Argument schemes for discussing Bayesian modellings of complex criminal cases

Henry PRAKKEN [1]
Utrecht University and University of Groningen

Abstract. In this paper two discussions between experts about Bayesian modellings of complex criminal cases are analysed on their argumentation structure. The usefulness of several recognised argument schemes is confirmed, two new schemes for interpretation arguments and for arguments from statistics are proposed, and an analysis is given of debates about the validity of arguments. From a practical point of view the case study yields insights into the design of support software for discussions about Bayesian modellings of complex criminal cases.

Keywords. Argument schemes, reasoning about evidence, Bayesian probability theory, argumentation support

1. Introduction

There is an ongoing debate on what is the best model of rational evidential reasoning in criminal cases. Both argumentation-based, story-based and Bayesian approaches have been proposed [3]. In this paper I remain neutral with respect to this debate. Instead I will argue that even if a Bayesian approach is adopted, there is still one clearly argumentative aspect of this form of reasoning, namely, debates about the merits of a proposed Bayesian model. 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 the experts' analyses, and argumentation support technology may be of use here.

To obtain insight in the requirements for argumentation-based add-ons to Bayesian-network software tools, this paper examines two recent Dutch criminal cases in which I was appointed by courts to comment on a Bayesian analysis of the entire case proposed by an expert of the prosecution. 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.

[1] Department of Information and Computing Sciences, Utrecht University, and Faculty of Law, University of Groningen, The Netherlands; E-mail: H.Prakken@uu.nl.

2. The cases

In the *Breda Six* case three young men and three young women were accused of killing a woman in the restaurant of her son after closing time, in 1993. The six were initially convicted in two instances, mainly on the basis of confessions of the three female suspects. In 2012 the Dutch Supreme Court reopened the case because of doubts about the truthfulness of these confessions. After a new police investigation the six were tried again by the court of appeal of The Hague and in 2015 they were again all found guilty, mainly on the ground that new evidence had confirmed the reliability of the confessions.

The prosecution in the case brought in an 80 page expert report by the climate physicist Dr. Alkemade (henceforth 'A') containing a Bayesian analysis of the entire case. A claimed that he could give a Bayesian analysis of the case since he had experience with using Bayesian probability theory in his work as a climate 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 at least 99,7%. The investigating judge in the case asked me to assess and evaluate A's report, which I did in a 41 page report. My main conclusion was that A's claims had no objective basis. 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, for which reason A's conclusions had to be disregarded.

In the *Oosterland* case a person was accused of being responsible for 16 small arson cases in the small town of Oosterland in a six-month period in 2013. Initially the suspect was acquitted, mainly on the grounds that the two main witness testimonies were unreliable. In the appeal case the prosecution again brought in a report by A, this time 79 pages long. A concluded that on the basis of the evidence considered by him the probability that the suspect was involved in at least a substantial number of the arson cases was at least 99,8%. The investigating judge in the appeal case asked me to asses the reliability of A's method and its application to the case. I delivered a 42 page report with essentially the same conclusions as in the Breda Six case. A then wrote a 47 page reply to my report, after which I wrote a 9 page reply to his reply. In 2016 the court of appeal convicted the suspect of 7 arson cases and acquitted him of the remaining 9 cases. The court stated that it had chosen to disregard A's report "considering" my criticism.

3. Theoretical background

Probability theory [2] 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 $Pr(A \mid B)$ stands for the conditional probability of A given B. In criminal cases we are interested in the conditional probability $Pr(H \mid E)$ 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 $Pr(A \mid C)$ and $Pr(\neg A \mid C)$ 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 $Pr(E_1 \mid H)$, i.e., if $Pr(E_1 \mid H \wedge E_2) = Pr(E_1 \mid H)$. The axioms of probability imply that such independence is symmetric. The axioms also imply the following theorems (here given in odds form). Let $E_1, \ldots, E_n$ be pieces of evidence and H a hypothesis. Then:

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

$$\ldots \times \frac{Pr(E_2 \mid H \wedge E_1)}{Pr(E_2 \mid \neg H \wedge E_1)} \times \frac{Pr(E_1 \mid H)}{Pr(E_1 \mid \neg H)} \times \frac{Pr(H)}{Pr(\neg H)}$$

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 $E_1, \ldots, E_n$ are statistically independent from each other given H, then the chain rule reduces to

$$\frac{Pr(H \mid E_1 \wedge \ldots \wedge E_n)}{Pr(\neg H \mid E_1 \wedge \ldots \wedge E_n)} = \frac{Pr(E_n \mid H)}{Pr(E_n \mid \neg H)} \times \ldots \times \frac{Pr(E_1 \mid H)}{Pr(E_1 \mid \neg H)} \times \frac{Pr(H)}{Pr(\neg H)}$$

which is *Bayes' theorem* (in odds form). 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 estimated 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_1 makes H more (less) probable.

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*. The more general chain rule is often also practically infeasible, because of the many combinations of pieces of evidence that have to be considered. As a solution, *Bayesian networks* have been proposed, which graphically display possible independencies with directed links between nodes representing probabilistic variables. For each value of each node, all that needs to be estimated 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.

Argumentation is the process of evaluating claims by providing and critically examining grounds for or against the claim. *Argument schemes* [6] 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. In formal approaches to argumentation, such as *ASPIC*$^+$ [5], argument schemes are often formalised as (deductive or defeasible) inference rules and critical questions as pointers to counterarguments. In the present paper argument schemes and their critical questions will be semiformally displayed, where critical questions asking whether the premises of the scheme are true will be left implicit.

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.

4.1. Text interpretation arguments

Some arguments are interpretation arguments, since they interpret the natural-language text of an expert report. In [6] two schemes for arguments from vagueness, respectively, arbitrariness of verbal classification are given, meant for criticising vagueness or arbitrariness in an argument. In the present case studies no such criticism was expressed but nevertheless issues arose concerning the correct interpretation of fragments of the reports. This gives rise to a new scheme of **Arguments from text interpretation**:

$$\frac{E \text{ says “}P\text{”} \qquad P \text{ means } Q}{E \text{ asserts that } Q}$$

This argument seems deductively valid (indicated by the single horizontal line) so it can only be criticised on its premises. Usually only the second premise will be controversial. In my reports I used this scheme several times as an introduction to an argument against *Q*. In one case, A convinced me in a private conversation afterwards that he had meant something else, after which I retracted my argument against *Q*.

4.2. Arguments from expert opinion

An obviously relevant scheme for modelling expert testimony is **arguments from expert opinion**. This especially holds for Bayesian modellings, since expert judgement is a recognised source of subjective probabilities. The following version of the scheme is modelled after [6].

E is an expert in domain *D*
E asserts that *P*
P is within *D*
══════════════════
P

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

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, his sharp distinction breaks down from both sides. To start with, 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 holds for the judgements involved in applying frequency information and statistics: if 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.

In the two cases, the question whether the scheme's 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 estimates. 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, 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, 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. The court in the Breda Six case instead ruled that A could be regarded as an expert for the purpose of the case. For space limitations an analysis of the court's justification of this decision has to be omitted. In the Oosterland case, the court did not discuss the issue of A's expertise but A himself discussed it 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 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. Against this I argued that even identifying the right questions in a complex criminal requires expertise in the relevant evidence domains.

Considering the critical questions of the scheme, personal bias (the second question) was not an issue. 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 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. 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 estimates were not based on any data or scientific knowledge.

Concluding, [6]'s 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.

4.3. Arguments from reasoning errors

In Section 4.2 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. Especially in probabilistic and statistical reasoning real or claimed reasoning fallacies can be frequent, so arguments from reasoning errors deserve to be studied.

In the two cases of the present case study, several arguments about argument validity were exchanged. For reasons of space I can discuss just one example. In his report in the Oosterland case, A first estimated that the probability of fifteen arson cases in a town like Oosterland in a six-months period given the hypothesis that they were not related is at most one in a million. He then concluded from this that the fifteen arson cases considered by him cannot have been coincidence and that they must have been related. 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.

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 $Pr(E \mid H) = Pr(E \mid \neg H) = 1/1.000.000$ so that the likelihood ratio of E with respect to H equals 1, so that the posterior probability $Pr(H \mid E)$ equals the prior probability $Pr(H)$, which can be any value.

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 estimated by an expert, the user could click on the statement and be able to inspect the following argument:

Expert E asserts that $\psi_1, \ldots, \psi_n$
Expert E asserts that $\psi_1, \ldots, \psi_n$ imply φ
Therefore, φ because of $\psi_1, \ldots, \psi_n$.

Our example can be modelled with a combination of two applications of the expert testimony scheme combined with a deductive inference from their conclusions:

$$\frac{\begin{array}{l} A \text{ is an expert on arson cases} \\ E \text{ asserts that } Pr(\mathit{incidents} \mid \neg \mathit{related}) \leq 1/1.000.000 \\ E\text{'s assertion is within the domain of arson cases} \end{array}}{Pr(\mathit{incidents} \mid \neg \mathit{related}) \leq 1/1.000.000}$$

$$\frac{\begin{array}{l} E \text{ is an expert in Bayesian reasoning} \\ E \text{ asserts that } P \text{ implies } Pr(\mathit{related} \mid \mathit{incidents}) >> 0.5 \\ E\text{'s assertion is within the domain of Bayesian reasoning} \end{array}}{P \text{ implies } Pr(\mathit{related} \mid \mathit{incidents}) >> 0.5}$$

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

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

$$\frac{\begin{array}{l} C \text{ implies that } P \text{ does not imply } Pr(related \mid incidents) >> 0.5 \\ C \end{array}}{P \text{ does not imply } Pr(related \mid incidents) >> 0.5}$$

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

4.4. Analogical arguments

In the two case studies, several analogical arguments were used. The following version of the **argument scheme from analogy** is fairly standard; cf. [6, pp. 58,315].

$$\frac{\begin{array}{l} \text{Case } C_1 \text{ and } C_2 \text{ are similar in respects } R_1,\ldots,R_k \\ R_1,\ldots,R_n \text{ are relevant similarities as regards } P \\ P \text{ is true in case } C_1 \end{array}}{P \text{ is true in case } C_2}$$

Its two **critical questions** are: (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?

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 estimated the likelihood ratio of this "coincidence". A first estimated the denominator of this likelihood ratio (the probability of the coincidence given innocence of all six accused) as 1 in 500 (on grounds that are irrelevant here). He then estimated the numerator of this likelihood ratio (the probability of the coincidence given his guilt hypothesis) as 1, thus arriving at a strongly incriminating likelihood ratio of 500. 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 guilt of the suspect 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 guilt 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 guilt hypothesis. So the coincidence cannot be regarded as a necessary element of the crime. I thus pointed at a relevant difference with A's hypothetical burglary case, in which possession of the key *is* a necessary element of the crime, thus using the first critical question of the analogy scheme.

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 estimates were not based on statistics; in just a few cases A used them to support his estimates. In some other cases A used a quasi-frequentist approach. For example, in the Oosterland case he estimated 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 estimating 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 estimates 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, **arguments from statistical frequencies** to an individual probability take the following form.

$$\begin{array}{l} \text{The proportion of } F\text{'s that are } G\text{'s is } n/m \\ a \text{ is an } F \\ \hline\hline Pr(Ga \mid Fa) \approx n/m \end{array}$$

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

$$\begin{array}{l} \text{The proportion of investigated } F\text{'s that are } G\text{'s is } n/m \\ \hline\hline \text{The proportion of } F\text{'s that are } G\text{'s is } n/m \end{array}$$

This scheme is not treated in the usual accounts of argument schemes, such as [6]. A full investigation of ways to criticise its use 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 requires expertise in the subject matter at hand. I then observed that there was no evidence that A possessed relevant criminological expertise, 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 first question. 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 estimates. For instance, in the Oosterland case, the probability estimated 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 estimated by A in his quasi-frequentist way.

Another scheme used by A in deriving probability estimates from statistics was the scheme from analogy. For example, in his report in the Oosterland case, A based his estimates 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 estimates. 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.

In sum, 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.

5. Related research

One motivation underlying this paper is the design of support software for discussions about Bayesian analyses of complex criminal cases. In the medical domain, [7] present 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 modellings of criminal cases. [1] 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 but on incorporating the information expressed in an argument in the BN.

The closest to the present paper is [4], who proposes a set of source-based argument schemes for modelling the provenance of probability estimates 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 modellings 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.

6. Conclusion

In this paper two discussions about Bayesian modellings of complex criminal cases were analysed on their argumentation structure. Since this is a case study, the question arises how general the results are. 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 two studied discussions, 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 modellings and the usefulness of several recognised argument schemes have been confirmed, two new argument schemes for interpretation arguments and arguments from statistics have been formulated, and a novel analysis of some subtleties concerning arguments from expert opinion has been given. From a practical point of view, the paper has identified a new use case for argumentation support tools, namely, support for argumentation about Bayesian probabilistic modellings of legal evidential reasoning.

References

[1] F.J. Bex and S. Renooij. From arguments to constraints on a Bayesian network. In P. Baroni, T.F. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument. Proceedings of COMMA 2016*, pages 96–106. IOS Press, Amsterdam etc, 2016.

[2] I. Hacking. *An Introduction to Probability and Inductive Logic*. Cambridge University Press, Cambridge, 2001.

[3] H. Kaptein, H. Prakken, and B. Verheij, editors. *Legal Evidence and Proof: Statistics, Stories, Logic*. Ashgate Publishing, Farnham, 2009.

[4] J. Keppens. On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning. *Artificial Intelligence and Law*, 22:239–290, 2014.

[5] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.

[6] D.N. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.

[7] B. Yet, Z.B. Perkins, N.R.M. Tai, and W.R. Marsh. Clinical evidence framework for Bayesian networks. *Knowledge and Information Systems*, 50:117–143, 2016.