An Argumentation-Based Analysis of the Simonshaven Case

Prakken, Hendrik

Published in:
Topics in Cognitive Science

DOI:
10.1111/tops.12418

TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 12 (2020) 1068–1091

ISSN: 1756-8765 online
DOI: 10.1111/tops.12418

This article is part of the topic "Models of Rational Proof in Criminal Law," Henry Prakken, Floris Bex and Anne Ruth Mackor (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

An Argumentation-Based Analysis of the Simonshaven Case

Henry Prakken[a,b]

[a] *Department of Information and Computing Sciences, Utrecht University*
[b] *Faculty of Law, University of Groningen*

Received 29 August 2018; received in revised form 28 January 2019; accepted 11 February 2019

Abstract

In an argumentation approach, legal evidential reasoning is modeled as the construction and attack of "trees of inference" from evidence to conclusions by applying generalizations to evidence or intermediate conclusions. In this paper, an argumentation-based analysis of the Simonshaven case is given in terms of a logical formalism for argumentation. The formalism combines abstract argumentation frameworks with accounts of the structure of arguments, of the ways they can be attacked and of ways to evaluate conflicting arguments. The purpose of this paper is not to demonstrate or argue that the argumentation approach to modeling legal evidential reasoning is feasible or even preferable but to have a fully worked-out example that can be used in the comparison with alternative Bayesian or scenario-based analyses.

Keywords: Evidential reasoning; Legal proof; Formal argumentation; Generalizations; Attack and defeat

Correspondence should be sent to Henry Prakken, Department of Information and Computing Sciences, Utrecht University, P.O. Box 80089, 3508 TB Utrecht, The Netherlands. E-mail: h.prakken@uu.nl

1. Introduction

In this paper, an argumentation-based analysis of the Simonshaven case will be given. Argumentation-based approaches are arguably close to how judges and lawyers think, so expressing principles of rational evidential reasoning in terms of argumentation may help in making these principles understandable for them. Having said so, the purpose of this paper is not to demonstrate or argue that the argumentation approach to modeling legal evidential reasoning is feasible or even preferable but to have a fully worked-out example that can be used in the comparison with alternative analyses. Argumentation is a broad term which can be made precise in various ways. The argumentation-theorist Douglas Walton defines the term "argument" as "the giving of reasons to support or criticize a claim that is questionable, or open to doubt" (Walton, 2006, p. 1). This nicely captures the dual constructive-critical nature of argumentation but says nothing about the relation between support and claim. In some sense, a Bayesian network can be regarded as support for a claim, for instance, for the claim that the suspect is guilty as charged. However, such a broad interpretation would deprive the term "argumentation" from most of its distinctive force. In this paper, the term is defined much narrower as the construction and attack of "trees of inference" from evidence to conclusions by applying generalizations to evidence or intermediate conclusions.

The argumentation approach to legal evidential reasoning ultimately goes back to Wigmore's charting method of legal evidence, with which alternative arguments from evidence to hypotheses can be graphically displayed and sources of doubt in these arguments can be revealed (Wigmore, 1931). Generalizations are important here as the "glue" that connects the evidence with the hypotheses and as important sources of doubt, both concerning their general plausibility and their possible exceptions. Wigmore's charting method has been extended and refined by the "New Evidence Scholars" (Anderson & Twining, 1991). In legal applications of artificial intelligence (AI), the argument-based approach has been developed in terms of logical frameworks for argumentation, adapted from general AI research on argumentation, which model reasoning as the construction and comparison of arguments and counterarguments, where the conclusions of the arguments that survive the competition with their counterarguments must be adopted as true (Bex, Prakken, Reed, & Walton, 2003; Verheij, 2003). Such adoption is not infallible, since new evidence or additional generalizations might give rise to new arguments that defeat the currently undefeated arguments.

This paper is organized as follows. In Section 2, the formal argumentation model assumed in this paper is summarized and briefly contrasted with scenario-based and Bayesian approaches. In Section 3, this paper's specific analysis method is explained, after which the analysis is introduced in Section 4, graphically displayed in Section 5, continued in Section 6, and discussed in Section 7.

2. Preliminaries

The argumentation model assumed in this paper is based on the standard AI theory of so-called abstract argumentation frameworks (Dung, 1995) as instantiated by the *ASPIC+*

framework (Modgil, 2014; Modgil & Prakken, 2013). In this section, an informal but still reasonably precise summary of these formalisms will be given as far as needed for present purposes. Readers who wish to see examples can look ahead to the analysis of the Simonshaven case.

2.1. Sketch of application of the ASPIC+ *framework for structured argumentation*

According to the instance of *ASPIC+* used in this paper, each step in an argument is of the form

$$P_1, \ldots, P_n, \text{therefore (qualifier) } Q$$

where

- Each *Pi* and *Q* is a statement in some language.
- The inference step is supposed to be based on a generalization *if P*1,…, *Pn then (qualifier) Q,* where the qualifier is an expression of strength of the generalization. A generalization is called *defeasible* if the qualifier is something like "normally," "typically," "usually," "most of the time," and *nondefeasible* if the qualifier is something like "always" or "certainly." A generalization can but need not be a logical inference rule. In most cases it will be a domain-specific generalization, like "people who flee from the crime scene when the police arrives are usually involved in the crime" or presumptive argument schemes (like "expert witnesses usually speak the truth").
- The qualifier of the inference step is "presumably" if the underlying generalization is defeasible and "deductively" otherwise. If the premises of the inference step are accepted, then if the inference is deductive, its conclusion must be accepted no matter what, while if it is defeasible, its conclusion must be accepted unless there are good reasons not to accept it (expressed in a convincing counterargument).
- Each *Pi* is either given (in AI terms "an element of the knowledge base") or a conclusion from a preceding inference in the argument.
- *Q* is the conclusion of the inference step.

Arguments are then trees of such inference steps,[1] where the leaves are its *premises* (pieces of evidence and/or other assumed statements), the links are inference steps, the root is the *final conclusion*, and all other nodes are *intermediate conclusions*. Subtrees of an argument are *subarguments* of the argument.

Arguments can be *attacked* in three ways: on their premises (*undermining attack*, where a subset of the premises, for example, the evidence, can be declared beyond attack), on their defeasible inference steps (*undercutting attack*), and on their defeasibly derived conclusions (*rebutting attack*). While a rebutting attack contradicts a conclusion ("no, this is not true but false"), an undercutting attack only states reasons why a conclusion cannot be drawn from the premises. Such a reason can be that there is an exception to the underlying generalization ("this expert is biased since some of his research is paid by the defendant") or that the generalization itself is unacceptable ("it is not the case that people who flee from the crime scene when the police arrives are usually involved in the crime"). Like all attacks, undercutting attacks can be counterattacked, so that arguments about generalizations can be modeled.

The *ASPIC*+ distinguishes between attack and defeat relations between arguments. Undercutting attacks always succeed as *defeats* while rebutting attacks can be resolved into defeats by applying some measure of relative strength or preference. If *A* (directly) rebuts or undermines *B,* then *A* defeats *B* if *B* is not preferred over *A*. This leaves open the possibility that two attacking arguments defeat each other, namely, if they are equally preferred or if their preference relation is not defined. If *A* defeats *B* but *B* does not defeat *A*, we say that *A strictly defeats B*. Finally, if *A* defeats a subargument of *B,* then *A* also (indirectly) defeats *B*.

The preference relation on arguments may in general be based on any ground. In the context of evidential reasoning, it usually reflects some notion of probability, but as yet there is no systematic theory of probabilistic argument strength. For this reason, the present criteria for argument strength as reflected in the preference relation on arguments will be somewhat *ad hoc* and largely based on intuitions. Having said so, it should be noted that there is current work in AI on integrating argumentation and Bayesian approaches, such as Timmer, Meyer, Prakken, Renooij, and Verheij (2017); Prakken (2018) and the work discussed therein. In my 2018 paper, I define notions of internal and external probabilistic strength of arguments on the basis of probabilistic strengths of generalizations interpreted as conditional probabilities. In future work it would be interesting to see how these ideas can be applied in the present context.

2.2. The theory of abstract argumentation frameworks

That an argument *A* strictly defeats an argument *B* does not yet settle the issue, since *A* may in turn be (strictly or weakly) defeated by an argument *C,* and so on. In some cases the outcome can still be easily determined. For example, a common pattern in judicial decisions, depicted in Fig. 1a, is that *A* is strictly defeated by *B*, which is strictly defeated by *C* and there are no other defeat relations, where *A* is the court's argument for its main decision (e.g., the suspect is guilty of manslaughter), *B* is a counterargument by the defense (for instance, based on the claim that the suspect killed in self-defense), and *C* is the court's refutation of the defense's counterargument (e.g., based on the judgment that the witnesses on which the claim of self-attack are unreliable). In this case it is clear that *C* defends *A* against *B* so both *A* and *C* can be accepted while *B* must be rejected. However, we can easily imagine more complex examples where our intuitions fall short. For instance, another argument *D* could be constructed such that *C* and *D* defeat each other, then an argument *E* could be constructed that defeats *D* but is defeated by *A*, and so on: which arguments can now be accepted and which should be rejected? Here, we cannot rely on intuitions but need a calculus. Its input is a set of arguments with a binary defeat relation, while its output is an assessment of the dialectical status of these arguments in terms of three classes (three and not two since some conflicts cannot be resolved). Intuitively, the *justified arguments* are those that survive all conflicts with their defeaters and so can be accepted, the *overruled arguments* are those that are defeated by

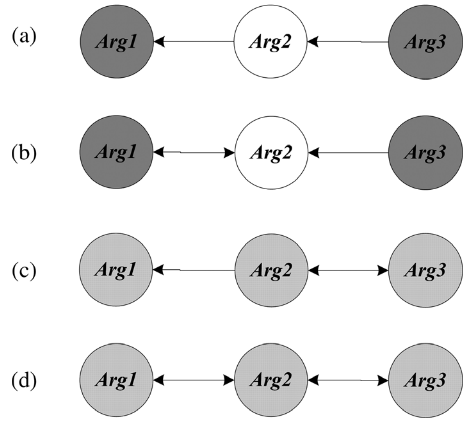

Fig. 1. Defeat patterns with three arguments.

a justified argument and so must be rejected, and the *defensible arguments* are those that are involved in conflicts that cannot be resolved. Furthermore, a *statement* is justified if it has a justified argument, it is overruled if all arguments for it are overruled, and it is defensible if it has a defensible argument but no justified arguments. In terms more familiar to lawyers, if a claim is justified, then a rational adjudicator is convinced that the claim is true; if it is overruled, such an adjudicator is convinced that the claim is false; while if it is defensible, s/he is neither convinced that it is true nor that it is false.

In AI, the standard calculus for determining the dialectical status of arguments is the theory of so-called abstract argumentation frameworks, going back to the seminal work of Dung (1995). This theory says nothing about the structure of arguments or the nature of defeat. In the present paper, arguments will take the form of inference graphs, but they could also, for instance, be a Bayesian network supporting a claim that the suspect is guilty; and a defeating argument could, for instance, be "the BN is of no value since its conditional probability tables (CPTs) are not based on any expertise." However, I will apply Dung's calculus to the theory of argumentation sketched above. I will do so in a labeling-based form, which is not in Dung's original paper but was later proved to be equivalent to Dung's account.

A *labeling* of an argumentation framework (a set of arguments with a binary defeat relation) is an assignment of either the label *in* or the label *out* (but not both) to zero or more arguments such that for all arguments *A*:

1. *A* is *in* iff all arguments defeating *A* are *out*;
2. *A* is *out* iff *A* is defeated by an argument that is *in*.

This leaves room for multiple labeling policies ("semantics"). In the present paper, so-called *preferred labelings*[2] are considered, which maximize the set of arguments labeled *in*. Then, an argument can be defined as *justified* if it is *in* in all labelings, as *overruled* if it is *out* in all labelings, and as *defensible* if it is *in* in some but not all labelings.

Let us consider a few simple cases. With just two arguments that have one or two defeat relations between them the outcome is obvious: If *A* strictly defeats *B*, then *A* is justified while *B* is overruled; while if *A* and *B* defeat each other, they are both defensible since there are two labelings: one with *A in* and *B out* and one with *B in* and *A out*. With three arguments the situation is more interesting. See Fig. 1, in which dark gray stands for "justified," light gray for "defensible," and white for "overruled."

- In (a), *Arg*3 has to be *in* because it has no defeaters. Then, *Arg*2 has to be *out* so *Arg*1 has to be *in*. This is the only labeling, so *Arg*1 and *Arg*3 are justified while *Arg*2 is overruled.
- In (b), the outcome is the same, which agrees with the intuition that if *Arg*3 defends *Arg*1 against a strict defeat by *Arg*2, then *Arg*3 *a fortiori* defends *Arg*1 against a mutual defeat by *Arg*2.
- In (c), *Arg*2 and *Arg*3 defeat each other and are not defeated by *Arg*1. So they can both be labeled *in* if the other is labeled *out*. Thus, there are two labelings: in the one *Arg*1 and *Arg*3 are *in* while *Arg*2 is *out*, and in the other *Arg*2 is *in* while *Arg*2 and *Arg*3 are *out*. So all three arguments are defensible. This is so since *Arg*3 only partly defends *Arg*1 against *Arg*2.
- Finally, (d) is similar to (c): All three arguments are defensible.

We need one further complication. In reality, preference relations between arguments are often not simply given but can be argued about. Modgil (2009) gives the following example:

A: It will rain in Calcutta since the BBC says so
B: It will not rain in Calcutta since CNN says so
C: The BBC is more reliable than CNN since it is a more trustworthy institution

Here, *A* and *B* attack each other, so without a preference relation specified they defeat each other. However, *C* then attacks the claim that *B* defeats *A*, so as a result *A* strictly defeats *B,* and *A* and *C* are justified, while *B* is overruled. The model is recursive: *C* could itself be defeated by a conflicting preference argument:

D: CNN is more reliable than the BBC as shown by these statistics

Then, without a preference relation specified between *C* and *D*, these two preference arguments defeat each other about whether *A* and *B* defeat each other. As a result, all four arguments are defensible. But suppose now that

E: Statistics are more reliable than trust

which defeats the claim that *C* defeats *D.* In result, *B, D,* and *E* are justified while *A* and *C* are overruled. So the conclusion that it will not rain in Calcutta is justified.

The full formalization of these ideas is quite complicated (Modgil, 2009), but the present analysis will only contain simple examples of preference arguments, in which intuitions suffice.

2.3. Argumentation versus Bayesian and scenario-based approaches

In this section, the argumentation approach is contrasted to scenario-based and Bayesian approaches with respect to the kinds of generalizations (or conditional probabilities) that they assume. Argumentation approaches usually work with generalizations that are expressed from evidence to hypothesis and that are applied to pieces of evidence in a "forward" manner to draw conclusions about a hypothesis given this evidence. If a generalization applies to a piece of evidence, then the hypothesis is accepted unless there is evidence to the contrary. In Prakken (2014), I called this "direct" probabilistic reasoning. In contrast, scenario-based approaches usually assume generalizations that are expressed from cause to effect, which usually is from hypothesis to evidence (although not always: an exception is, for example, when the evidence is about a motive, in which case the direction is from evidence to hypothesis). Such generalizations from hypotheses to evidence are then applied in a "backward" manner to see whether the hypothesis can explain the evidence. If more than one hypothesis can explain the evidence, then the best explanation is chosen. In Prakken (2014), I called this "indirect" probabilistic reasoning. Likewise, in Bayesian networks, arcs and conditional probabilities are often specified from hypothesis to evidence, and then probability theory is used to compute which hypothesis is the most probable given the evidence. Both forms of "indirect" or "backward" reasoning can be seen as inference to the best explanation.

To illustrate the difference with an example, consider "smoke means fire"; that is, if there is smoke, then there is fire, and suppose we see smoke. Then, we conclude presumptively that there is fire. This is "forward" reasoning from evidence to cause and it can naturally be modeled in the present approach. Note that what "smoke means fire" in fact says is that fire is the *usual* cause of smoke; this is what licenses the presumptive inference. In a scenario-based approach, the reasoning is different. There we say "fire causes smoke" and conclude that fire can explain the observed smoke. But this is weaker than concluding that there presumably is fire: All that is concluded is that the smoke can have been caused by fire. To draw the stronger conclusion that the smoke was presumably caused by fire, we must compare fire as a cause of smoke with alternative causes, for example, a smoke machine or a malfunctioning mechanical device. The problem is that all these causes explain the observation, so we need an additional criterion for the comparison. For example, we could make predictions of discriminating facts from the alternative hypotheses ("if there is fire, then there must be heat but not if there is a smoke machine"). Or, if no new investigations are possible, we could look at the a priori probability of these causes, as is done in Bayesian approaches. One could say that accepting the evidential generalization "fire causes smoke" is based on the opinion that the outcome of a Bayesian calculation will usually be that the probability of fire given smoke is very high. Thus, evidential generalizations as used in argumentation could be seen as compressions, or summaries of probabilistic reasoning.

In the argumentation approach, alternative explanations to the usual one are regarded as exceptions to the generalization "if observation then usually cause." Undercutting arguments based on these exceptions can only be constructed if there is evidence for an exception. This is how the argumentation approach counters conspiracy theories. However, this approach is not generally applicable, since often there will not be a single usual cause of an observation, so accepting the evidential generalization "if observation then usually cause" is not justified. This is a genuine limitation of the argumentation approach.

This limitation shows itself, for instance, with witness testimonies. The usual approach in argumentation theory is to accept that what witnesses say is usually true and to treat cases where the witness is insincere, misremembers, or observed incorrectly as exceptions to this evidential generalization. This is also the approach that judges often take, since they often ask themselves whether they can believe the witness. In the argumentation-based approach, undercutting counterarguments to the argument that accepts the testimony as true can, given the just-listed exceptions, only be constructed if there is evidence that the witness is insincere, misremembers, or observed incorrectly. If there is no such evidence, then the testimony will be accepted as truthful by default. In contrast, in the scenario-based approach there is no a priori assumption that the truth of what the witness says is the usual cause of the witnesses' testimony. In a scenario-based approach this explanation of the testimony must be put on a par with alternative explanations, such as that the witness is insincere, misremembers, or observed incorrectly. In consequence, a judge who reasons according to the scenario-based approach cannot ask any more whether s/he can believe the witness. Instead, s/he should ask how likely the testimony is under the various hypotheses. Judges might find this a less natural way of evaluating witness testimony. Note that they should ask this question even if there is no evidence that the witness is insincere, misremembers, or observed incorrectly. As we just saw, this is different in the argumentation approach.

An interesting question is whether in criminal cases witness testimonies are generally so reliable that accepting the evidential generalization "what witnesses say is usually true" is justified or whether reasoning about witness testimonies should always take a scenario-based form, where the judge asks how likely the testimony is under the various hypotheses.

3. Method of the analysis

I designed my analysis in the following way. I did not construct my analysis from scratch, but I started with the court of appeal's verdict. I first considered its list of intermediate conclusions from the evidence (p. 74 of the court of appeal's verdict, the beginning of section e) and tried to group these by introducing more abstract conclusions. I then combined these with the court's conclusion that there is no likely alternative defender into the main argument for the conclusion that the accused is guilty. Then, for each of the court's intermediate conclusions, I investigated how the court supported it. This led to some small changes in the result of the initial phase. More seriously, I thought that the court's reasoning in supporting the intermediate conclusions was not entirely convincing, so I began to reshape it according to my own insights. I then considered the defense's counterarguments and how they can be attacked. In this phase, I also looked at the prosecution's requisitory. The result of all this was my own analysis.

In some cases when I did not regard an argument of one of the participants convincing, I simply ignored it. For example,

- The court's mobile phone evidence argument;
- The defense's arguments that the suspect did not have the physical ability to carry out the attack;
- Whether there could have been an unknown alternative attacker;
- The defense's arguments against the motive (the relationship had been improving recently);
- That no gun was found and the prosecution's suggestion that the suspect had gotten rid of the gun;
- The court's argument that the suspect has given no plausible alternative explanation of various observations.

When support of a claim with evidence was uncontroversial, I did not model this as a separate inference. Also, in general, I tried to model the reasoning from evidence to hypothesis as explained in Section 2.3, but in several cases I deviated from this approach. I especially did this in cases where the court observed that some observation can be explained by a particular cause and that there is no other likely explanation. Schematically:

(1) E is observed; (2) E can be explained by C; (3) there is no other likely explanation of E; therefore, (presumably) E was caused by C.

A reconstruction from evidence to cause would in such cases yield

(1) E is observed; (3) E is usually caused by C; therefore, (presumably) C is the case.

However, this reconstruction seems less natural in the examples in the case. For instance, it is less similar to the actual wordings used by the court. Examples in which I took this deviating approach are in the supporting arguments of premises p_2, p_3, and p_4. In Section 7, I will discuss what would happen in the "standard" argumentative approach.

4. Introducing the analysis

In this section, the analysis is introduced, while in the next section the full analysis is displayed in diagrammatic form. The names of arguments in this section correspond to their names in Fig. 9 in the next section. In its main argument A, the court essentially argues that the suspect killed the victim since he had the opportunity, the means, and a motive to kill the victim while several observations can be explained if the suspect is the attacker. This becomes explicit in the support of premise p_4 of the main argument. The main attack on A is argument B for the conclusion that Perry S. (from now on "Perry") is a plausible attacker of the victim, supported by the suspect's testimony that a man had suddenly jumped from the bushes and had attacked him and his wife, by the map of the crime scene area found with Perry with, among other things, a cross at the location where a pit was found and by some other evidence. This attacker is then rebutted by argument C on the grounds that no traces of Perry were found on the crime scene and that several observations cannot be explained if Perry attacked the victim.

No gun was found at the crime scene, even though the claim that the suspect had carried out the attack with the gun was an important ground for the court's decision. In argument *A*, this is captured in its subargument for f_9 that the suspect had the means to kill the victim. Based on the generalization that if no gun was found at the crime scene then usually no gun was used in the crime, a rebuttal *G* of *A*'s subargument for the conclusion $f_{9.1}$ can be constructed. This rebuttal was attacked by the prosecution in a detailed analysis of how the suspect had the time and means to get rid of the gun. For simplicity, only the top-level part of this argument (called *H*) is given below. In my reconstruction, argument *H* undercuts argument *G*.

The remaining arguments are essentially about the evaluation of arguments *B* and *C* on whether Perry is a plausible alternative attacker of the victim. Arguments *F* and F' attack two premises of argument *C* by offering explanations of observations assuming that Perry was the attacker. Arguments *D* and *E* are two priority arguments about the rebutting conflict between arguments *B* and *C* on whether Perry is a plausible alternative attacker. The premises of argument *D* are not added as premises to argument *C* for the conclusion that Perry is not a plausible attacker, since they are not directly reasons for this conclusion; instead, they say something about the strength of the reasons for or against this conclusion. Note also that the reasons concerning the pit are not formulated as attacks on $f_{13.3.3}$ since they are about the conclusive force of the combination of $f_{13.3.3}$ with the other premises. Without the other premises it is not very relevant that the pit could have been of Perry. Finally, arguments *I, J, K,* and *L* are attacks and counterattacks concerning the priority arguments *D* and *E.* Note that in argument *K,* witness reliability as interpreted by the court of appeal clearly is the probability of *P* given that the witness says that *P*.

5. The analysis

This section contains the actual analysis of the Simonshaven case, in the form of visualizations of all arguments plus the resulting extended argumentation framework and some of its labelings (Figs. 2–11). In the argument diagrams, the following notational conventions have been used. Solid links between boxes are inferential relations, and tick dashed arrows display attack relations. Statements in arguments (whether premises, intermediate, or final conclusions) that are attacked are displayed with thick borders and a gray color. The notation in the pictures with extended argumentation frameworks (Figs. 9–11) is explained below in the Outcome section.

6. The outcome

The outcome will now be analyzed in terms of Fig. 9, which displays the analysis as an extended argumentation framework in the sense of Modgil (2009). That is, it displays

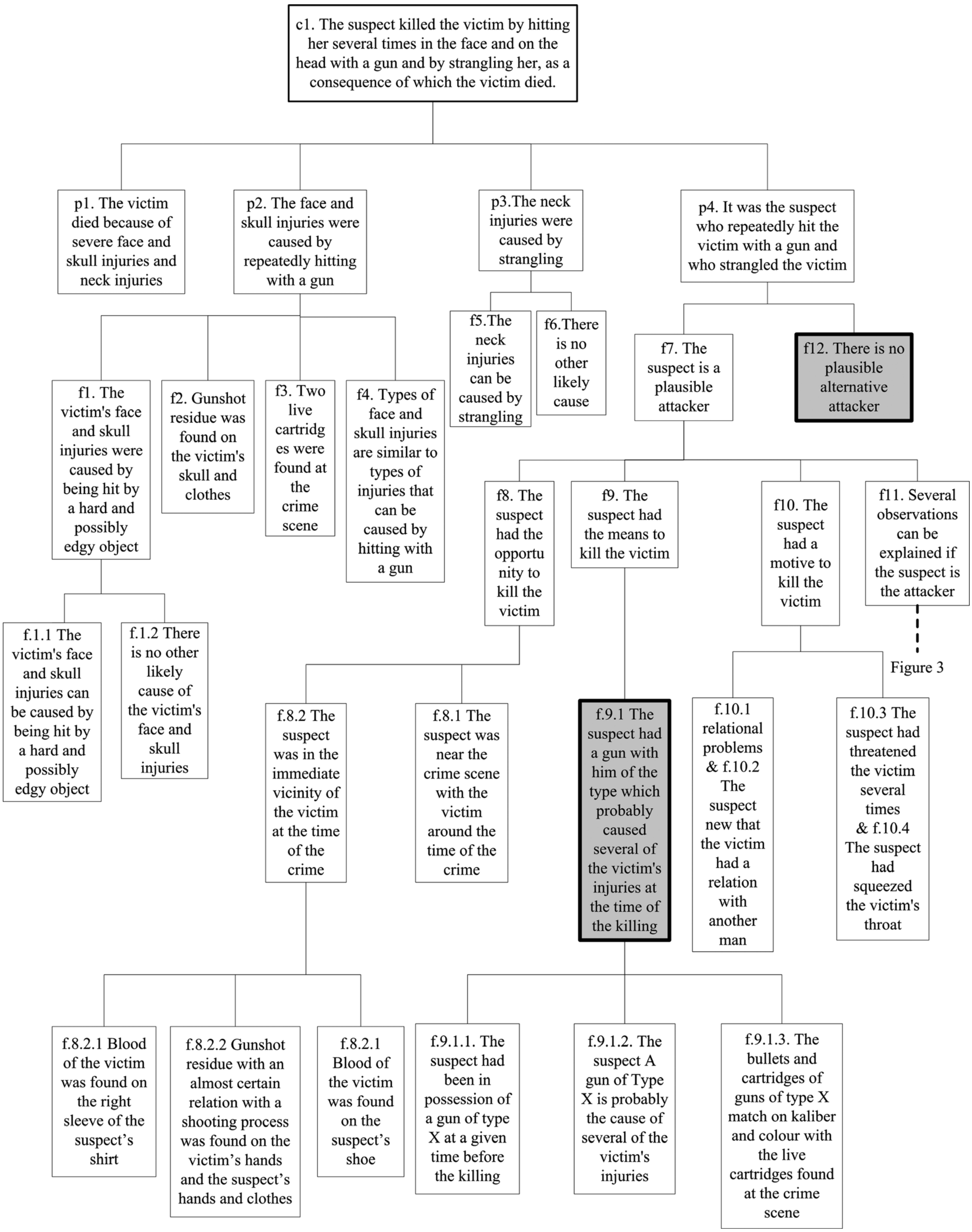

Fig. 2. The main argument *A* for guilt (incomplete).

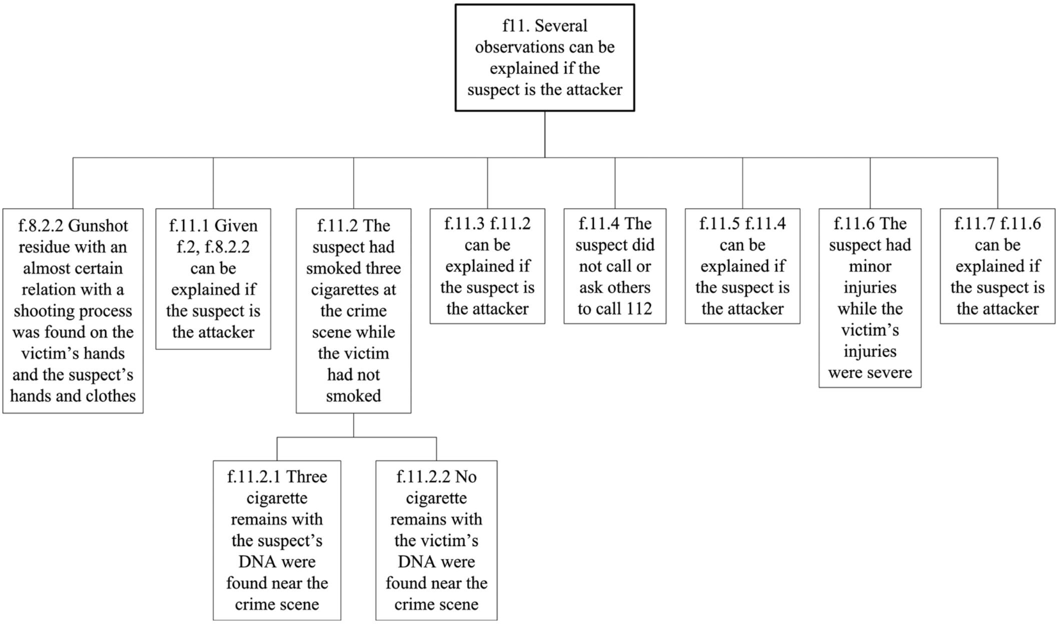

Fig. 3. The remaining part of the main argument *A* for guilt.

all attack relations between the arguments plus the attack of the priority argument *D* on the rebutting attack of argument *B* (Perry is a plausible attacker) on argument *C* (Perry is not a plausible attacker). To this end, the symmetric attack between arguments *B* and *C* is displayed with two separate arrows. Strictly speaking, this should also be done with all other symmetric attacks, but for simplicity they have been visualized with a single bidirectional arrow. The figure is further simplified in that the subarguments of the arguments are not separately displayed. In a few cases this means that it is not visualized that an attack of an argument is in fact an attack on one of its subarguments. For example, argument *G* attacks argument *A* by rebutting its subargument for the intermediate conclusion that the suspect was wearing a gun. Since this direct rebuttal is symmetric, the attack between *A* and *G* has also been visualized as symmetric, even though strictly speaking the indirect attack of *G* on *A* is asymmetric. The same holds for the attacks between *E* and *I* (where *I* undermines *E* on a premise and this premise in turn rebuts *I)*, for the attacks between *C* and *F* and *F′* (where *F* and *F′* symmetrically rebut a subargument of *C)*, and the attacks between *J* and *L* (where *L* symmetrically undermines *J)*.

So far, no explicit argument preferences have been specified, except the two priority arguments *D* and *E.* Therefore, all attacks in the figure succeed as defeats except that the attack of *B* on *C* succeeds or does not succeed as defeat depending on whether argument *D* is labeled *in* or *out*. Fig. 10 displays a labeling in which argument *D* is accepted, so that the attack of *B* on *C* does not succeed as defeat. In consequence, the main argument *A* that the suspect is guilty has to be labeled *in* (note that its defeater argument *G* is *out* in all labelings since it is undercut by argument *H*, which has no attackers and

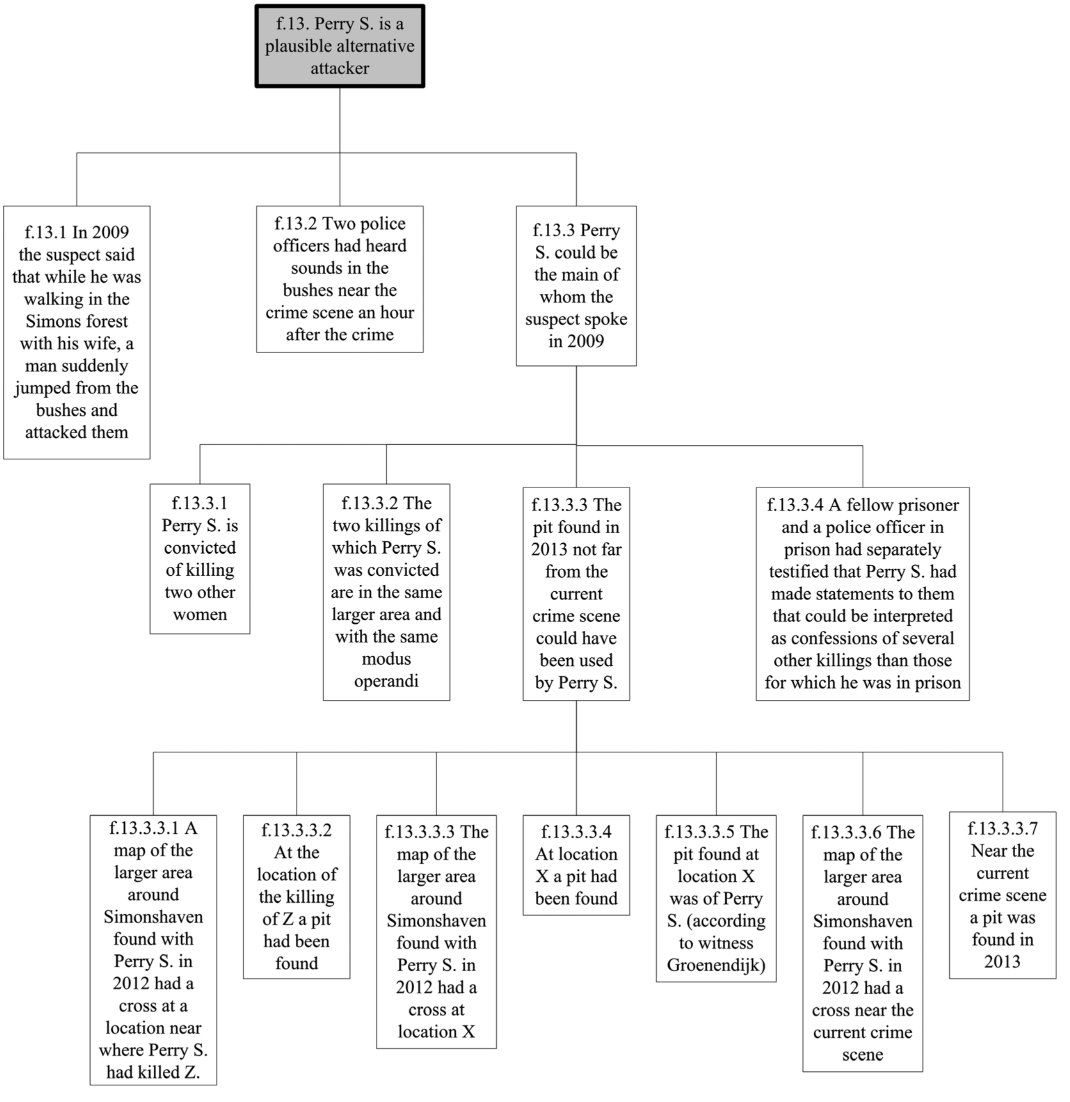

Fig. 4. The main argument *B* against guilt.

undercutting attack always succeeds as defeat). Note that making argument *D* (and thus argument *A*) *in* requires that either argument *L* or argument *K* is labeled *in* in order to make *D*'s defeater *J out*.

However, there are also labelings in which the main argument *A* for the suspect's guilt is *out*. One of them is displayed in Fig. 11, in which *B* is accepted as *in* at the expense of *C*. This requires that *E*'s underminer *I* is rejected as *out*. There are further labelings of the graph, but the just discussed two labelings suffice to conclude that the main argument *A* that the suspect is guilty is defensible since it is *in* in some but not in all labelings. So on the basis of the present analysis, a conviction is not justified.

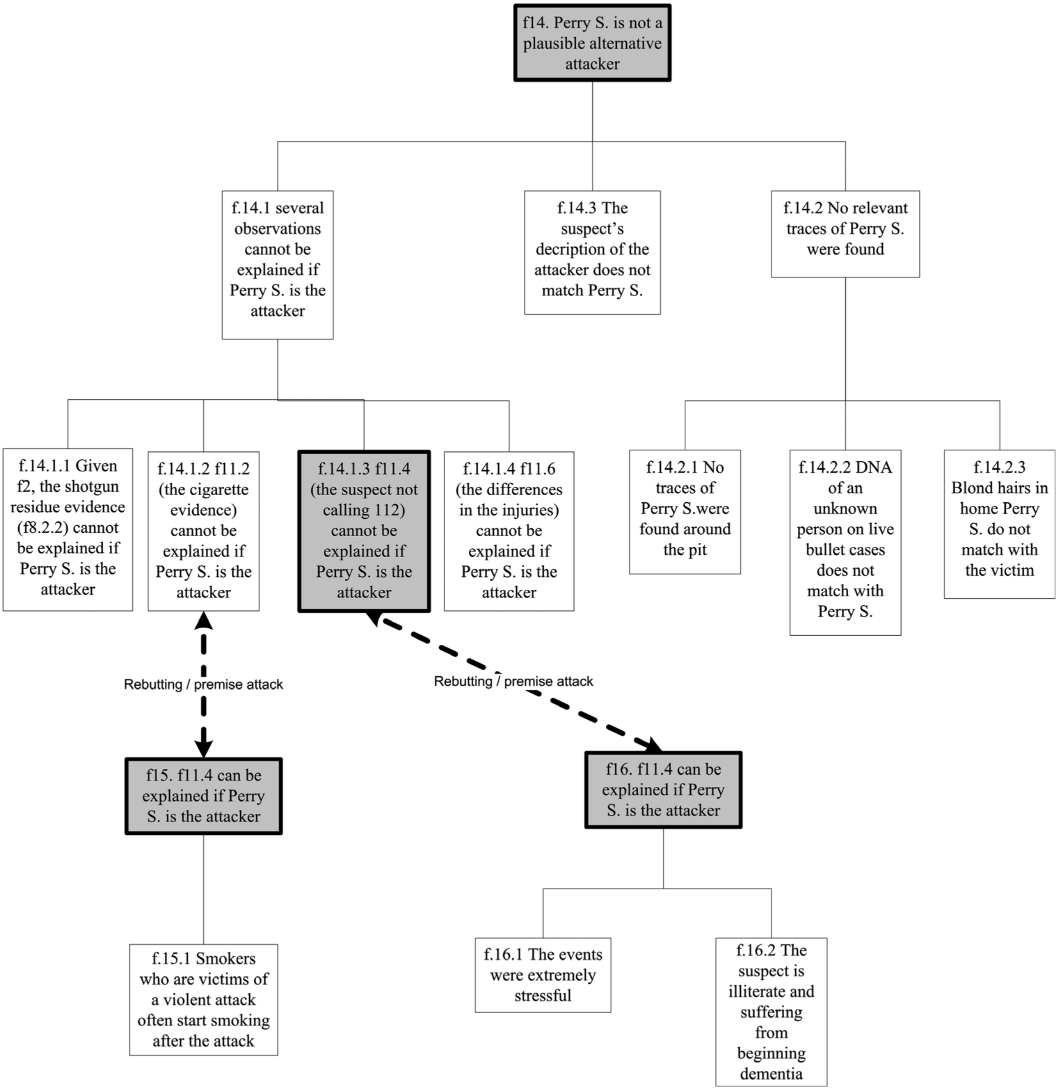

Fig. 5. The main attack C on the main argument against guilt.

However, this is not the end of the analysis. It is interesting to see which preference relations on rebutting or undermining arguments can make argument A justified. This can be done by selecting a labeling in which A is *in* and by then adding preference relations that together make this the only possible labeling. Such preference relations turn symmetric into asymmetric defeats by saying that one of the two corresponding attacks does not succeed as defeat. So to make A justified, the labeling of Fig. 10 has to be made the only one. The following four sets of preference relations each suffice to make argument A justified. In all of them, C's attacked subarguments are preferred over their attackers F and F', and furthermore either:

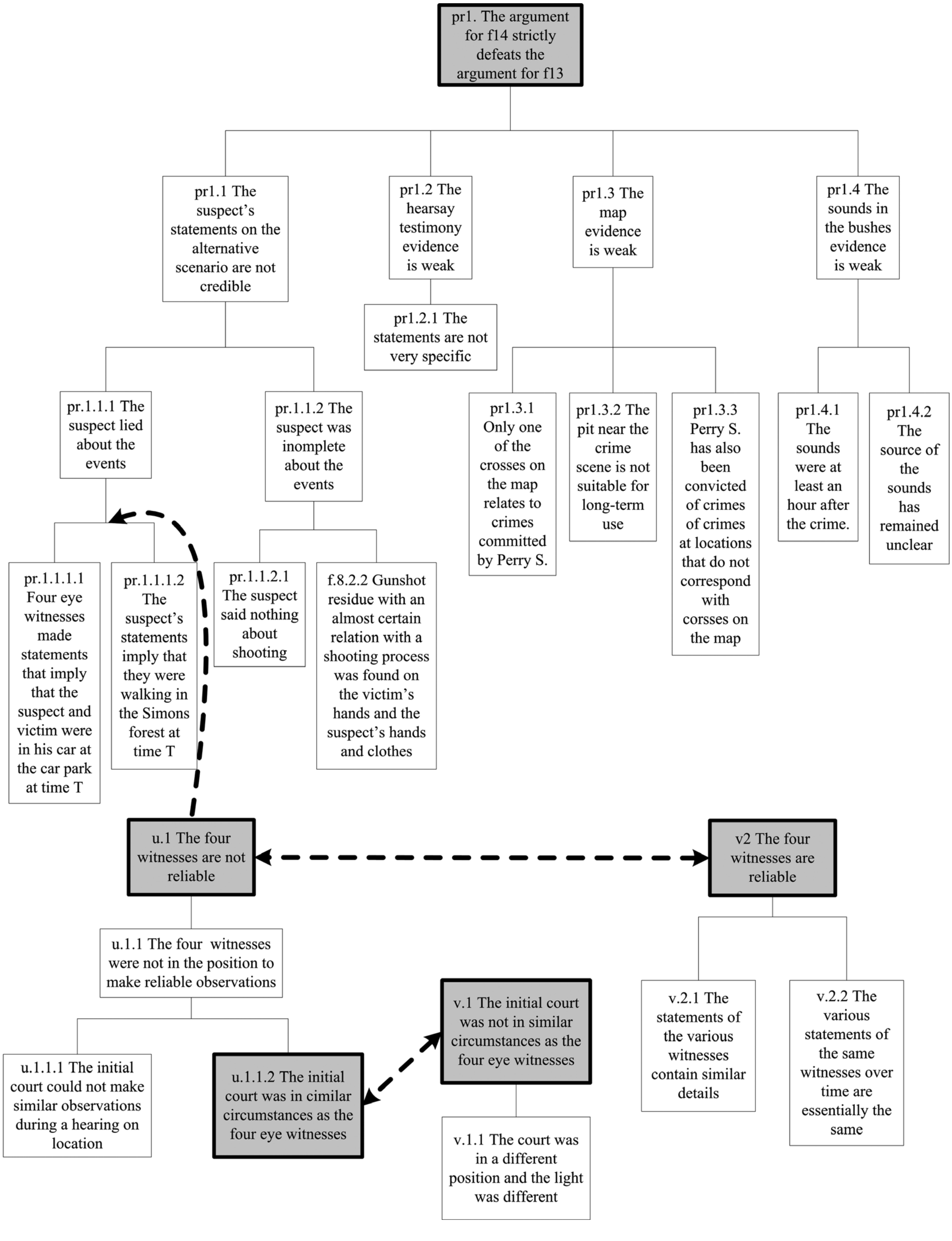

Fig. 6. Priority argument *D* and (counter) attackers *J, K, L*.

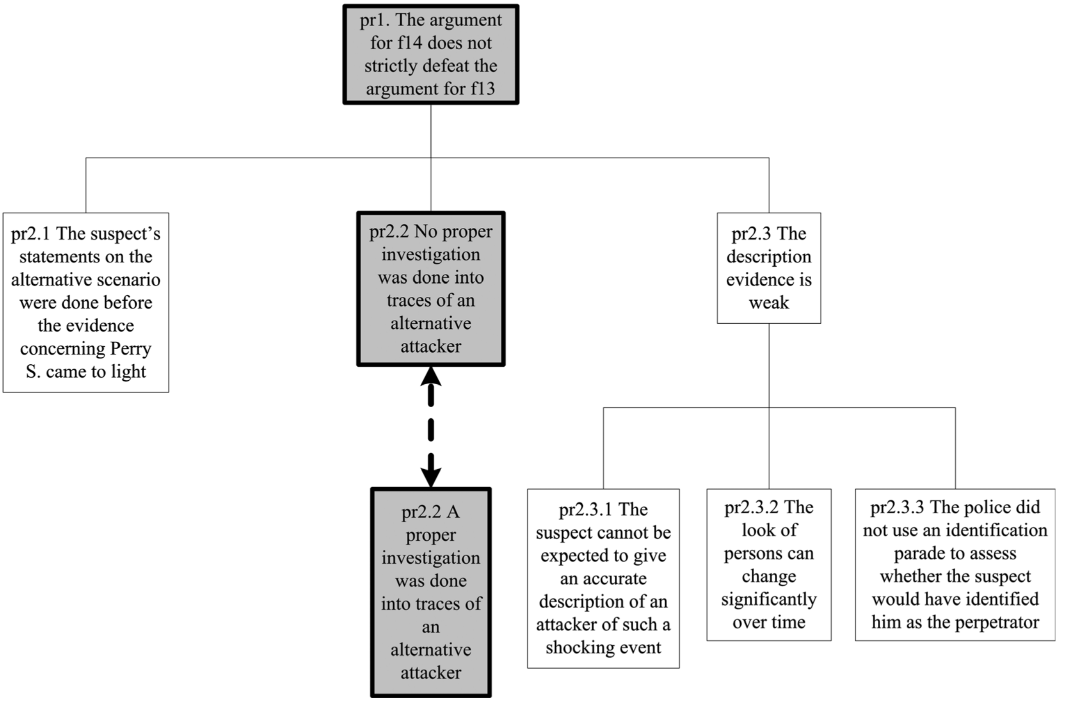

Fig. 7. Conflicting priority argument *E* and attacker *I*.

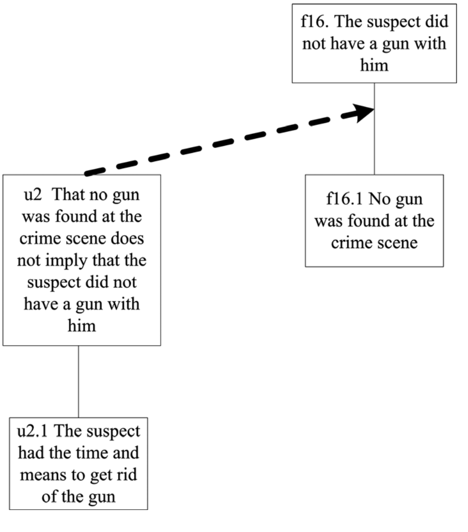

Fig. 8. Arguments *G* and *H* about the gun.

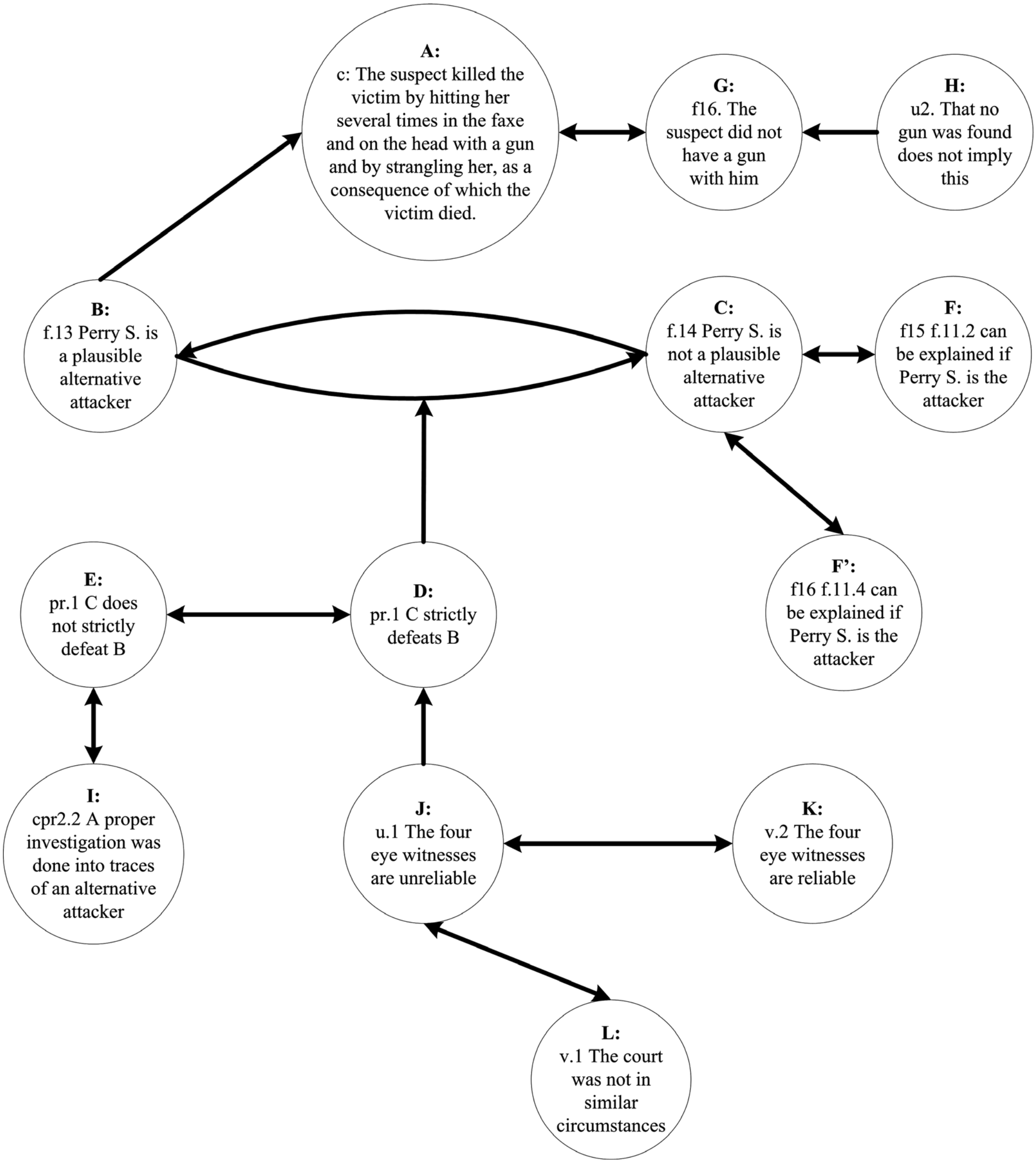

Fig. 9. An extended argumentation framework.

- *D* is preferred over *E* and *K* is preferred over *J;* or
- *D* is preferred over *E* and *L* is preferred over *J*'s attacked premise; or
- *I* is preferred over *E*'s attacked premise and *K* is preferred over *J;* or
- *I* is preferred over *E*'s attacked premise and *L* is preferred over *J*'s attacked premise.

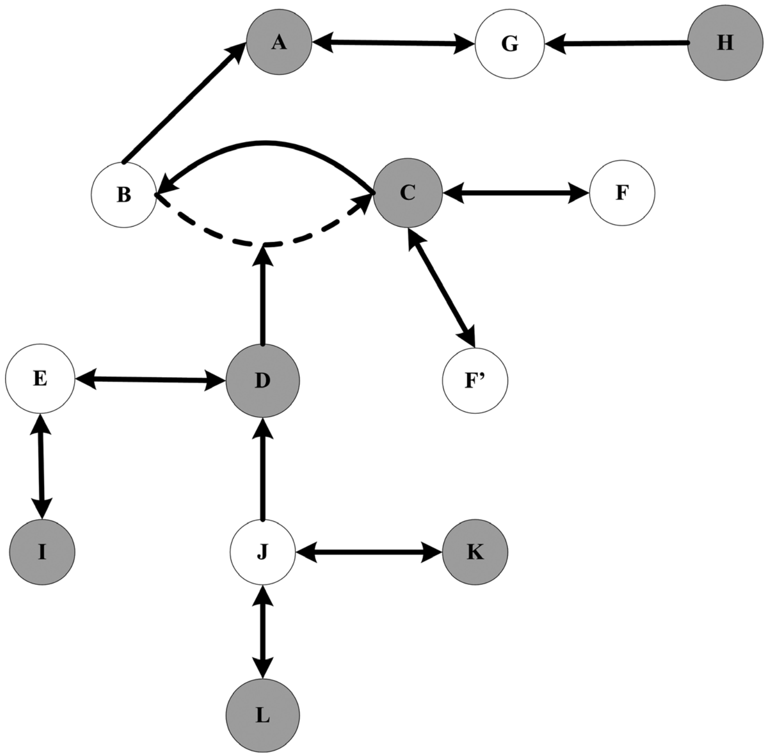

Fig. 10. A labeling in which the court's main argument is accepted.

Thus, given the arguments and their attack relations listed in this section, the main issues are:

- Whether the argument that Perry is a plausible alternative attacker of the victim is sufficiently strong to be able to defeat the argument for the opposite conclusion;
- Whether a proper investigation was done into traces of an alternative attacker;
- Whether the suspect's behavior after the crime can be explained if Perry is the attacker.

This illustrates one potential benefit of the present approach. It reveals on which issues preference relations have to be expressed, and it allows us to analyze the consequences of particular preference relations, analogous to sensitivity analysis in Bayesian networks. Note that further arguments can change the result. For example, any nonevidence premise can be attacked and any argument can be undercut on its underlying generalization, either on the grounds that in the present case it has an exception or on the grounds that it is not justified at all. Such arguments can be constructed even if no further evidence becomes available, since they are in fact ways to evaluate the premises and inferences of the above arguments.

One might criticize this kind of analysis on the grounds that since defeat is an all-or-nothing matter, relatively minor issues can have too much weight. For instance, one

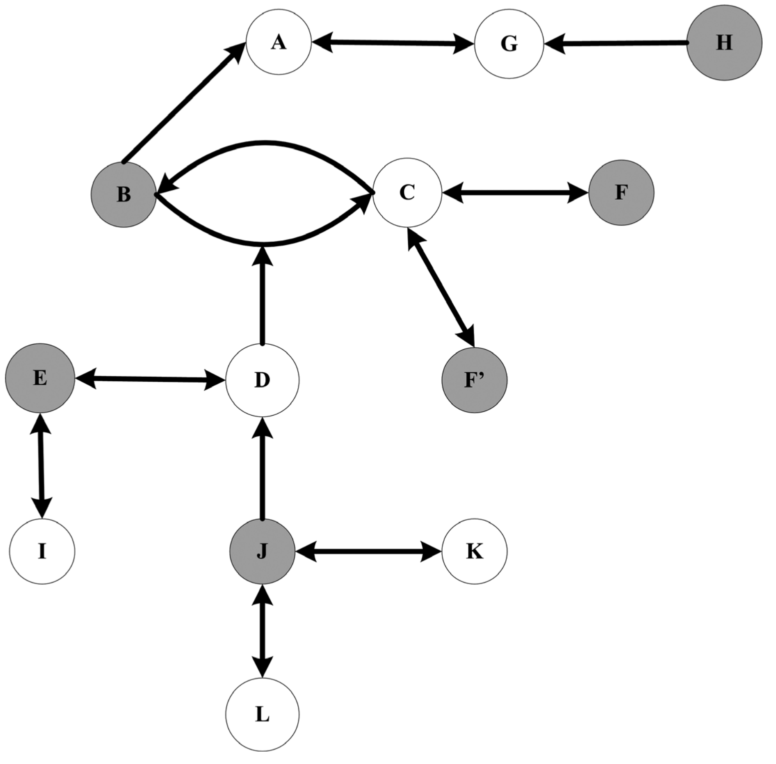

Fig. 11. A labeling in which the court's main argument is not accepted.

might argue that just accepting that the suspect's strange behavior after the crime can be explained if Perry is the attacker should not be sufficient to turn the status of the main argument for guilt from justified to defensible. However, there is a way to make the analysis more fine grained. Looking at the above arguments, it can be seen that many of them in fact aggregate various reasons for their conclusion into one conjunctive reason for the same conclusion. For instance, in argument *C,* for the conclusion that Perry is not a plausible attacker, each of the subconclusions, $f_{14.1}$, $f_{14.2}$, and $f_{14.3}$, on its own provides a reason for the conclusion, but their combination is a (possibly) stronger reason. And the same holds for the three premises of subconclusion $f_{14.1}$ that several observations cannot be explained if Perry is the attacker: Each of these premises on its own supports the conclusion, but together they do so more strongly. So if one or two of these premises are rejected by accepting argument F and/or F', then the remaining premise(s) can still be used in an alternative, usually somewhat weaker argument for subconclusion $f_{14.1}$ with a subset of the premises of the original argument for the same conclusion. This might in turn make the support for the main conclusion of argument C somewhat weaker, which might increase the possibility that the evaluator will accept priority argument E and reject priority argument *D,* thus accepting that B defeats *C.* A similar analysis can be given of the relation between arguments E and *I*: If I is accepted, then a version of E can be constructed without premise $pr_{2.2}$ and possibly somewhat weaker than the original version of E.

In the literature on formal argumentation, this is known as the issue of accrual of arguments (Prakken, 2005). Various formal treatments of this phenomenon exist. The present treatment is based on the idea that if a generalization "If P_1 and... and P_n then (qualifier) Q" is adopted in which each P_i on its own is a reason for Q, then any generalization for Q with some (but not all) P_i's deleted can also be adopted. For reasons of space this treatment of accrual is left implicit in Fig. 9 in that this figure leaves various versions of argument C possible. But this analysis can be made precise in the way just explained. In particular, the phrase

> In all of them, C's attacked subarguments are preferred over their attackers F and F'...

has to be refined to

> In all of them, C's attacked subarguments are preferred over their attackers F and F' or else C is replaced by a version without the attacked subargument(s)...

7. Discussion

The present case study has by and large confirmed the often-claimed strong and weak points of the argumentation approach to evidential reasoning (Kaptein, Prakken, & Verheij, 2009; Verheij et al., 2016). Strong points are the possibility to clearly identify the support and attack relations between evidence and hypotheses considered by a fact finder, and to clearly identify the points of disagreement in a case. As in Wigmore charts, the tree-structured arguments explicitly show how evidence is relevant to hypotheses. Moreover, the distinction in *ASPIC+* between several kinds of attack relations between arguments allows to state precisely on which issues preference relations have to be expressed. The labeling-based semantics of *ASPIC+* then allows to analyze the consequences of particular preference relations, analogous to sensitivity analysis in Bayesian networks, which tests the influence of changes in input probabilities on the posterior probability of interest. Weak points are the lack of overview over a case (which is a strong point of scenario-based approaches) and the lack of a systematic account of degrees of uncertainty (which is a strong point of Bayesian approaches; but recall that recent work tries to integrate argumentation and probability theory).

I conclude by discussing the present case study four guiding questions from the special-issue proposal on the Simonshaven case.

To what extent is the analysis objective and to what extent is it based on subjective beliefs, assumptions, and choices?

This question can be rephrased as the question how likely it is that other analysts applying the same method would produce the same or a similar analysis. While this question can strictly speaking only be answered in a controlled experiment, there is much reason to believe that my analysis is not very reliable in this sense, even though I have abstained from committing to preference relations between arguments. Above all, different analysts might have different opinions on which evidence is relevant and/or on how various pieces of evidence relate. These two issues are related to the issue of which generalizations can or should be accepted. Also, an analysis can be more or less fine-grained and can vary on the intermediate conclusions that are assumed between the evidence and the conclusions. This also relates to the issue of which generalizations can or should be accepted. Moreover, alternative ways are possible to group pieces of evidence into object-level, undercutting, or priority arguments. Finally, there seems no fully objective way to select nonevidential premises, such as, for example, the various premises that some observation can or cannot be explained if some person is or is not the attacker.

I now make some specific observations in this respect.

At first sight, my analysis would seem to commit to the general argument scheme that if a suspect had the opportunity, the means, and a motive to commit the crime and there is no other plausible offender, then the suspect is guilty (namely, in the support of premise $p4$ that it was the suspect who repeatedly hit the victim with a gun and who strangled the victim). In the workshops during which the Simonshaven case was discussed, Christian Dahlman argued that in his country, Sweden, this scheme is generally by the judiciary regarded as too weak to justify a conviction. However, upon closer inspection it turns out that my analysis applies a strengthened version of this scheme. The support of p_4 also contains premise p_{11} that several observations can be explained if the suspect is the attacker, while, moreover, the indirect defense of premise p_{12} that there is no plausible alternative attacker states that these observations cannot be explained if Perry attacked the victim. See the argument for $\neg f_{13}$ that Perry is not a plausible attacker, which defends the argument for p_4 against the attack on its premise f_{12} by the argument for f_{13} that Perry is a plausible alternative attacker. So a more accurate summary of the main argument scheme is that if a suspect had the opportunity, the means, and a motive to commit the crime, *several observations can be explained if the suspect is the offender*, and there is no other plausible offender, then the suspect is guilty. The main issue in my analysis, whether Perry is a plausible alternative offender, includes the subissue whether the observations that can be explained if the suspect is the offender can also be explained if Perry is the offender.

In my analysis, I consider Perry as the only plausible alternative offender. Christian Dahlman in his analysis of the case during the London 2017 workshop, also considers a "mobster" scenario, in which the accused, who is known to have been involved in criminal activities, was attacked by another criminal for some reason related to his involvement in criminal activities. I have left this out of my analysis as too implausible to deserve consideration. The alternative opinion can be modeled by generalizing f_{13} to "There are plausible alternative attackers" and by having the current argument B and an argument based on the mobster hypothesis as two accruing subarguments supporting the generalized f_{13}.

Finally, an issue that I did not analyze in detail in my analysis is whether it is relevant that no gun was found at the crime scene. Presumably a detailed analysis is possible about whether premise $u_{1.1}$ is justified. I left this analysis out because of space constraints.

How natural is the analysis from a cognitive and legal point of view?

In my analysis, I have not modeled how a court could *reach* its decision but how it could *justify* its decision. In this respect, it is interesting to see how close my reconstruction is to the wording of the court of appeal's decision. The court's main argument contains almost no structure. First, there is a lengthy discussion of the evidence, then there is a list of conclusions drawn from this discussion (p. 74 of the court's decision), and then it is said without further reasoning that, thus, the charge has been proven. Finally, two possible objections are briefly discussed. This top-level structure of the court's justification is argumentative. However, since it is quite coarse, we cannot say for sure that the court reasoned in an argumentative way; it may well be that in going from one stage to another it reasoned in a story-like or even Bayesian way but left this implicit in its justification. In some places it seems obvious that the court reasoned in a scenario-based way, namely, when it considers explanations of observations. However, the court does not always show awareness that the possibility of alternative explanations has to always be considered. In my analysis, I have tried to avoid this mistake by always when a premise "*X* can be explained if *Y* is the case" also including a premise "*X* cannot be explained if *Y* is not the case." In other places, the court clearly reasons in an argumentative way from evidence to a hypothesis, in particular, when it concludes that the suspect lied since he contradicted various witnesses. Here, the court clearly reasons from "the witnesses said that *P* is the case" to "*P* is the case." Moreover, when the court discusses the reliability of these witnesses, it clearly interprets it as the question whether *P* is the case given that the witnesses said that *P* is the case.

In conclusion, my analysis is not far from how the court in this case has justified its decision, although my analysis contains more detail, especially on specifying the nature of the relation between the evidence and the conclusions. Whether this is good or bad depends, of course, on one's opinion on how courts should justify their decisions. In this respect, it is relevant that criminal courts in the Netherlands increasingly phrase their decisions in complex cases in part in terms of scenario comparison, although large parts of their decisions are still phrased in an argumentative way (Stevens, 2014).

Did your analysis identify errors or biases in the reasoning of the judge, prosecutor, or defense?

I see one point at which this may be the case. An important issue in the case is whether the suspect lied about the events. The court concludes this from the contradictions of the suspect's statements with the statements of the four witnesses. In my analysis, this is taken into account in the priority discussion about the arguments that Perry is or is not a plausible alternative attacker. It might be argued that the court here fails to take into account that the probability that the suspect lies is reduced by the evidence concerning Perry. In my analysis, this is taken into account, namely, in the rebutting priority argument that the suspect's statements on the alternative scenario were done before the evidence concerning Perry came to light (premise $pr_{2.1}$). However, it might be argued that this is not the most principled analysis and that a (qualitative or quantitative) scenario-based analysis can give a more principled analysis.

Does your analysis respect the legal constraints, such as the burden and standard of proof and the right to remain silent?

Prakken and Sartor (2009) argue that in an argumentation approach the burden of proof should be modeled as the requirement to have a justified argument for the conviction. My analysis in Section 6 respects this. The right to remain silent was, in my analysis, respected by not attaching any value to the fact that the accused remained silent after the first interrogation. In fact, Dutch procedural criminal law does not forbid judges to draw inferences from the fact that by being silent, the accused did not deny the facts nor offer an alternative reading of the facts that are held against him. So it might be argued that my application of this principle is legally too strict. However, this is not an inherent feature of the argumentation approach. Arguments based on the accused's silence that are judged admissible according to Dutch procedural law can be incorporated in the analysis.

Acknowledgments

This work has partly been made possible by financial support and hospitality of the Isaac Newton Institute for Mathematical Sciences, Cambridge, supported by EPSRC grant no EP/K032208/1. I thank the participants of the Cambridge 2016 and London 2017 workshops on the Simonshaven case for their feedback on earlier versions of my analysis.

Notes

1. If a statement is used more than once in an argument, then the argument strictly speaking is a directed acyclic graph; I will ignore this complication for simplicity.
2. From now on the qualifier "preferred" will be omitted if there is no danger for confusion.

References

Anderson, T., & Twining, W. (1991). *Analysis of evidence. How to do things with facts based on Wigmore's science of judicial proof.* Boston: Little, Brown and Company.

Bex, F., Prakken, H., Reed, C., & Walton, D. (2003). Towards a formal account of reasoning about evidence: Argumentation schemes and generalisations. *Artificial Intelligence and Law*, *12*, 125–165.

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.

Kaptein, H., Prakken, H., & Verheij, B. (eds.) (2009). *Legal evidence and proof: statistics, stories, logic*. Farnham, UK: Ashgate.

Modgil, S. (2009). Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, *173*, 901–934.

Modgil, S. (2014). The ASPIC+ framework for structured argumentation: A tutorial. *Argument and Computation*, *5*, 31–62.

Modgil, S., & Prakken, H. (2013). A general account of argumentation with preferences. *Artificial Intelligence*, *195*, 361–397.

Prakken, H. (2005). A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law* (pp. 85–94). New York: ACM Press.

Prakken, H. (2014). On direct and indirect probabilistic reasoning in legal proof. *Law, Probability and Risk*, *13*, 327–337.

Prakken, H. (2018). Probabilistic strength of arguments with structure. In M. Thielscher, F. Toni, and F. Wplter (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference* (pp. 158–167). AAAI Press: To appear.

Prakken, H., & Sartor, G. (2009). A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, & B. Verheij(Eds.), *Legal evidence and proof: Statistics, stories, logic* (pp. 223–253). Farnham, UK: Ashgate.

Stevens, L. (2014). Bewijs waarderen. hoe doen strafrechters dat? *Nederlands Juristenblad*, *89*(40), 2842–2850. In Dutch (Evaluating evidence: how do judges in criminal cases do this?).

Timmer, S., Meyer, J.-J., 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. (2003). Dialectical argumentation with argumentation schemes: An approach to legal logic. *Artificial Intelligence and Law*, *11*, 167–195.

Verheij, B., Bex, F., Timmer, S., Vlek, C., Meyer, J.-J., Renooij, S., & Prakken, H. (2016). Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, *15*, 35–70.

Walton, D. (2006). *Fundamentals of critical argumentation*. Cambridge, UK: Cambridge University Press.

Wigmore, J. (1931). *The principles of judicial proof* (2nd ed.). Boston: Little, Brown and Company.