entered a bar (Location III) around 1:40 a.m. The prosecution claimed the suspect walked from Location I to Location III via Location II (Route

1 - 1600 meters). The suspect told the police that he walked directly from Location I to Location III (Route 2 - 900 meters) and that he, therefore, did not pass the house that was set on fire. The computation in the paper by Vink et al. (2022) rendered a likelihood ratio of 5 in favour of Hypothesis(a). The assumptions concerning steps and travelled distance underpinning this result rest primarily on findings from experiments conducted by van Zandwijk and Boztas (2019).

An additional element was added to the original scenario: Suspect Y agreed to participate in a reconstruction of Route 1 and Route 2. He was asked to carry the iPhone 7 in the same place and walk the route as similarly as possible as he did the questioned night. He walked both routes at an even pace and carried the iPhone in the left front pocket of his trousers. The registered Apple Health app data was:

- Route 1 (1600 m): 1478 steps / 1390 meters and 18 minutes of movement.
- Route 2 (900 m): 810 steps / 730 meters and 13 minutes of movement.

4.1.1. *Proposed method applied to Case 1*

This section applies the method introduced in Section 3 to the case.

Step 1: Define the investigative question and opposing propositions

- The investigative question is: *Was Suspect Y at the Location II when the fire started?*
- Hypothesis(a): Suspect Y walked Route 1 (1600 meters from location I–II–III).
- Hypothesis(b): Suspect Y walked Route 2 (900 meters from location I-III).

Step 2: Identify arguments' components

- E1: Apple Health app data secured from Suspect Y's iPhone 7: 1450 steps / 1250 meters between 1:21 a.m. and 1:40 a.m.
- E2: Eyewitness X stated that the fire started at 1:30 a.m. at Location II (supported by other evidence).
- E3: Eyewitness Z stated that Suspect Y entered the bar at Location III around 1:40 a.m. (supported by other evidence).
- E4: Information from public registry → Suspect Y lives near Location I.

Hypothesis(a):

- G1a: Suspect Y walked from his home at 1:21 a.m., started the fire at Location II at 1:30 a.m. and walked further to the bar at Location III where he arrived at 1:40 a.m.
- I1a: The Apple Health App data of 1250 meters makes Route 1 of 1600 meters corresponds with the research showing that the Apple Health app tends to register a lesser distance than the actual when walking.
- I2a: The results from the reconstruction support the research showing that the Apple Health app registers fewer steps/meters than the actual distance when walking.
- I3a: Reconstruction of Route II with iPhone 7 and Apple Health app led to a registration of 730 meters travelled distance. This result corresponds with the research and makes Route II more likely.

Hypothesis(b):

- G1b: Suspect Y walked from his home at 1:21 a.m. and directly to location III (the bar), which is 900 meters apart, where he arrived at 1:40 a.m.
- I1b: Since the Apple Health app tends to register a lesser distance than the actual when walking, the registration of 1250 meters indicates that something may have influenced the registration, leading to more registered steps and longer travelled distance than Route II. The exceeded registration was caused by influences such as walking style, where and how the iPhone was carried, or some other unknown source of interference.

Step 3: Assess the probative value of arguments

- The probative value of Hypothesis(a) is *high*.
- The probative value of Hypothesis(b) is *weak*.

Reasoning:

The credibility of G1a under Hypothesis(a) is high (value C4) based on the evidential support from E1-E4. The relevance of I1 -I3 is high (value R4). Therefore, they make Hypothesis(a) very likely due to the research foundation combined with the reconstruction where similar tendencies were observed.

The credibility of G1b under Hypothesis(b) is high (value C4) based on the evidential support from E1-E4. The relevance of I1 is rated weak (value R2) since it is – based on research experiments (van Zandwijk and Boztas, 2019), less likely that walking 900 meters would render a higher registered distance. The reconstruction also supports the experimental findings and makes the Hypothesis(b) less likely.

Step 4: Identify attacks to arguments' components

- AcE1: There is no supporting information concerning Suspect Y's walking style or where he carried the iPhone on the night in question other than Suspect Y's account. It is uncertain whether Suspect Y carried the iPhone consistently during the walk on the night in question (hand, pocket, bag, etc.). It is also uncertain whether Suspect Y walked or ran the distance from Location I-III on the night in question. Hence, walking style and carrying location may have influenced the registered distance in the Apple Health app. Running is more likely than walking to increase the registered distance above the actual distance, especially when the iPhone is held in hand or kept in a trouser pocket (van Zandwijk and Boztas, 2019, Pages 129-130).

Attacks on Hypothesis(a):

- Ar1a: There is no supporting evidence that Suspect Y was at Location II when the fire started.

Attacks on Hypothesis(b):

- Ar1b: There is no supporting evidence that Suspect Y was at another place than Location II when the fire started.

Fig. 3 summarises the structure of the argument for Case 1 visually.

Step 5: Re-assess the probative value of arguments

- The probative value of Hypothesis(a) is *good* when attacks are considered.
- The probative value of Hypothesis(b) is *very weak* when attacks are considered.

Reasoning:

The probative value of Hypothesis(a) is changed from high to good based on the uncertainties introduced by the attacks. Although the scenario of Suspect Z walked from Location I via Location II before being observed at Location III is consistent with the iPhone Health app registration, the lack of credible information about walking style and where the phone was carried, combined with the absence of observations along Route 1 justifies to a decrease the probative value from high to good.

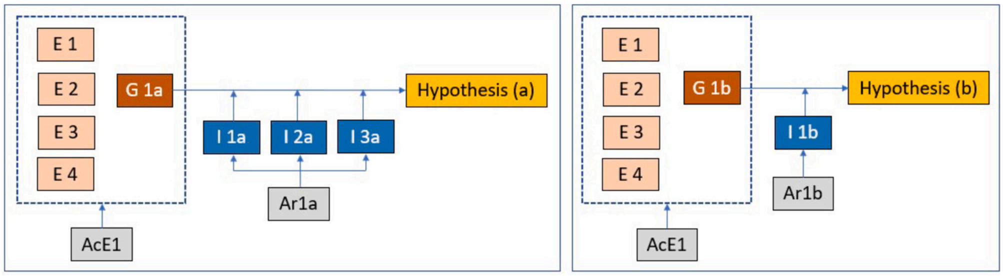

Fig. 3. Visual representation of the argument structure for Case 1.

Therefore, the attacks are equally relevant to Hypothesis(b) and justify decreasing the probative value from weak to very weak.

Step 6: Evaluate arguments

The evidence-based argument is of *significantly higher* probative value under Hypothesis(a) than under Hypothesis(b) (see Table 4). The reasoning behind this opinion is the following.

Reasoning:

The registered data in the iPhone's Apple Health app for the night in question deviates equally from the distance in both Hypothesis(a) and Hypothesis(b). However, research on how data is registered during different walking styles and iPhone carrying locations shows that a discrepancy from the actual distance is typical. Research shows that, when walking while carrying the iPhone 7 in the trouser's pocket, the registered distance in the Apple Health app is more likely to be lower than the actual distance as opposed to higher. Under the reconstruction, Suspect Y demonstrated that he had walked the distance and kept the iPhone in the front left pocket. Based on these grounds, the registration of 1250 meters has a higher probative value under Hypothesis(a) compared to Hypothesis(b). The evidential foundation could, however, be more robust. There is no supporting evidence, such as witness observations or CCTV evidence, supporting that Suspect Y took either Route 1 or 2 during the relevant timeframe, which equally weakens the probative value of the evidence-based argument under Hypothesis(a) and Hypothesis(b).

4.2. *Case 2: Theft from purse caught on CCTV*

Case description:

The example case is largely inspired by a real case from 2009 concerning the theft of a wallet with 800 NOK (approx. 70 Euro) from an older woman's purse while she was shopping at a grocery store. The theft was recorded by the CCTV camera in the store, and although the CCTV footage was blurry, a police officer identified the suspect as Suspect A (anonymised). The police officer recognised Suspect A since he had seen her before. Suspect A was arrested and later sentenced in the city court based mainly on the CCTV evidence (TSTRO-2009-172386); she appealed, but the sentence was upheld by the appeal court (LF-2010-40043). An appeal to the supreme court was not accepted (HR-2010-01789-U). She applied for reopening the case to the Norwegian Criminal Cases Review Commission (the commission), and after two rejections (GK-2014-183; GK-2017-26), the commission decided to reopen the case after the third application (GK-2017-138). The woman was acquitted in 2019 (LE-2019-24145). For the purpose of illustration, some case details have been added or modified in the example below.

4.2.1. *Method applied to Case 2*

This section applies the method introduced in Section 3 to the case.

Step 1: Define investigative question and opposing hypotheses

- The investigative question is: *Who is the person taking the wallet from an older woman's purse?*
- Hypothesis(a): Suspect A is the person on the CCTV footage.
- Hypothesis(b): Suspect A is not the person on the CCTV footage.

Step 2: Identify arguments' components

- E1: CCTV footage obtained from the CCTV system in the grocery store. The footage is of low quality due to low resolution and deviations in the colour representation (e.g. red colour appears as pink).
- E2: The police officer who has seen the suspect in person has provided an identification report – identifying Suspect A as the woman in the footage, referring to former knowledge of her.
- E3: Photograph of Suspect A taken by the police on the day of the arrest, the same day as the E1 was recorded. The image is of high resolution and provides an accurate representation of the facial features of Suspect A.

Hypothesis(a)

- G1a: The visual facial features of the person on the CCTV footage correspond to those of Suspect A.
- I1a: The identification by the police officer is a recognition based on former knowledge of Suspect A (as opposed to an identification based on comparison of images).

Hypothesis(b)

- G1b: The person's facial features on the CCTV footage correspond with those of Suspect A but are inconclusive due to the ambiguity caused by low quality.
- I1b: The police officer misidentified the suspect as the woman in the footage due to low resolution and colour deviations of the CCTV footage.
- I2b: It is a higher risk of misidentification when the footage on which it is based is ambiguous due to low quality (see, e.g., Keval and Sasse, 2008; Lee et al., 2009).
- I3b: The low quality of the CCTV footage (E1) does not provide for an accurate evaluation of the facial features against several of the visible features in the high-resolution photograph (E3).

Step 3: Assess the probative value of arguments

- The probative value of Hypothesis(a) is *weak*.
- The probative value of Hypothesis(b) is *weak*.

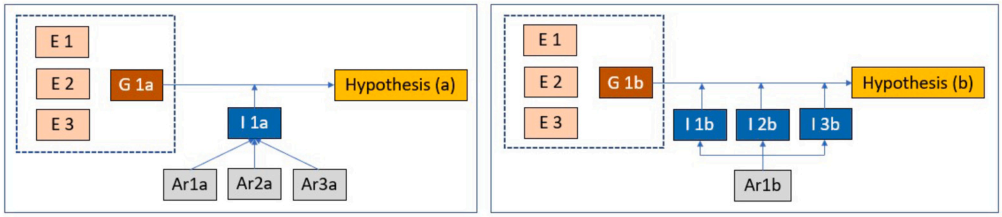

Fig. 4. Visual representation of the argument structure for Case 2.

Reasoning:

The credibility of G1a under Hypothesis(a) is weak (value C2) since, although it is based on evidence from a source of very high credibility, the CCTV evidence is elastic despite some empirical support in the E2. Considering the relevance, the inferential leap I1a makes Hypothesis(a) more likely than unlikely (value R3) due to the police officer's former knowledge and the corresponding facial features between E1 and E3. Therefore, the probative value of Hypothesis (a) based on the credibility and relevance in context is weak.

The credibility of G1b under Hypothesis(b) is weak (value C2). It is based on evidence from a source of very high credibility, although the CCTV evidence is elastic despite some empirical support in E2. Considering the relevance under Hypothesis(b), the inferential leaps (I1b-I2b-I3b) make Hypothesis(b) more likely than unlikely (value R3). However, the low resolution in the CCTV footage increases the risk of misidentification due to ambiguous features. Therefore, the probative value of Hypothesis(b) based on the credibility and relevance in context is also weak.

Step 4: Identify attacks to arguments' components

Attacks on Hypothesis(a):

- Ar1a: There is no empirical support from other information sources, e.g., CCTV footage or location data linking Suspect A to the time and place of the theft.
- Ar2a: A systematic evaluation of facial image comparison would probably be infeasible due to the low quality of the CCTV.
- Ar3a: Although the visible features correspond with Suspect A's features, there are no significant or rare features such as scars or tattoos, and the features thus correspond with many other similar-aged women.

Attacks on Hypothesis(b):

- Ar1b: There is no empirical support suggesting that the suspect was at a different location during the theft.

Fig. 4 summarises the structure of the argument for Case 2 visually.

Step 5: Re-assess the probative value of arguments

- The probative value of Hypothesis(a) is *very weak* when attacks are considered.
- The probative value of Hypothesis(b) is *very weak* when attacks are considered.

Reasoning:

The attacks show that both hypotheses lack empirical support that would increase the credibility of the grounds, and both are adjusted from weak to very weak.

Step 6: Evaluate arguments

Hypothesis(a) that Suspect A is the person in the CCTV footage has *equal* probative value as Hypothesis(b) that Suspect A is not the person in the CCTV footage (see Table 4). The reasoning behind this opinion is the following:

Reasoning:

The probative value of the evidence-based arguments under both hypotheses is very weak, mainly due to the low quality of the CCTV evidence. Although the police officer states in his report that he recognises the person in the CCTV footage as Suspect A, there is no other evidence supporting or contradicting that Suspect A was in the store at the time of the theft. The footage is of too low-quality to perform an identification based on the high-resolution image (E3). The blurred CCTV footage corresponds with the facial features of many similar-aged women and does not provide a sufficient foundation to discriminate between Hypothesis(a) and Hypothesis(b).

5. Discussion

The proposed *Argument-Based Method for Evaluative Opinions* provides a structured way of evaluating evidence and arriving at an opinion and aids practitioners to be transparent about the reasoning behind their assessments. As the method obliges articulation of the reasoning from evidence/ground to hypothesis, implicit assumptions, tacit knowledge and generalisations become explicit not only to the practitioners themselves but also become available for peer scrutiny. This may reveal flawed judgements based on insufficient knowledge or flawed assumptions which otherwise would remain hidden. It serves as a method that gives the legal decision-maker insights into each component of the argument, the credibility of each piece of evidence, the inferential leaps that make the evidence relevant to the case, and to which of these components the identified uncertainty relates. Increased transparency also facilitates cross-examination in court, which is a crucial safeguard for the fair administration of justice.

The proposed method may be used alone or combined with other methods. It does not, in itself, require any computations. However, support for inferential leaps may very well be based on calculations and statistics, as shown in Case 1 (Section 4.1) – where the original work by Vink et al. (2022) calculated the likelihood ratio for the opposing hypotheses. As individual pieces of evidence may be subject to Bayesian probabilistic analyses, the proposed argument-based method can be applied as an "add-on" (Di Bello and Verheij, 2020; Prakken, 2020). It is flexible and enables the combination of multiple evidence of similar kinds as well as different types of evidence.

A strength of the proposed method is the ability to consider and describe the robustness of the evidence foundation and the empirical uncertainty introduced by missing pieces of evidence. As shown in Step 5 of Case 1, the probative value of both arguments is lowered after consideration of attacks highlighting evidential gaps. However, the adjustment does not affect their relative probative value. This allows the legal decision-makers insight into essential aspects when considering whether the evidence is sufficient to meet the standard often expressed as “proven beyond any reasonable doubt”.

The dialectic process of taking one perspective at a time may minimise cognitive bias. The process of articulating argument components may counter the confirmation bias, which is a tendency not only to look for information that corresponds with one's beliefs and to overlook and explain away contradicting information (Nickerson, 1998). The argument-based method proposed may help the practitioner to avoid fast and intuitive “system 1 thinking” (Kahneman, 2011), and rather apply analytical and slow “system 2 thinking” by scrutinising their own assumptions and decision making, explicating the components and logic within each argument, and systematically identifying uncertainty (attacks) under each hypothesis. Case 2 (Section 4.2) serves as an illustration of this point. Intuitively, one would put much weight on a police officer claiming to recognise a person which he or she has prior knowledge of. The argument-based method directs the attention, forcing an evaluation of the digital evidence upon which the identification is founded. A systematic evaluation reveals that, despite the police officer's strong conviction, the poor quality footage diminishes its probative value – or, as the adage warns: “you can't make a silk purse out of a sow's ear”.

It should be emphasised that, as the proposed method involves subjectivity and discretion, it does not entirely eliminate cognitive bias influencing the many judgements and decisions made throughout the process. Despite that, the explicit reasoning will make a possible bias more visible to peers, which increases the possibility of bias detection and correction.

The uncertainty is articulated in qualitative terms and encompasses not only the credibility of evidence but also a separate evaluation of the relevance of inferences. Since the digital forensics domain is challenged by changing technology, many issues subject to evaluation do not have an associated reference database. Therefore, the evaluations must often rely on the practitioners' expertise combined with rigorous experimentation and testing. Under such circumstances, the proposed argument-based method may be a helpful supplement in their toolbox for performing subjective and non-numerical probabilistic analyses as a basis for evaluative opinions.

A limitation of the proposed method is that it becomes complex if many pieces of evidence and grounds are to be considered under one hypothesis, with different inferential leaps and supporting tests and experimentation. To avoid losing oversight, visualisation may be helpful.

The granularity of the values for assessing probative value may also be a limitation. As argued by Casey (2020), a detailed scale grasps the result of smaller adjustments. The argument matrix (Table 3) may thus sometimes not capture minor changes in relevance or credibility. However, since the method highlights transparency in reasoning, such adjustments may be expressed in qualitative terms when conveying the reasoning behind the selected values.

As with other hypothesis-based methods, the fairness of the outcome depends on whether the hypotheses are balanced towards the ultimate investigative question of guilt. Suppose both hypotheses are framed as guilt hypotheses in Step 1 of the proposed method (Fig. 2). In that case, the outcome will render a probative value of one guilt hypothesis vs. the other and, consequently, not the probative value of a plausible hypothesis pointing towards innocence. To avoid a misleading outcome, the practitioner should pay close attention to this issue when defining hypotheses in Step 1 and include hypotheses such as “Suspect Y walked Route 2 (900 meters from location I-III)” entailing that the suspect did not pass the house that was set on fire (Case 1), or “Suspect A is not the person in the CCTV footage” (Case 2) meaning that the suspect did not commit the theft that was caught on CCTV.

In addition to evaluation, the proposed argument-based method may be used as an investigative tool at the pre-trial stage. If information collection is still ongoing, the method may be employed to assess what needs to be investigated further through, for example, additional collection of evidence or through testing and experimentation to reduce uncertainty related to the existing evidence – as a logical reconstruction mechanism. It can also supplement other analytical approaches used during the investigation stage, such as the Analysis of Competing Hypotheses (ACH) (Heuer Jr., 1999, 2005). ACH is primarily used for assessing and eliminating hypotheses; however, it does not inform the probative value of the evidence under each hypothesis or relative to another hypothesis in qualitative terms. ACH applies an atomic approach to evidence evaluation, and assesses the consistency of each piece of information (evidence) with each hypothesis. Therefore, it fails to capture the dependencies between individual pieces of evidence. For example, the credibility of a ground may be conditioned by another ground, an aspect which the “atomic consistency test” in ACH fails to capture. The proposed argument-based method may thus supplement the ACH method, providing insights into the logical connection between components under a hypothesis and the probative value of the evidence under one hypothesis vs. the other.

6. Conclusion

The paper presents the *Argument-Based Method for Evaluative Opinions*, which is a novel method for a balanced, non-numerical, evaluation of evidence. The method is primarily founded on general concepts from argumentation theory combined with key principles and concepts from probability and narrative/scenario-based approaches and meets three objectives stated at the end of Section 1.

The proposed method may mitigate bias by obliging the practitioner to assess the argument under two opposing hypotheses. To ensure fairness and adherence to the presumption of innocence, at least one of the hypotheses should be framed as an “innocence” hypothesis, as demonstrated in example Case 1 and Case 2 in this paper. In both cases, digital evidence is central, and the application of the method shows that it is useful for determining the probative value of the evidence under one hypothesis vs. another. However, as with any novel method, application and testing in a real life setting is essential for examining its effectiveness.

The proposed method guides practitioners, through several steps, to follow a rationale from identifying opposing hypotheses, identifying argument components and possible attacks to them, and, ultimately, assessing the probative value of the arguments. This leads to the elaboration of an evaluative expert opinion which follows logically from the assessment of the opposing hypotheses. Therefore, documentation of all the steps promotes transparency in reasoning about the credibility of evidence and grounds and the reliability of the inferential leaps from these steps and onwards to a hypothesis. Transparency concerning reasoning around the value and uncertainty associated with the evidence is crucial to facilitate audit, peer scrutiny and effective cross-examination in court.

The proposed method applies a qualitative approach for argument evaluation, providing a scale for assessment of the credibility of evidence and ground (Table 1) and of the relevance of inferential leaps (Table 2). Such support scales allow practitioners without competency and experience in quantitative approaches to use the proposed method. Despite the qualitative foundation, the proposed method is flexible and can be combined with quantitative methods for uncertainty assessment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

We would like to thank Professor Eivind Kolflaath for the insightful comments and valuable feedback on drafts of this article. We also thank the anonymous reviewers for their helpful comments and suggestions.

References

Anderson, T., Schum, D., Twining, W., 2005. Analysis of Evidence. Cambridge University Press.

Bex, F., 2021. Philosophical Foundations of Evidence Law. Oxford University Press, pp. 183–198. Chapter Argumentation and Evidence.

Boddington, R., 2012. A case study of the challenges of cyber forensics analysis of digital evidence in a child pornography trial. In: Conference on Digital Forensics, Security and Law. ADFSL, pp. 155–172.

Boddington, R., 2016. Practical Digital Forensics. PACKT Publishing.

Casey, E., 2011. Digital Evidence and Computer Crime, 3 ed. Elsevier Press.

Casey, E., 2020. Standardization of forming and expressing preliminary evaluative opinions on digital evidence. Forensic Sci. Int., Digit. Investig. 32, S1–S10. https://doi.org/10.1016/j.fsidi.2019.200888.

Casey, E., Jaquet-Chiffelle, D.O., Spichiger, H., Ryser, E., Souvignet, T., 2020. Structuring the evaluation of location-related mobile device evidence. Forensic Sci. Int., Digit. Investig. 32, S1–S9. https://doi.org/10.1016/j.fsidi.2020.300928.

Cederblom, J., Paulsen, D.W., 2012. Critical Reasoning: Understanding and Criticizing Arguments and Theories, 7th ed. Cengage Learning, Boston, Wadsworth.

Cohen, L.J., 1977. The Probable and the Provable. Oxford University Press, Oxford, UK.

Cohen, L.J., 1989. An Introduction to the Philosophy of Induction and Probability. Oxford University Press, Oxford, UK.

Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J., 1998a. A hierarchy of propositions; deciding which level to address in casework. Sci. Justice 38, 231–239. https://doi.org/10.1016/S1355-0306(98)72117-3.

Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J., 1998b. A model for case assessment and interpretation. Sci. Justice 38, 151–156. https://doi.org/10.1016/S1355-0306(98)72099-4.

Di Bello, M., Verheij, B., 2020. Evidence & decision making in the law: theoretical, computational and empirical approaches. Artif. Intell. Law 28, 1–5. https://doi.org/10.1007/s10506-019-09253-0.

European Convention on Human Rights (ECHR), 1950. Article 6: Right to a fair trial. European convention on human rights. [Online]. https://www.echr.coe.int/Documents/Convention_ENG.pdf. (Accessed 3 April 2023).

European Network of Forensic Science Institutes, 2015. ENFSI guideline for evaluative reporting in forensic science: strengthening the evaluation of forensic results across Europe (STEOFRAE). [Online]. http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf. (Accessed 3 April 2023).

Forensic Science Regulator, 2021. FSR-C-118 development of evaluative opinions. [Online]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_Interpretation_Appendix_Issue_1__002_.pdf. (Accessed 3 April 2023).

Franqueira, V., Horsman, G., 2020. Towards sound forensic arguments: structured argumentation applied to digital forensics practice. Forensic Sci. Int., Digit. Investig. 32, S1–S8. https://doi.org/10.1016/j.fsidi.2020.300923.

GK-2014-183, March 19, 2015. Rejection in the Criminal Cases Review Commission; Lovdata Pro.

GK-2017-138, November 15, 2018. Decision to reopen case by the Criminal Cases Review Commission.

GK-2017-26, March 8, 2017. Rejection in the Criminal Cases Review Commission.

Helm, R., 2022. Wrongful conviction in England and Wales: an assessment of successful appeals and key contributors. Wrongful Conviction Law Rev. 3, 196–217. https://doi.org/10.29173/wclawr79.

Heuer Jr., R.J., 1999. The Psychology of Intelligence Analysis. Center for Study of Intelligence, Central Intelligence Agency, Washington, DC. https://apps.dtic.mil/sti/pdfs/ADA500078.pdf.

Heuer Jr., R.J., 2005. How does analysis of competing hypotheses (ACH) improve intelligence analysis? Retrieved from http://www.pherson.org/wp-content/uploads/2013/06/06.-How-Does-ACH-Improve-Analysis_FINAL.pdf.

HR-2010-01789-U, October 21, 2010. Verdict by The Supreme Court; Lovdata Pro.

Kahneman, D., 2011. Thinking Fast and Slow. Macmillan.

Keval, H.U., Sasse, M.A., 2008. Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In: Proceedings of Mobile Multimedia/Image Processing, Security, and Applications (SPIE'2008), pp. 189–203.

King, G.L., 2006. Forensics plan guide. https://www.giac.org/paper/gcfa/283/forensic-investigation-plan-cookbook/108356.

Kolflaath, E., 2023. Ordlyd og resonnement. Fagbokforlaget.

Kwan, M., Overill, R., Chow, K., Tse, H., Law, F., Lai, P., 2011. Sensitivity analysis of Bayesian networks used in forensic investigations. In: Proceedings of the 2011 IFIP International Conference on Digital Forensics. Springer, pp. 231–243.

LE-2019-24145, May 10, 2019. Verdict by The Eidsivating Appeal Court; Lovdata Pro.

Lee, W.J., Wilkinson, C., Memon, A., Houston, K., 2009. Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage. AXIS Online J. CAHId 1, 19–28.

LF-2010-40043, August 10, 2010. Verdict by The Frostating Appeal Court; Lovdata Pro.

Mackor, A.R., Van Koppen, F., 2021. Philosophical Foundations of Evidence Law. Oxford University Press, pp. 215–230. Chapter the Scenario Theory about Evidence in Criminal Law.

Newman, S.E., Marshall, C.C., 1991. Pushing Toulmin Too Far: Learning from an Argument Representation Scheme. Technical Report SSL-92-45. Xerox PARC.

Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2, 175–220. https://doi.org/10.1037/1089-2680.2.2.175.

Overill, R.E., Silomon, J.A.M., Chow, K.P., Tse, R., 2013. Quantification of digital forensic hypotheses using probabilistic theory. In: Proceedings of the 8th International Workshop on Systematic Approaches to Digital Forensics Engineering (SADFE'13). IEEE, pp. 71–75.

Pennington, N., Hastie, R., 1993. The Story Model for Juror Decision Making. Cambridge University Press.

Prakken, H., 2020. A new use case for argumentation support tools: supporting discussions of Bayesian analyses of complex criminal cases. Artif. Intell. Law 28, 27–49. https://doi.org/10.1007/s10506-018-9235-z.

Prakken, H., Bex, F., Mackor, A.R., 2020. Ditors' review and introduction: models of rational proof in criminal law. Top. Cogn. Sci. 12, 1053–1067. https://doi.org/10.1111/tops.12519.

Reitan, M., 2016. Argumentasjonsteori. Fagbokforlaget.

Schum, D.A., 2001. The Evidential Foundations of Probabilistic Reasoning. Northwestern University Press, Evanston, IL.

Schum, D.A., 2002. The Dynamics of Judicial Proof. Studies in Fuzziness and Soft Computing. Springer, pp. 307–336. https://doi-org.chain.kent.ac.uk/10.1007/978-3-7908-1792-8_16. Chapter Species of Abductive Reasoning in Fact Investigation in Law.

Smit, N.M., Morgan, R.M., Lagnado, D.A., 2018. A systematic analysis of misleading evidence in unsafe rulings in England and Wales. Sci. Justice 58, 128–137. https://doi.org/10.1016/j.scijus.2017.09.005.

Sunde, N., 2021. What does a digital forensics opinion look like? A comparison study of digital forensics and forensic science reporting practices. Sci. Justice 61, 586–596. https://doi.org/10.1016/j.scijus.2021.06.010.

Tart, M., 2020. Opinion evidence in cell site analysis. Sci. Justice 60, 363–374. https://doi.org/10.1016/j.scijus.2020.02.002.

Tecuci, G., Schum, D.A., Marcu, D., Boicu, M., 2016. Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots. Cambridge University Press.

Toulmin, S.E., 1958. The Uses of Argument, 1 ed. Cambridge University Press.

Toulmin, S.E., 2003. The Uses of Argument, 2 ed. Cambridge University Press.

TSTRO-2009-172386, December 15, 2009. Verdict in Sør-Trøndelag City Court; Lovdata Pro.

van Zandwijk, J.P., Boztas, A., 2019. The iphone health app from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence? Digit. Investig. 28, S126–S133. https://doi.org/10.1016/j.diin.2019.01.021.

Verheij, B., 2012. Integrating argumentation, narrative and probability in legal evidence (position paper). In: 3rd Workshop on Computational Models of Narrative (CMN 2012), Massachusetts Institute of Technology, pp. 174–175. http://narrative.csail.mit.edu/ws12/proceedings.pdf.

Verheij, B., Bex, F., Timmer, S.T., Vlek, C.S., Meyer, J.J.C., Renooij, S., Prakken, H., 2016. Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. Law Probab. Risk 15, 35–70. https://doi.org/10.1093/lpr/mgv013.

Vink, M.M., Sjerps, M.M., Boztas, A.A., 2022. Likelihood ratio method for the interpretation of iphone health app data in digital forensics. Forensic Sci. Int., Digit. Investig. 41, 1–16. https://doi.org/10.1016/j.fsidi.2022.301389.

Wagenaar, W.A., Van Koppen, P.J., Crombag, H.F., 1993. Anchored Narratives: The Psychology of Criminal Evidence. St Martin's Press.

Walton, D.N., 2002. Legal Arguments and Evidence. Penn State University Press.