

A New Use Case for Argumentation Support Tools: Supporting Discussions of Bayesian Analyses of Complex Criminal Cases

Henry Prakken

Received: date / Accepted: date

Abstract In this paper a new use case for legal argumentation support tools is considered: supporting discussions about analyses of complex criminal cases with the help of Bayesian probability theory. By way of a case study, two actual discussions between experts in court cases are analysed on their argumentation structure. In this study the usefulness of several recognised argument schemes is confirmed, a new argument scheme for arguments from statistics are proposed, and an analysis is given of debates between experts about the validity of their arguments. From a practical point of view the case study yields insights into the design of support software for discussions about Bayesian analyses of complex criminal cases.

Keywords Argument schemes · reasoning about evidence · probability · argumentation support

1 Introduction

There is an ongoing debate on models of rational evidential reasoning in criminal cases. Both argumentation-based, story-based and Bayesian approaches have been proposed (Pardo and Allen, 2008; Kaptein et al, 2009; Fenton and Berger, 2016; Verheij et al, 2016). In this paper¹ I remain neutral with respect to this debate. Instead I want to argue that even if Bayesian thinking is adopted as the overall model of legal evidential reasoning, there is still one aspect of this form of reasoning that is clearly argumentative in nature, namely, debates about a proposed Bayesian analysis of a case or some aspects of it. This observation is theoretically interesting but also has practical implications for support systems for legal proof and crime investigation. Forensic experts increasingly use Bayesian probability theory as their theoretical framework and they increasingly use software tools for designing Bayesian networks. In crime investigation or in court the need may arise to record the pros and cons of the various design decisions embodied in these analyses and argumentation support technology may be of use here. This holds both for Bayesian analyses of specific aspects of a case

Department of Information and Computing Sciences, Utrecht University, and Faculty of Law, University of Groningen, The Netherlands
E-mail: h.pakken@uu.nl

¹ This paper is an extended version of Prakken (2017).

(such as assessments of the evidential value of a piece of evidence) and for more general uses of Bayesian probability theory to say something about the probability of guilt given the available evidence.

In AI & Law and related fields various argumentation support systems have been proposed; cf. e.g. Van den Braak (2010); Scheuer et al (2010). Such systems do not themselves produce arguments but support humans in formulating, structuring and evaluating their own arguments or arguments of others. Some supposed benefits of such support systems are that the human user's thinking can be improved, that arguments can be drafted in a better way and be communicated more easily to others and that arguments can be connected to textual sources (such as case files) to make these sources more transparent. Moreover, computational tools may be used to evaluate debates in a more precise way than is possible with unstructured natural-language arguments. Most argumentation support systems proposed thus far are for quite general application domains, such as e-democracy (Wardeh et al, 2013). This paper instead studies a quite specific use case of argumentation support: argumentation about Bayesian analyses of complex criminal cases as attempts to say something about the probability of guilt given the available evidence. To obtain insight in the requirements for such support system, in particular in the form of argumentation-based add-ons to Bayesian-network software tools, it is important to examine actual discussions between experts about Bayesian analyses of complex criminal cases. Doing so is the purpose of this paper. A side effect may be increased understanding of probabilistic reasoning about evidence in criminal cases, which is worthwhile from a theoretical perspective. Note that the relevance of the present study is independent of the much debated question whether Bayesian probability theory is applicable to legal proof at all. Its relevance is instead given by the fact that Bayesian analyses are as a matter of fact increasingly being presented by forensic experts in court, so that discussions about the merits of such analyses inevitably arise, and such discussions are essentially argumentative in nature.

The analysis will take the form of a case study of two recent Dutch criminal cases in which I was appointed by courts to comment on a Bayesian analysis proposed by an expert of the prosecution. In both cases his analysis concerned not just a specific aspect of the case but the entire case. This raises the issue to which extent the studied cases are typical, since Bayesian analyses of entire complex criminal cases are still rare in the courtroom. The usual uses of Bayes in the courtroom concern individual pieces of evidence, especially random match probabilities of forensic trace evidence (DNA, tyre marks, shoe prints, finger prints, glass pieces).

In the present paper I analyse to what extent our expert reports and written replies contain arguments that can be classified as instances of argument schemes or as applications of critical questions of these schemes (Section 4). I then use this analysis to formulate requirements for an implemented support system (Section 5). But first I present the cases to be studied (Section 2) and introduce formal preliminaries concerning probability theory and argumentation (Section 3).

2 The cases

In the *Breda Six* case three young men and three young women were accused of jointly having killed a woman in the restaurant of her son in the evening or night after closing time, in the year of 1993. The six were initially convicted in two instances in 1994 and 1995, mainly on the basis of confessions of the three female suspects. The three male suspects have always claimed to be innocent while during the appeal case one of the female suspects retracted her

confession. In 1998 the case was brought to the attention to the Dutch committee for evaluating closed criminal cases, because of doubts about the truthfulness of the confessions of the three convicted women. This committee referred the case to the Dutch Supreme Court, which in 2012 decided to reopen the case. After a new police investigation the six were tried again by the court of appeal of The Hague and on 14 October 2015 they were again all found guilty, mainly on the ground that new evidence had confirmed the reliability of the confessions.

The prosecution (the Advocate General, or AG for short) had on 17 March 2015 brought in an 80 page expert report written by Dr. Alkemade, containing a Bayesian analysis of the entire case. Alkemade (henceforth referred to as A) is an atmospheric and climate physicist who until October 2015 worked at SRON, The Netherlands Institute for Space Research, based in Utrecht. A claimed that he was able to give a Bayesian analysis of the case since he had experience with using Bayesian probability theory in his work as a physicist. In his report, he concluded that on the basis of the evidence considered by him the probability that at least one of the six suspects was involved in the crime was 99,7%.

On 28 April 2015 I was appointed as an expert witness in the case by the investigating judge in the case, with the task to assess and evaluate A's report. I delivered my 41 page report on 28 June 2015. In my report I was critical about both A's level of expertise and his method. In its final verdict, the court ruled that A could be regarded as an expert for the purpose of the case but that his method cannot be regarded as a reliable method for analysing complex criminal cases. The court therefore decided to disregard A's conclusions. These wordings, with "conclusions" instead of "report", suggest that the court may have wanted to give itself the freedom to use some elements of A's report, although its reasoning in the final verdict was not expressed in terms of Bayesian probability theory.

In the *Oosterland* case a person was accused of being responsible for 18 (in appeal 16) small arson cases in the small town of Oosterland in a six-month period in 2013. In the initial case on 13 February 2014 the suspect was acquitted, mainly on the grounds that the two main testimonies (of a witness and of another suspect in the same cases) were unreliable. In the appeal case the prosecution again brought in a report by A (on 1 October 2015), this time 79 pages long. This time A concluded that on the basis of the evidence considered by him the probability that the suspect was involved in several of the arson cases was at least 99,8%.

On 19 January 2016 I was appointed as an expert witness in the case by the investigating judge in the appeal case, now with the more specific task to assess the reliability of A's method and of the way he had applied his method to the case. On 30 June I delivered my 42 page report, with essentially the same conclusions as in my report for in the Breda Six case. A then wrote a reply to my report and I wrote a reply to his reply. On 22 November 2016 the court of appeal convicted the suspect of 7 arson cases and acquitted him of the remaining 9 cases. Again the court's reasoning was not expressed in terms of Bayesian probability theory. The court stated without further explanation that it had chosen to disregard A's report "considering" my criticism.

3 Formal preliminaries

3.1 Bayesian probability theory

Probability theory (Hacking, 2001) defines how probabilities between 0 and 1 (or equivalently between 0% and 100%) can be assigned to the truth of statements. As for notation,

$Pr(A)$ stands for the unconditional probability of A while $Pr(A | B)$ stands for the conditional probability of A given B . In criminal cases the court is (on a Bayesian account) interested in the conditional probability $Pr(H | E)$ of a hypothesis of interest (for instance, that the suspect is guilty of the charge) given evidence E (where E may be a conjunction of individual pieces of evidence). For any statement A , the probabilities of A and $\neg A$ add up to 1. The same holds for $Pr(A | C)$ and $Pr(\neg A | C)$ for any C . Two pieces of evidence E_1 and E_2 are said to be statistically independent given a hypothesis H if learning that E_2 is true does not change $Pr(E_1 | H)$, i.e., if $Pr(E_1 | H \wedge E_2) = Pr(E_1 | H)$. The axioms of probability imply that such independence is symmetric. The axioms also imply the following theorems (here given in odds form). Let E_1, \dots, E_n be pieces of evidence and H a hypothesis. Then:

$$\frac{Pr(H | E_1 \wedge \dots \wedge E_n)}{Pr(\neg H | E_1 \wedge \dots \wedge E_n)} = \frac{Pr(E_n | H \wedge E_1 \wedge \dots \wedge E_{n-1})}{Pr(E_n | \neg H \wedge E_1 \wedge \dots \wedge E_{n-1})} \times \dots$$

$$\dots \times \frac{Pr(E_2 | H \wedge E_1)}{Pr(E_2 | \neg H \wedge E_1)} \times \frac{Pr(E_1 | H)}{Pr(E_1 | \neg H)} \times \frac{Pr(H)}{Pr(\neg H)}$$

This formula is often called the *chain rule* (in odds form). The fractions on the extreme right and left are, respectively, the *prior* and *posterior odds* of H and $\neg H$. Given that probabilities of H and $\neg H$ add up to 1, the *prior*, respectively, *posterior probability* of H can be easily computed from them.

If all of E_1, \dots, E_n are statistically independent from each other given H , then the chain rule reduces to

$$\frac{Pr(H | E_1 \wedge \dots \wedge E_n)}{Pr(\neg H | E_1 \wedge \dots \wedge E_n)} = \frac{Pr(E_n | H)}{Pr(E_n | \neg H)} \times \dots \times \frac{Pr(E_1 | H)}{Pr(E_1 | \neg H)} \times \frac{Pr(H)}{Pr(\neg H)}$$

This is the formula used by A in his reports. Its attractiveness is that to determine the posterior odds of a hypothesis, it suffices to, respectively, multiply its prior odds with the so-called likelihood ratio, or evidential force, of each piece of evidence. For each piece of evidence E_i all that needs to be specified is how much more or less likely E_i is given H than given $\neg H$. If this value exceeds (is less than) 1, then E_i makes H more (less) probable compared to before E_i was known, while if the value equals 1, the probability of H remains the same, so E_i is irrelevant for H .

Elegant as this way of thinking is, it is usually not applicable since often the global independence assumption concerning the evidence is not justified. Hence the name *naive Bayes*. Resorting to the general version of the theorem, the chain rule, is often also cumbersome, because of the many combinations of pieces of evidence that have to be taken into account. As a solution, *Bayesian networks* (Fenton and Neil, 2013) have been proposed, which graphically display possible independencies with directed links between nodes representing probabilistic variables (e.g. statements that can be true or false). For each value of each node, all that needs to be specified is its conditional probability given all combinations of all values of all its parents. Evidence can be entered in the network by setting the probability of the value of the corresponding node to 1, after which the probabilities of the values of the remaining nodes can be updated. For present purposes the most relevant observation is that to specify a Bayesian network, not only probabilities but also specific (in)dependencies have to be asserted.

3.2 Argumentation

Argumentation is the process of evaluating claims by providing and critically examining grounds for or against the claim. An important notion here is that of argument schemes (Walton et al, 2008), which capture typical forms of arguments as a scheme with a set of

premises and a conclusion, plus a set of critical questions that have to be answered before the scheme can be used to derive conclusions. If a scheme is deductively valid, that is, if its premises guarantee the conclusion, then all critical questions of a scheme ask whether a premise is true. If a scheme is defeasibly valid, that is, if its premises create a presumption in favour of its conclusion, then the scheme also has critical questions pointing at exceptional circumstances under which this presumption is not warranted. One formal approach to argumentation, *ASPIC*⁺ (Modgil and Prakken, 2014), formalises argument schemes as (deductive or defeasible) inference rules and critical questions as pointers to counterarguments: underminers attack an argument’s premise, undercutters state that there is an exception to a defeasible rule, and rebuttals have a conclusion that contradicts the conclusion of a defeasible inference. Arguments can be formed by chaining inference rules into directed graphs (which are trees if no premise is used more than once). Conflicts between arguments can be resolved with a given relative notion of argument strength, to see which arguments defeat each other. Then Dung’s (1995) abstract theory of argument evaluation can be used to determine which arguments are acceptable.

In the present paper argument schemes and their critical questions will be semiformally displayed. Critical questions asking whether the premises of the scheme are true will be left implicit. As the formal background *ASPIC*⁺ is assumed but the analysis will be such that it can also be formalised in similar argumentation formalisms or in related formalisms such as Defeasible Logic (Governatori et al, 2004). A formal background can provide a semantics for the notations and it can support automatic evaluation of reconstructed discussions. For example, as described by Bex et al (2013), the reconstruction could be stored in the Argument Interchange Format format and then be exported to implementations of an argumentation logic, such as the online TOAST implementation of *ASPIC*⁺ (<http://toast.arg-tech.org>).

4 The case study

In this section I discuss arguments from the written expert reports, the written replies and (when relevant) the verdicts that can be classified as instances of argument schemes or as applications of critical questions of these schemes. Most of the schemes are taken from the literature but in two cases a new scheme will be proposed.

All schemes will be presented semiformally in the following format:

Scheme name

Premise 1

...

Premise n

Conclusion

Critical questions:

Question 1

...

Question m

The double horizontal line indicates that the scheme is presumptive. A few times deductive schemes will be listed; they will be displayed with a single horizontal line.

4.1 Arguments from expert opinion

When modelling expert testimony, an important scheme is, of course, the scheme of arguments from expert opinion. This also holds for Bayesian analyses, since expert judgement is a recognised source of subjective probabilities. So there is every reason to discuss the expertise issue in detail. The scheme below is modelled after Walton et al (2008).

Argument scheme from expert opinion

$$\frac{\begin{array}{l} E \text{ is an expert in domain } D \\ E \text{ asserts that } P \\ P \text{ is within } D \end{array}}{P}$$

The double horizontal line indicates that the scheme is presumptive. Therefore, the scheme has **critical questions**:

1. How credible is E as an expert source?
2. Is E personally reliable as a source?
3. Is P consistent with what other experts assert?
4. Is E 's assertion of P based on evidence?

Question (1) is about the level of expertise while question (2) is about personal bias. With respect to question (3) an implicit condition of use becomes relevant, namely, that the expert opinion scheme can only be used by those who are not themselves experts in domain D , such as the judges in the cases. I could, of course, not defeat A 's arguments by saying that I am also an expert and I say $\neg P$.

In probability theory sometimes a sharp distinction is made between frequentist (objective) and epistemic (subjective) Bayesian probability theory. Probabilities based on frequencies as reported by statistics would be objectively justified, while probabilities reflecting a person's degrees of belief would be just subjective. However, selecting, interpreting and applying statistics involves judgement, which could be subjective. Moreover, a person's degrees of belief could be more than just subjective if they are about a subject matter in which s/he is an expert. The same actually holds for the judgements involved in applying frequency information and statistics: if these judgements are made by someone who is an expert in the problem at hand, these judgements may again be more than purely subjective. So the issue of expertise is crucial in both 'objective' (frequentist) and 'subjective' (epistemic) Bayes (likewise Biedermann et al (2017)).

It should be noted that in many cases, an expert asserts not a proposition but an argument. In this section I confine myself to assertions of statements; in Section 4.2 I will discuss how expert assertions of arguments can be modelled as sequences of assertions of statements.

4.1.1 The truth of the premises

Before the witness testimony scheme can be applied, first its premises have to be established. In the two cases, the question whether the first premise is true was very relevant. In this respect the cases highlight the importance of a distinction: P can be a specific statement made by the expert about a specific piece of evidence but it can also be a collection of similar statements or even the entire expert report. What A did was formulating hypotheses, making decisions about relevance of evidence to these hypotheses, about statistical independence between pieces of evidence given these hypotheses and, finally, about probability judgements. I claimed that all these decisions can only be reliably made by someone who is

an expert in the domains of the various aspects of the case at hand. In the Breda Six this concerned, among other things, the time of rigor mortis, reliability of statements by the suspects and witnesses (including hearsay evidence and anonymous witnesses), information concerning prior convictions and prior criminal investigations, evidence of various traces like DNA, blood stains and hairs, statistical evidence concerning confession rates among various ethnic groups and various common-sense issues, such as the relevance of the fact that two of the six suspects worked in a snack-bar next door to the crime scene. In the Oosterland case the main evidence concerned statements of the suspects and witnesses, statistics and other general knowledge about arson cases, information concerning prior convictions and prior criminal investigations and again various commonsense issues, such as how communities might turn against individuals and the relevance of friendships between suspects.

Let us now consider the case where D is the domain of Bayesian analysis of complex criminal cases, understood as comprising all the above issues. In my report, I formulated two general arguments against the truth of the first premise that A is an expert in this domain. First, expertise in the mathematics of Bayesian probability theory does not imply expertise in applying Bayes to a domain and, second, expertise in applying Bayesian probability theory in the domain of climate physics does not imply expertise in applying Bayes to the domain of complex criminal cases.

In his requisitory (pp. 184-5), the advocate general in the Breda Six case argued that A is an expert in Bayesian reasoning by mentioning that A had given tutorials about this topic to judges and prosecutors at the Dutch national training center for the judiciary (SSR) and that he had given various guest lectures on this topic at Dutch universities. In my report in the Oosterland case (one year after the Breda Six case) I anticipated similar arguments by stating that such activities do not make someone an expert but that rather one is invited to give tutorials and guest lectures because one is regarded to be an expert on the basis of other evidence, and this other evidence was, in my opinion, lacking.

The AG in the Breda Six case did not give arguments for A 's expertise on any of the evidence domains in the case, while yet he used A 's analyses of blood stain evidence, hair evidence and rigor mortis issues in quite some detail.

The court in the Breda Six case ruled that A could be regarded as an expert on the following grounds. First the court stated the relevant criteria: profession, education and experience of the claimed expert, the relevance of his expertise for the case, the nature of the method used by the claimed expert, whether this method is reliable and whether the claimed expert is able to apply this method skillfully. The court then referred to A 's education and PhD in physics and remarked that he had experience in the application of Bayesian thinking, "albeit in fields of science and research different than the law"². The court then mentioned the SSR tutorials also mentioned by the advocate general in his requisitory, and mentioned that A had advised the prosecution in one earlier case, "although not as an appointed expert". Finally, the court noted that A had in his report and in the court session described his method and his way of applying it in quite some detail, and had argued why he thought this method was reliable.

In my opinion, this argument can be criticised in many ways. First, it is remarkable that the court did not explicitly apply its own criterion whether A 's expertise was relevant to the case. The court did in its conclusion not even state in which field A could be regarded as an expert; it just stated that he could be regarded as an expert "for the present procedure". As noted above, I had in my report argued that expertise in Bayesian thinking and its application

² English translations from the case documents are my translations, HP.

in one domain does not imply expertise in the application of Bayesian thinking in another domain. Just like the advocate general, the court chose not to reply to this argument.

Second, as I noted above in my comments on the AG's requisitory, the relevance of A's SSR tutorials for his expertise can be questioned. In fact, the courts mentioning of A's work as an advisor for the prosecution in an earlier case is subject to the same criticism. Third, there is no evidence for the court's claim that A had experience in Bayesian analysis in his work as a climate physicist. In fact, some evidence suggests that the opposite may be the case. For my report in the Oosterland case I did a search with Google scholar and I could find no publications of A on climate physics.

Fourth, the court did not apply its own criteria concerning the method used by A. All the court did was mentioning that A described his method and way of using it and that A had argued why the method is reliable. From this nothing follows about whether the method is indeed reliable and whether A is indeed able to apply it. To be honest, the court did address the question whether A's method is reliable, but not as an aspect of the expertise question. I will discuss this part of the court's ruling below in Section 4.3.

In Figure 1 this analysis is visualised, with as top-level conclusion that A is an expert in Bayesian analysis of complex criminal cases (the first premise of the witness testimony scheme). In the figure, final conclusions of arguments are displayed in thick boxes. When the way in which several grounds for the same conclusion are combined is unclear, this is visualised with separate single arrows pointing to the same conclusion. Thus further interpretation about whether the premises are linked or accrue is left to the reader. Next, specific generalisations are left implicit and attacks on them are visualised as attacks on the inference. Note that in an implemented support system it might instead be useful to visualise generalisations as premises, to support arguments about whether they are justified.

In the Oosterland case, court did not discuss the issue of A's expertise but the issue was discussed by A in his written reply to my report. He admitted that he has no expertise in any of the relevant evidence domains of the case and he argued that the value of his report did not lie in providing reliable posterior probabilities but in showing which questions had to be answered by the court. To understand this, it is relevant that A's analyses in the two cases were applications of naive Bayes, with specifications of the prior odds concerning the hypotheses and the likelihood ratio's of each individual piece of evidence given these hypotheses. A presented this as a "spreadsheet" approach, suitable for showing to the court which probability judgements they had to make, after which all these probabilities could be multiplied according to Bayes' rule to provide the posterior odds. My first argument against this (to be discussed further below in Section 4.3) was that naive Bayes is too simplistic to be applicable to complex criminal cases, so that the spreadsheet metaphor is misleading. My second counterargument was that even identifying the right questions in a complex criminal requires expertise in the relevant evidence domains. In my report in the Oosterland case I backed this with a medical analogy, further discussed below in Section 4.4.

This completes the discussion of the first premise of the witness testimony scheme. The second premise was not really an issue in the case, while the third premise is irrelevant given the way the discussion about the first premise was summarised above.

4.1.2 *The critical questions*

Considering the critical questions of the scheme, personal bias (the second question) was not considered as an issue in the cases. The first question (how credible is E as an expert source) is in fact a weaker version of the question whether the first premise (is E an expert in domain D ?) is true: if the court in the Breda Six is followed in its decision that A can be regarded as

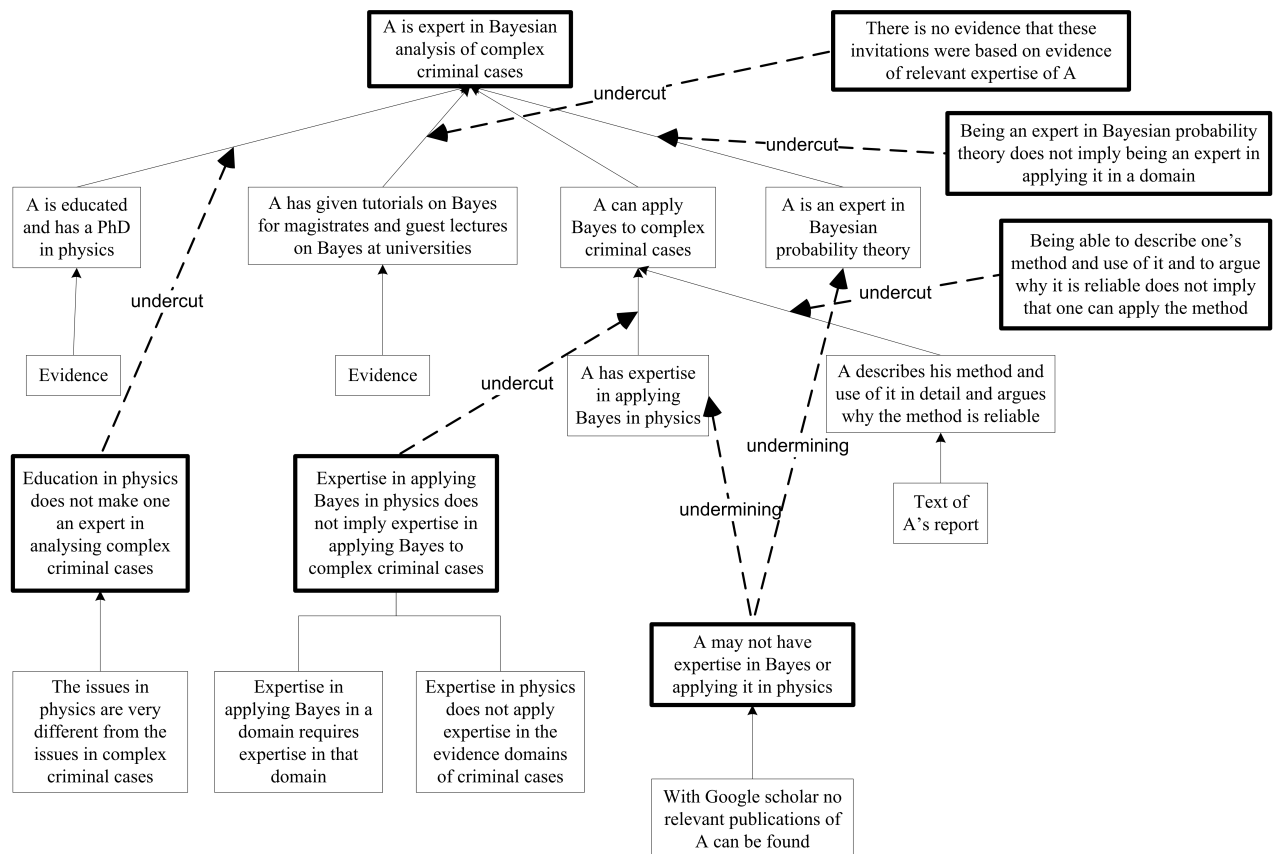


Fig. 1 Arguments about expertise.

an expert for the purpose of the case, then the arguments against this decision now become arguments that A's level of expertise is low. Such arguments are especially relevant when dealing with the third critical question (Is P consistent with what other experts assert?). In fact, A and I disagreed on a number of issues, so the court arguably had to assess the relative level of our respective expertise, and doing so is a kind of metalevel argumentation about the strength of arguments as formalised by e.g. Modgil and Prakken (2010). Finally, the fourth question (Is E 's assertion of P based on evidence?) was used by me in forming arguments that most of A's probability judgements were not based on any data or scientific knowledge and in an argument against A's specific assertion, not backed by any reference, that "internationally accepted estimates" yield a specific likelihood ratio of confessions.

4.1.3 Conclusions on expertise arguments

Concluding, Walton et al's (2008) argument scheme from expert opinion is a good overall framework for analysing the debates about expertise in the two cases. On the other hand, most interesting argumentation is not at the top level of this scheme but deeper down in the detailed arguments concerning the scheme's premises and critical questions. A support system should therefore not confine itself to giving support at the top level of an argument

scheme. If more specific knowledge is available (as, for instance, in the criteria for assessing expertise that a legal system may have), then this should preferably be incorporated. As for the arguments in Figure 1, some of them are quite specific to the case but others are more generic, such the arguments on whether expertise in Bayes implies expertise in applying Bayes and the arguments on whether expertise in applying Bayes in one domain implies expertise in applying it in another domain. In the present cases (with A being a climate physicist and having no relevant education, work experience or publications) the counter-arguments are quite strong but other cases could arise in which, for example, statisticians or philosophers of science give Bayesian analyses. In such cases Figure 1 points at relevant issues to be discussed.

Finally, it should be noted that the significance of this section is not confined to Bayesian expert analysis. Many observations from this sections also apply to expert analyses with other reasoning approaches, such as argumentation or story-based approaches. Here too, expertise in a mode of reasoning must be combined with expertise on the matters at hand.

4.2 Arguments from reasoning errors

In Section 4.1 I assumed that an expert asserts propositions but often an expert will assert an argument. Asserting an argument includes but goes beyond asserting its premises and conclusion: the expert also claims that the conclusion has to be accepted because of the premises. In many cases such an argument can be attacked by rebutting, undercutting or undermining it. However, sometimes a critic might want to say that the argument is inherently fallacious. This is not the same as stating an undercutting argument, since an undercutter merely claims that there is an exception to an otherwise acceptable inference rule. Although any expert may make reasoning errors, such errors will especially in the present domain, with many complex probabilistic arguments and sometimes complex statistical arguments, be frequent. So arguments from reasoning errors deserve to be studied in this paper.

In the two cases of the present case study, several arguments about argument validity were exchanged. An example from the Breda Six case is when I claimed that contrary to what A claimed, a probability concluded by A does not follow in probability theory from other probabilities assumed by A. After some discussion I had to admit that I was wrong and that A's argument was deductively valid.

In his report in the Oosterland case, A first estimated the probability of fifteen arson cases in a town like Oosterland in a six-months period given the hypothesis that they were not related as at most one in a million. He then concluded from this that the fifteen arson cases that he considered in his report cannot have been coincidence and that they must have been related. More formally, this argument can be rendered as 'The probability that the fifteen incidents happen given that they are unrelated is at most 1 in a million, therefore the probability that the fifteen incidents are unrelated given that they happen is very low'. From this he in turn concluded that serial arsonists must have been active, since no other relation would be realistically possible. In my report I claimed that this argument is an instance of the prosecutor fallacy, since it confuses the probability that the fifteen incidents happen given that they are not related with the probability that the fifteen incidents are not related given that they happen (this is sometimes also called the fallacy of the transposed conditional). A full argument would not just state but show that given the axioms of probability theory the second probability is not implied by the first. However, I left this part of my argument implicit since I assumed it to be generally known.

One way to show that A 's argument is fallacious is by giving a simple formal counterexample, for example, to specify for some E and H that $Pr(E | H) = Pr(E | \neg H) = 1/1.000.000$ so that the likelihood ratio of E with respect to H equals 1, so that the posterior probability $Pr(H | E)$ equals the prior probability $Pr(H)$, which can be any value. As a reply, it might be argued that in the present case $Pr(E | \neg H)$ (i.e. the probability of the 15 incidents given that they are related) is much higher than $1/1.000.000$ and that, moreover, the prior is such that an application of Bayes' rule would yield a very low posterior $Pr(H | E)$. However, this would not reinstate the fallacious argument but instead replace it with a valid argument.

From the point of view of argument visualisation one would like to have the following. For a given probabilistic statement ϕ , such as a link or probability in a Bayesian network, or a probability that is part of a likelihood ratio specified by an expert, the user could click on the statement and be able to inspect the following argument:

Expert E asserts that ψ_1, \dots, ψ_n
 Expert E asserts that ψ_1, \dots, ψ_n imply ϕ
 Therefore, ϕ because of ψ_1, \dots, ψ_n .

Discussions about reasoning errors by experts can then schematically be represented as follows with several applications of the witness testimony scheme combined with an inference from their conclusions:

E is an expert in domain D
E asserts that $\psi_i (1 \leq i \leq n)$
ψ_i is within domain D
ψ_i
...
E is an expert in domain D'
E asserts that ψ_1, \dots, ψ_n imply ϕ
That ψ_1, \dots, ψ_n imply ϕ is within domain D'
ψ_1, \dots, ψ_n imply ϕ

Then if the fact finder both accepts ψ_1, \dots, ψ_n and that ψ_1, \dots, ψ_n imply ϕ on the basis of the expert testimony, the fact finder should also accept ϕ .

This approach can be used to model any debate about reasoning errors by experts. I now illustrate it with a probabilistic example from the present case studies, namely, the dispute in the Oosterland case about my claim that A had committed the prosecutor fallacy. In terms of the just-sketched approach, this dispute was about the final application of this sequence of expert testimony schemes and can be modelled as follows.

A is an expert on arson cases
E asserts that $Pr(incidents \neg related) \leq 1/1.000.000$
E 's assertion is within the domain of arson cases
$Pr(incidents \neg related) \leq 1/1.000.000$
E is an expert in Bayesian reasoning
E asserts that P implies $Pr(related incidents) \gg 0.5$
E 's assertion is within the domain of Bayesian reasoning
P implies $Pr(related incidents) \gg 0.5$

Here P is the conclusion of the first argument and \gg means 'much greater than'. The conclusions of these two arguments deductively imply $Pr(related | incidents) \gg 0.5$.

My counterargument can be modelled as follows, where C stands for a description of the above-given counterexample:

$$\frac{C \text{ implies that } P \text{ does not imply } Pr(\text{related} \mid \text{incidents}) \gg 0.5}{C} \\ \hline P \text{ does not imply } Pr(\text{related} \mid \text{incidents}) \gg 0.5$$

In *ASPIC*⁺ and similar argumentation systems this argument defeats the preceding one, since it is a deductive argument with universally true premises while its target is defeasible.

4.3 Methodological issues

The reliability of the method used by A was an issue in both cases. One might expect that this issue will frequently arise in debates between experts, so a discussion is in order, even though the arguments in the two cases did not clearly instantiate recognised argument schemes. Neither courts defined what they meant with reliability. I defined it as the question whether different analysts applying the same method to the same problem will arrive at the same or at least similar result. In both my reports I argued that Bayesian analysis is not a reliable method in this sense, since in the academic literature there is no consensus on the right way of using Bayes for analysing complex criminal cases. In fact, there seems to be consensus on two things: that naive Bayes is too simplistic for complex criminal cases and that further research is needed before reliable methods can be offered to courts. Much current research concentrates on applying Bayesian networks (Fenton and Neil, 2011; Lagnado et al, 2013; Vlek et al, 2016; de Zoete et al, 2015) but the results are still preliminary while, moreover, it is almost exclusively concerned with the structure of Bayesian networks and leaves aside the question how reliable probabilities can be established. For these reasons, it seems quite likely that different analysts will produce quite different Bayesian analyses of the same case. This warrants the conclusion that A's method is therefore not reliable in the sense defined above. This conclusion generalises to any use of Bayesian probability theory to analyse complex criminal cases.

In neither of the two cases did A respond to this criticism. The court in the Six of Breda case agreed with my analysis and therefore disregarded the conclusions of A's report. The court in the Oosterland case did not comment on this issue except for summarising my argument. Since the court chose to disregard A's report "considering" my criticisms, this indicates that they may have agreed with my argument.

Like with the scheme from expert testimony, methodological issues can concern either the entire report or specific issues. An example of the latter was the debates in both cases between A and me about the appropriateness of the global independence assumption implied by A's adoption of naive Bayes. For example, in both cases A used quite specific pieces of evidence to assess the prior probabilities of his hypotheses. The axioms of probability theory then imply that in every likelihood ratio the evidence has to be conditioned not just on the hypotheses but also on the evidence used for assess their priors, unless the pieces of evidence can be regarded as statistically independent given these hypotheses. A did not condition the likelihood ratios in this way but neither provided arguments for the required independence assumptions. I criticised him on the grounds that according to the axioms of probability theory he should have done either one or the other of these things. This is general methodological criticism that translates into specific criticism of A's final argument that the posterior odds can be calculated by multiplying his prior odds and likelihood ratios. There are two ways to formally model this specific criticism. One is to interpret A's final argument as having as additional premises the required independence assumptions and then to observe that these additional assumptions are not backed by evidence or argument. The other way is

to leave these assumptions out of the interpreted argument and then to build an argument in the style of Section 4.2 that A's final argument is deductively invalid.

4.4 Analogical arguments

In the two case studies, several analogical arguments were used. The following version of the scheme for such arguments is fairly standard; cf. (Walton et al, 2008, pp. 58,315).

Argument scheme from analogy

Case C_1 and C_2 are similar in respects R_1, \dots, R_k

R_1, \dots, R_n are relevant similarities as regards P

P is true in case C_1

P is true in case C_2

Critical questions:

1. Do cases C_1 and C_2 also have relevant differences ?
2. Is Case C_2 relevantly similar to some other case C_3 in which P is false?

As is well-known, for special domains the scheme can be concretised, such as for case-based reasoning in the law (Bench-Capon, 2017). However, for present purposes the above version will do.

One use of analogy was in the Breda Six case, concerning the evidence that two of the three accused women worked in a snack-bar next door to the crime scene. In his report, A specified a likelihood ratio of the "coincidence" between this piece of evidence and other evidence that these two suspects knew three other suspects who by an anonymous informant of the criminal intelligence unit of the police (CID) had been mentioned as being involved in the crime. A first specified the denominator of this likelihood ratio (the probability of the coincidence given innocence of all six accused) as 1 in 500 to 1 in 1000 (on grounds that are irrelevant here). He then said "this agrees with a likelihood ratio of the coincidence of 500 to 1000". Taking this literally, this is a simple error, since before this conclusion can be drawn, first the probability of the coincidence given A's offender hypothesis has to be determined. However, other parts of A's report reveal that he set this probability to 1. Here he used an analogy with a hypothetical case in which a burglar breaks into a house by using a key of the house. Suppose a suspect is caught in possession of the key. According to A, possession of the key is a necessary element of the crime, so given that of the suspect committed the burglary, the probability that he possesses the key is 1. In the same way, A argued, the coincidence in the Breda Six case is a necessary element in the crime, since A's offender hypothesis was that at least some of the six accused were involved in the crime, *where one or more female accused lured the victim to the restaurant where the crime took place*. I criticised this on the grounds that, firstly, such luring can also be done by someone who does not work next door to the restaurant, such as the third female suspect; and, second, that the joint innocence of the two female suspects working next door to the restaurant is consistent with A's offender hypothesis. So the coincidence cannot be regarded as a necessary element of the crime. In pointing this out, I observed that this is a relevant difference with A's hypothetical burglary case, in which possession of the key *is* a necessary element of the crime. Thus I criticised A's analogy by using the first critical question of the analogy scheme.

Arguably another use of analogy in the Breda Six case is in the reasoning why A can be regarded as an expert in Bayesian analysis of complex criminal cases (Section 4.1 above). A reasoning step that is arguably implicit in the court's ruling on this issue is that experts in Bayesian analysis in physics are also experts in Bayesian analysis of complex criminal cases. This argument can be seen as an argument from analogy, referring to the supposed similarity between Bayesian analysis of problems in physics and of complex criminal cases. My argument that the one does not imply the other can be regarded as another application of the first critical question of the analogy scheme. One might expect that in debates about the use of Bayes by experts in court such analogical arguments referring to expertise in supposedly similar fields will arise more often.

In fact, the same holds for arguments used in combination with arguments from statistics, such as the arguments discussed above that statistics on arson in Japan and the UK also apply to the Netherlands. These arguments, too, can be seen as analogical arguments so they, too, can be criticised with the critical questions of the analogy scheme.

Finally, in my reply to A's reply in the Oosterland case I made use of a medical hypothetical to criticise A's claim that he is able to show the court which questions it has to consider. My hypothetical was part of a rhetorical question whether a medical specialist investigating a seriously ill patient would ask a climate physicist to tell him which medical investigations he had to perform.

4.5 Arguments from statistics

One might expect that in a probabilistic analysis of a complex criminal case, arguments from statistics to individual probability statements are frequent. Yet in my two cases most probability judgements were not based on statistics; in just a few cases A used them to support his judgements. In some other cases A used a quasi-frequentist approach. For example, in the Oosterland case he assessed the probability that the suspect and someone else (a suspect in a related case) were best friends given the innocence hypothesis by first observing that Oosterland has 2400 inhabitants and then supposing that for men like the suspect there were 200 candidates in Oosterland for being his best friend, thus arriving at a probability of 1 in 200 given innocence of both. This illustrates that even if probability judgements are based on data, the step from data to probabilities can involve subjective assumptions (in this case that there were 200 candidates for being the suspect's best friend).

In its most basic form, an argument from a statistical frequency to an individual probability takes the following form.

Argument scheme from statistical frequencies

The proportion of F 's that are G 's is n/m

a is an F

$Pr(Ga | Fa) \approx n/m$

It is important to note that this scheme is presumptive: there is no necessary relation between a frequency statement about a class and a conditional probability statement about a member of that class. So the scheme has critical questions other than whether the premises are true. Hacking (2001) calls the scheme the 'frequency principle'. He notes that it is justified on the assumption that nothing else relevant is known besides the frequency and that a is an F ; he also notes that a lot of judgement can go into the question which other information can be relevant.

Before considering the scheme's critical questions, let us look at how the first premise can be established. One way is by **statistical induction**:

The proportion of investigated F 's that are G 's is n/m

The proportion of F 's that are G 's is n/m

This scheme is not treated in the usual accounts of argument schemes, such as Walton et al (2008). A full investigation of ways to criticise uses of the scheme would lead us to the field of statistics, which is beyond the scope of this paper. For now it suffices to list two obvious **critical questions**: whether the sample of investigated F 's is biased and whether it is large enough.

In my cases, A derived some statistical information from sources. For example, in the Breda Six case he used statistics reported in a criminological publication on the frequencies of confessions of denials among various ethnic groups in the Netherlands. The reasoning then becomes:

E says that *S* is a relevant statistic, *E* is expert on this, therefore (presumably), *S* is a relevant statistic. Furthermore, *S* says that the proportion of investigated F 's that were G 's is n/m , therefore (presumably) the proportion of investigated F 's that were G 's is n/m .

The final conclusion then feeds into the scheme from statistical frequencies. In my report on the Breda Six case, I did not criticise A's specific selection of statistics on confessions and denials but I did note in general that selection of relevant and reliable statistics from the research literature requires expertise in the subject matter of that literature. I then observed that there was no evidence that A possessed criminological expertise of the relevant kind, thus in fact attacking the second premise of this line of reasoning. All this illustrates that even in reasoning from statistics the argument scheme from expert opinion is relevant.

I now turn to three possible **critical questions** of the scheme from statistical frequencies (there may be more).

1. *Is there conflicting frequency information about more specific classes?* This is the well-known issue of choosing the most specific reference class.
2. *Is there conflicting frequency information about overlapping classes?* This is a variant of the issue of choosing the most specific reference class. If a belongs to two non-overlapping but non-inclusive classes F and H , then in general the proportion of F -and- H 's that are G does not depend on the respective proportions of F 's and H 's that are G . So without further information nothing can be concluded on $Pr(Ga \mid Fa \wedge Ha)$.
3. *Are there other reasons not to apply the frequency?* For example, a might belong to some subclass for which commonsense or expert judgement yields different frequency assessments. For instance, in the Oosterland case, the probability assumed by A that the suspect and the other person were best friends given the innocence hypothesis ignored that both were outsiders in the community, that they had similar life styles and that one was previously convicted and the other was previously suspected of serial arson. Even if no statistics about these subclasses of adult male inhabitants of Oosterland exist, commonsense says that given these characteristics the probability of being best friends given innocence may be considerably higher than as assumed by A in his quasi-frequentist way.

Another scheme that was used by A in deriving probability judgements from statistics was the scheme from analogy (cf. Section 4.4). For example, in his report in the Oosterland case, he based his assessment of the probability of fifteen arson cases in a town like Oosterland in

a half-year period given that no serial arsonist was active in Oosterland in that period among other things on statistics on arson in Japan and the United Kingdom. Applying this statistic to The Netherlands assumes that Japan and the United Kingdom are relevantly similar to the Netherlands as regards (serial) arson. This seems a quite common way of using statistics for deriving probability judgements. Here again the expertise issue comes up, since judging whether two countries are relevantly similar as regards (serial) arson requires domain expertise relevant to that question. Here too my general criticism was that there was no evidence that A, being a climate physicist, possesses such relevant expertise.

Summarising, reasoning from statistics can be a combination of at least the following presumptive argument schemes: arguments from statistical frequencies, arguments from statistical induction, arguments from expert opinion and arguments from analogy. In addition, specific methodological concerns from the field of statistics can arise. Therefore, a full model of evaluating arguments from statistics cannot be developed without involving statisticians.

Finally, the analysis in this section is also relevant for argumentative and story-based reasoning approaches. In both these approaches qualitative defeasible generalisations are very important. For example, in argumentative approaches evidential generalisations connect evidence to conclusions, such as *witnesses usually speak the truth*. And in story-based approaches causal generalisations connect hypotheses to evidence, such as *extreme jealousy can cause feelings of murderous revenge*. A qualitative version of the statistical induction scheme could be used to argue for such generalisations. For example:

The great majority of investigated *F*'s are *G*'s
F's are usually *G*'s

Moreover, the critical questions of the scheme from statistical frequencies have their counterparts in critical examinations of arguments applying defeasible generalisations. The first two questions indicate the possibility of rebutting arguments (arguments for contradictory conclusions) on the basis of conflicting generalisations while the third question points at undercutting on uses of the qualitative variant of the statistical induction scheme.

5 Requirements for an argumentation support system

By way of summary I now list the requirements for a support system for argumentation about Bayesian analyses of criminal cases suggested by the present case study.

First, the system should provide support for use of the common argument schemes in this domain, including the schemes discussed in this paper. In addition, support should be provided for formulating and criticising arguments not conforming to such schemes. One way to provide such support is to utilise legal knowledge, regulations or policies about evaluating specific types of evidence, such as the criteria that several jurisdictions have for determining whether someone can be regarded as an expert witness. Parts of the present case study may also be reusable, such as the analysis of expertise arguments visualised in Figure 1. In addition, collaboration with statisticians may result in useful refinements of the critical questions of the scheme from statistics.

Capturing ownership of arguments and claims is important, especially with respect to evaluating the expertise of the owner or assessing the relative strength of conflicting arguments of different experts. For the latter, the system should also support metalevel argumentation about argument strength.

The system should also support arguments about the (deductive or defeasible) validity of other arguments. For this reason, the system should support explicit representation of the

logical nature of an argument and it should not prevent the formulation of arguments that are neither deductively nor defeasibly valid. Note that a system like Carneades (Gordon et al, 2007) which abstracts from the relation between premises and conclusion, does not fully satisfy these requirements.

The system should somehow make a natural distinction between discussions about general issues or expertise and methodology and discussions about expertise and method concerning specific expert assertions. On the one hand, it would be nice if a user of a tool for designing Bayesian networks could simply click on any element of a network to inspect the arguments exchanged about that element. However, this might not cover all relevant arguments, since as we have seen, many arguments are about more general issues.

Finally, there should be ways to express that no evidence for a claim or premise has been provided, since that is a usual way of criticising arguments in general and expert arguments in particular. If the software provides automatic means for evaluating a debate, then this issue should be taken into account, as is, for instance, done in Carneades.

6 Related research

One motivation underlying the present paper was the design of support software for modelling discussions about Bayesian analyses of complex criminal cases. In the medical domain, Yet et al (2016) have recently presented a similar system, which relates a medical BN to the clinical evidence on which it is based. Both supporting and conflicting evidence of a BN element can be represented in and shown by the system, as well as evidence related to excluded variables or relations. Three sources of evidence are modelled: publications, experts and data. Despite its argumentative flavour, the system is not based on an explicit argumentation model.

There is some earlier research on argumentation related to Bayesian analyses of criminal cases. Bex and Renooij (2016) provide a translation from *ASPIC*⁺-style arguments to constraints on Bayesian networks (BN). Their focus is different from the present paper in that their arguments are not about how to justify elements of BN. Instead, the translation method aims to translate the information expressed in an argument in the BN. For example, one constraint says that propositions in arguments should have corresponding nodes in the BN and another constraint says that inferences in an argument and attacks between arguments should have a corresponding ‘active’ chain of links in the BN.

Timmer et al (2017) does the opposite as Bex and Renooij (2016), namely, translating the information contained in a BN into an *ASPIC*⁺ argumentation framework, in order to explain the BN with arguments.

The closest to the present paper is Keppens (2014), who proposes a set of source-based argument schemes for modelling the provenance of probability judgements in likelihood approaches. Among other things, Keppens proposes schemes for expert opinion (a special case of the one in the present paper), for reasoning from data sets (not unlike the present scheme for reasoning from statistics) and for reasoning from generally accepted theories. In addition, Keppens proposes a set of schemes for relating source-based claims concerning the nature of subjective probability distributions (such as ‘B has a [non-negative/non-positive] effect on the likelihood of C’) to formal constraints on the probability distributions. Yet there is a difference in approach. Keppens primarily aims to build a formal and computational model, while this paper primarily aims to analyse how discussions about Bayesian analyses actually take place. Thus the present study complements Keppens’ research. Also, the focus

of Keppens' model is more limited than the present study in that it only models arguments about specific probability distributions.

7 Conclusion

In this paper a new use case for legal argumentation support tools has been considered: supporting discussions about Bayesian analyses of complex criminal cases. By way of a case study, two actual discussions between experts in court cases have been analysed on their argumentation structure. Since this is a case study, the question arises how general the results are. As stated in the introduction, it is hard to say to which extent the studied cases are typical, since Bayesian analyses of entire complex criminal cases are still rare in the courtroom. The usual uses of Bayes in the courtroom concern individual pieces of evidence, especially random match probabilities of forensic trace evidence (DNA, tyre marks, shoe prints, finger prints, glass pieces). Also, since I was involved in the debate about the Bayesian analysis, my analysis in the present paper may have been affected by a personal view.

Nevertheless, with this in mind, the case study still warrants some preliminary conclusions. From a theoretical point of view the richness of argumentation about Bayesian analyses and the usefulness of several recognised argument schemes have been confirmed, a new argument scheme for arguments from statistics has been formulated, and a novel analysis of some subtleties concerning arguments from expert opinion has been given. In particular, the rich variety of argument schemes that turned out to be applicable in our case studies suggests that rational discussions about Bayesian analyses of complex criminal cases need not conform to strict statistical or forensic-science methodology but may employ many techniques from argumentation theory. From a practical point of view the case study has resulted in requirements for support software for discussions about Bayesian analyses of complex criminal cases. The actual design and usefulness of such software are issues for future research, as well as the refinement of critical questions of some argument schemes, in particular the schemes for arguments from expert opinion and from statistics. Moreover, some contributions of this paper may also be relevant for other domains in which Bayesian analysis is used, such as medicine.

Finally, at several places in this paper we saw that the significance of the present analysis is not confined to Bayesian approaches but extends to discussions about expert reports using any reasoning method. To briefly summarise, expertise in any mode of reasoning must be combined with expertise on the matters at hand (Section 4.1) and arguments from statistics can in qualitative forms be found in arguments about defeasible generalisations in argumentative or scenario-based analysis (Section 4.5). And presumably, analogical arguments can be made about any kind of analysis (Section 4.4).

References

- Bench-Capon T (2017) HYPO's legacy: introduction to the virtual special issue. *Artificial Intelligence and Law* 25:205–250
- Bex F, Renooij S (2016) From arguments to constraints on a Bayesian network. In: Baroni P, Gordon T, Scheffler T, Stede M (eds) *Computational Models of Argument*. Proceedings of COMMA 2016, IOS Press, Amsterdam etc, pp 96–106
- Bex F, Modgil S, Prakken H, Reed C (2013) On logical specifications of the argument interchange format. *Journal of Logic and Computation* 23:951–989

- Biedermann A, Bozza S, Taroni F, Aitken C (2017) The meaning of justified subjectivism and its role in the reconciliation of recent disagreements over forensic probabilism. *Science and Justice* 57:477–483
- Van den Braak S (2010) Sensemaking Software for Crime Analysis. Doctoral dissertation Department of Information and Computing Sciences, Utrecht University
- Dung P (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n -person games. *Artificial Intelligence* 77:321–357
- Fenton N, Berger D (2016) Bayes and the law. *Annual Review of Statistics and Its Application* 3:51–77
- Fenton N, Neil M (2011) Avoiding legal fallacies in practice using Bayesian networks. *Australian Journal of Legal Philosophy* 36:114–151
- Fenton N, Neil M (2013) *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, Boca Raton, FL
- Gordon T, Prakken H, Walton D (2007) The Carneades model of argument and burden of proof. *Artificial Intelligence* 171:875–896
- Governatori G, Maher M, Antoniou G, Billington D (2004) Argumentation semantics for defeasible logic. *Journal of Logic and Computation* 14:675–702
- Hacking I (2001) *An Introduction to Probability and Inductive Logic*. Cambridge University Press, Cambridge
- Kapteijn H, Prakken H, Verheij B (eds) (2009) *Legal Evidence and Proof: Statistics, Stories, Logic*. Ashgate Publishing, Farnham
- Keppens J (2014) On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning. *Artificial Intelligence and Law* 22:239–290
- Lagnado D, Fenton N, Neil M (2013) Legal idioms: a framework for evidential reasoning. *Argument and Computation* 4:46–63
- Modgil S, Prakken H (2010) Reasoning about preferences in structured extended argumentation frameworks. In: Baroni P, Cerutti F, Giacomini M, Simari G (eds) *Computational Models of Argument*. Proceedings of COMMA 2010, IOS Press, Amsterdam etc, pp 347–358
- Modgil S, Prakken H (2014) The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation* 5:31–62
- Pardo M, Allen R (2008) Juridical proof and the best explanation. *Law and Philosophy* 27:223–268
- Prakken H (2017) Argument schemes for discussing Bayesian modellings of complex criminal cases. In: Wyner A (ed) *Legal Knowledge and Information Systems. JURIX 2017: The Thirtieth Annual Conference*, IOS Press, Amsterdam etc., in press
- Scheuer O, Loll F, Pinkwart N, McLaren B (2010) Computer-supported argumentation: A review of the state-of-the-art. *International Journal of Computer-Supported Collaborative Learning* 5:43–102
- Timmer S, Meyer JJ, Prakken H, Renooij S, Verheij B (2017) A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning* 80:475–494
- Verheij B, Bex F, Timmer S, Vlek C, Meyer JJ, Renooij S, Prakken H (2016) Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk* 15:35–70
- Vlek C, Prakken H, Renooij S, Verheij B (2016) A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law* 24:285–324
- Walton D, Reed C, Macagno F (2008) *Argumentation Schemes*. Cambridge University Press, Cambridge

-
- Wardeh M, Wyner A, Atkinson K, Bench-Capon T (2013) Argumentation based tools for policy-making. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, ACM Press, New York, pp 249–250
- Yet B, Perkins Z, Tai N, Marsh W (2016) Clinical evidence framework for Bayesian networks. *Knowledge and Information Systems* 50:117–143
- de Zoete J, Sjerps M, Lagnado D, Fenton N (2015) Modelling crime linkage with Bayesian networks. *Science and Justice* 55:209–217