Law, Probability and Risk (2007) **6**, 145−168 doi:10.1093/lpr/mgm007
Advance Access publication on July 7, 2007

Sense-making software for crime investigation: how to combine stories and arguments?

FLORIS BEX
Faculty of Law, University of Groningen, Groningen, The Netherlands

SUSAN VAN DEN BRAAK & HERRE VAN OOSTENDORP
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

HENRY PRAKKEN†
Department of Information and Computing Sciences, Utrecht University, and Faculty of Law, University of Groningen, Groningen, The Netherlands

BART VERHEIJ
Department of Artificial Intelligence, University of Groningen, Groningen, The Netherlands

AND

GERARD VREESWIJK
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

[Received on 3 April 2007; revised on 17 April 2007; accepted on 20 April 2007]

Sense-making software for crime investigation should be based on a model of reasoning about evidence that is both natural and rationally well-founded. A formal model is proposed that combines artificial intelligence formalisms for abductive inference to the best explanation and for defeasible argumentation. Stories about what might have happened in a case are represented as causal networks and possible hypotheses can be inferred by abductive reasoning. Links between stories and the available evidence are expressed with evidential generalizations that express how observations can be inferred from evidential sources with defeasible argumentation. It is argued that this approach unifies two well-known accounts of reasoning about evidence, namely, anchored narratives theory and new evidence theory. After the reasoning model is defined, a design is presented for sense-making software that allows crime investigators to visualize their thinking about a case in terms of the reasoning model.

Keywords: crime investigation; sense-making; explanation; stories; abduction; argumentation.

† Email: henry@cs.uu.nl

1. Introduction

Crime investigation is a difficult and laborious process, and the costs of mistakes can be high. Especially in large crime investigations, investigators are faced with a mass of unstructured evidence of which they have to make sense. They have to map out the possible hypotheses about what happened and assess the potential relevance of the available evidence to each of these hypotheses. Because of the difficulty of this task and the high costs of mistakes, it is worth investigating how crime investigators could benefit from support tools. It has been suggested that software could offer such tools by supporting crime investigators in expressing their reasoning about a case in terms of arguments on the relevance of evidence to the various hypotheses using common knowledge (Schum and Tillers, 1991). With such a programme, the user could visualize his reasoning in various ways (graphs, tables, forms) and explore its consequences. Thus, the programme would support investigators in seeing patterns, discovering new relationships or inconsistencies and identifying missing evidence. Such software would also facilitate transfer of the case files to others by increasing the transparency of the files, so that subsequent investigators, prosecutors and fact finders could gain a quicker and better understanding of the case.

In the current practice of crime investigation and similar fact-finding processes, software for managing and visualizing evidence is already being used. Well-known examples are Analyst's Notebook (2006) and HOLMES 2 (2006). In the Netherlands, the BRAINS system has been experimentally used in crime investigation with some success (van der Schoor, 2004). However, a limitation of such software is that it does not allow for expressing the reasons why certain pieces of evidence support or attack a certain hypothesis. Thus, the 'core business' of structuring evidence and assessing the relation between hypotheses and evidence is still, in fact, wholly dependent on human reasoning, and the structures resulting from such reasoning cannot be recorded and analysed by the software.

This article reports on an ongoing research project on software that could add these so-called 'sense-making' capabilities to current evidence management software. The aim of this project is to develop proof-of-concept software in which a human crime investigator can visualize possible stories about what happened and can link these stories through arguments with available supporting or attacking evidence. We anticipate that to make such software useful in practice, it should be combined with systems like HOLMES 2, Analyst's Notebook and BRAINS, so that the visualizations are linked with the documents containing the available evidence and thus make these documents more manageable. To this end, our software already supports links between components of the visualizations and pieces of text within source documents. However, a full integration with current professional document management software is outside the scope of this project.

In this article, we focus on one aspect of our project, namely, the design of the reasoning model underlying our system. We will investigate this aspect from both a theoretical and a software perspective, and we will especially focus on the role of causal information in reasoning about evidence. It has been argued that the only viable manner in which police investigators can structure the information that they gather is through constructing stories about what might have happened (Poot *et al.*, 2004). The stories should explain the available evidence by explaining what might have caused the evidence to be available. The same has been argued for triers of fact and others involved in criminal proceedings (Wagenaar *et al.*, 1993). However, others have advocated a method of building arguments from the available evidence to the hypotheses by applying commonsense generalizations (Anderson and Twining, 1991; Schum and Tillers, 1991). In this method, causal statements (for instance, the shooting has caused the appearance of gun powder) are often inverted to evidential statements (gun powder indicates that a shooting took place). In this paper, we want to investigate to which extent these two reasoning models (stories and arguments) can be used for our purposes. Very briefly, we will claim that there are reasons to combine storytelling in the form of abductive inference to the best explanation (IBE) with defeasible argumentation for linking the stories to the available evidence. After having outlined a formal model that combines both forms of reasoning, we will then describe our current design of software that visualizes reasoning based on this combined model.

2. Current theoretical approaches

In this section, we sketch the main current theoretical approaches to modelling legal reasoning about evidence. To be usable in practice, our software design should be based on concepts that are actually used in such reasoning, while to improve the quality of crime investigations, the design should be based on a rationally founded theory of these concepts. Methods founded on probability theory (such as probabilistic networks) satisfy the second criterion but arguably not the first since in the vast majority of legal cases, reliable statistics are not available, while human experts are often unable or reluctant to provide quantitative estimates. This project therefore takes its starting point in three important concepts of current legal evidence theory, viz. arguments, generalizations and stories (also called ‘narratives’). Two legal theoretical approaches that take these notions seriously are new evidence theory (NET) and anchored narratives theory (ANT). In artificial intelligence (AI), mainly models of abduction and IBE have been applied and sometimes also logics for default-style non-monotonic reasoning.

2.1 *NET and ANT*

The importance of the concepts of arguments, stories and generalizations has been emphasized in two recent research strands in legal theory, ‘NET’ (e.g. Anderson and Twining, 1991; Schum and Tillers, 1991) and ‘ANT’ (Wagenaar *et al.*, 1993). Both approaches stress the importance of empirical generalizations in reasoning about evidence but they differ on their precise role. While NET regards generalizations as the ‘glue’ in evidential arguments from evidence to hypotheses, ANT regards them as the ‘anchors’ of evidential stories in the available evidence.

NET takes its inspiration in Wigmore’s (1931) charting method, in which alternative arguments from evidence to hypotheses can be graphically displayed. Schum and Tillers take Wigmore’s method as the basis of their ideas on evidential visualization software. As observed by Prakken (2004) and Bex *et al.* (2003), Wigmore charts are similar to AI models of argumentation (Pollock, 1995; Loui, 1987; Vreeswijk, 1997; Prakken and Vreeswijk, 2002), and the notion of a generalization (or anchor) is essentially the same as AI’s notion of a default (Reiter, 1980). In consequence, the logical aspects of NET are now well understood. However, the relation between arguments and stories still needs to be clarified since, although NET acknowledges the importance of stories in fact-finding and evidence analysis (see, e.g. Twining, 1999), it does not incorporate this in its use of Wigmore charts.

ANT stresses that the only viable way in which crime investigation and judicial proof can proceed is by constructing alternative stories about what happened in a case, by comparing their quality as stories and by comparing how well they are ‘anchored’ in commonsense generalizations. A story can be anchored in two ways.[1] The first is *internal anchoring*: stories at least contain a sequence of events

[1] This distinction is not explicit in ANT but is made in Bex *et al.* (2006).

on a time line and stories become stronger if the connections between the events it contains are not just temporal but also causal (e.g. shooting a gun causes a sound) or intentional (a man possessing a gun who is assaulted will shoot the attacker). The second type of anchoring is *external anchoring*: elements of a story can be anchored in the available evidence by sources of information, such as observation, memory or testimony. This also involves commonsense generalizations. For instance, a witness testimony supports a belief only by virtue of the common knowledge that witnesses usually tell the truth. Clearly, the general knowledge involved in anchoring stories can have exceptions and therefore anchors must be critically examined and refined when the facts indicate a possible exception. For instance, if two witnesses know each other, they may have influenced each other's testimonies: to discard this possibility, a refined anchor may be needed, such as that if two witnesses agree but did not confer, they usually tell the truth.

Although ANT is more explicit than NET about the role of stories in reasoning about evidence, it does not give a detailed account of how stories can be connected to the available evidence. In the present paper, we aim to provide such an account in terms of the notion of argumentation as used in NET. This is motivated by the fact that, although the proponents in ANT focus on the story-based perspective in their choice both of wording and of research background, several of their central claims have a more argumentative than story-based flavour (cf. Verheij, 2000; Bex *et al.*, 2007). Especially, the role of generalizations (or anchors), exceptions to these generalizations and the dynamics of developing and refining an analysis of the evidence in a case are characteristics for an argumentative approach.

2.2 *Causation in reasoning about evidence*

Reasoning with causal information clearly is an important aspect of reasoning about evidence. Basically, it can take two forms. Using familiar AI terminology, in *prediction*, one observes or assumes a certain event and tries to predict what will happen as a consequence of this event, while in *explanation*, one observes an event or state of affairs and tries to explain how it could have been caused by other events. Both forms of reasoning are, of course, of prime importance in reasoning about evidence, whether story or argument based. Often, an attempted proof that a certain crime took place is constructed by saying that an observed fact (the evidence) holds since something else (the crime) happened which caused it. Such an explanation can then be tested by predicting what else must have been caused by the crime if it has taken place and by trying to find evidence concerning the predicted facts. In this paper, we will, as far as stories are concerned, only model explanation and leave its combination with prediction for future occasions. As remarked above, the causation involved in reasoning about evidence may be both physical (fire causes smoke) or mental (wanting to be rich makes one steal). Clearly, reasoning with causal information is defeasible in several ways: causal generalizations may have exceptions (striking a match will cause fire except if the match is wet) and observed evidence may be explained by several alternatives (the grass is wet since it rained or since the sprinkler was turned on).

2.3 *Current approaches in AI*

In AI, it is well-known that causal knowledge can be represented in two alternative ways, namely, from cause to effect (fire causes smoke) and from effect to cause (smoke means fire) (see, e.g. Pearl, 1988; Poole, 2001). Generally, a choice is made for one of these two representation methods. For instance, the well-known MYCIN medical expert system (Buchanan and Shortliffe, 1984) expressed empirical associations between symptoms and diseases from effect to cause (if symptoms then disease), while later, model-based approaches to medical diagnosis such as CASNET (Weiss *et al.*, 1978) represented the relevant knowledge from cause to effect (if disease then symptom). As for terminology, in this paper we call cause-to-effect statements *causal generalizations* and effect-to-cause statements *evidential generalizations*. With evidential generalizations, explanatory reasoning can be modus-ponens style: if the antecedent (the symptom) is known, the consequent (the disease) can be inferred. When thus alternative explanations can be derived, a choice should be made with some priority mechanism: MYCIN used certainty factors for this. With causal generalizations, prediction can also be modus-ponens style but explanation must be abductive: given the consequent (the symptom), the antecedent (the disease) is inferred since if true it would imply the symptom by modus ponens on the causal generalization. Of course, in the same way alternative explanations may be found. If so, then the best explanation must be determined either by trying to find evidence concerning predicted facts or (if gathering further evidence is impossible) with some priority mechanism on abductively inferred causes (such as a probability distribution, as in Poole, 2001).

In AI, the most popular way to model explanatory reasoning is with causal generalizations and abduction (see, e.g. Lucas, 1997; Poole, 2001; Josephson, 2001). Logics for defeasible argumentation and other non-monotonic logics, which only model modus-ponens-style reasoning, are less popular, presumably since they require the use of evidential generalizations, which arguably is cognitively less natural (Pearl, 1988). Moreover, these logics abstract from the nature of the relation between antecedent and consequent of a generalization, which creates the danger of a careless mixing of causal and evidential generalizations. Pearl (1988) showed that this may give rise to counterintuitive consequences. For instance, if we know that the smoke machine was on and that turning the smoke machine on causes smoke, we should not go on to infer from ‘there is smoke’ and the evidential generalization that smoke means fire that there is a fire. Pearl proposed his ‘c-e-constraints’ for default logics, which if respected, prevent such unwanted inferences. In this paper, we will allow for mixed reasoning with evidential and causal generalizations in a way that respects these constraints.

3. Our approach

As observed in Section 2, in AI usually one single approach to explanatory reasoning is taken. In most approaches, all causal knowledge is represented from cause to effect and the reasoning is abductive, and in some approaches all causal knowledge is represented from effect to cause and the reasoning is modus-ponens style. However, we will argue that there are reasons to combine these two approaches. We will propose a model in which stories about what happened are represented as networks of causal generalizations, while the relation between the available evidence and the events in the causal network is represented as evidential generalizations. The reasons for this approach are as follows.

A reason not to represent all causal information from effect to cause has to do with the fact that crime investigators very often draw time lines and causal-network-like structures. Since we want to build software for supporting crime investigators, we want to support this habit. This explains why for our purposes representing all causal information from effect to cause is less desirable.

However, there are also reasons not to represent all causal information from cause to effect. The first has to do with the treatment of witness testimonies. In an abductive approach as taken in, e.g. Poole (2001) and as further explained below in Section 5.2, the relation between a witness testimony and its content must be represented as causal generalizations, in which the witness testimony is regarded as caused by something else. One possible cause of a witness testimony is, of course, the truth of the event to which the witness testifies. Schematically,

g_1: $p \Rightarrow w$ `said` 'p'.

(To be truly realistic, this generalization should have some auxiliary conditions like 'the witness was interrogated' and so on, but for simplicity we will leave such conditions implicit.) However, there may be other possible causes of the witness testimony. For instance,

g_2: w `hallucinated and thought he saw` $p \Rightarrow w$ `said` 'p',
g_3: w `has reason to lie about` $p \Rightarrow w$ `said` 'p'.

As regards g_3, one reason to lie about p could be that the witness wants to protect the offender, another reason could be that by speaking the truth he would compromise himself (such as when speaking the truth would reveal a visit to the red-light district) and so on.

However, this creates the following problem. In a causal theory with the above three generalizations, the observation that w `said` 'p' can be explained in three ways: it can be explained in the 'normal' way by supposing that the witness speaks the truth, i.e. by p, but it can also be explained by supposing that w `hallucinated and thought he saw` p and by supposing that w `has reason to lie about` p. Therefore, in the approach applied here some measure of strength is needed to discriminate between them, or if further investigation is possible, further evidence should be gathered to discriminate between the two explanations. However, intuitively it seems that in the absence of further evidence for these alternative explanations, they are not worth considering. They are exceptions to the general default statement that witnesses usually speak the truth and they should therefore be assumed false as long as there is no evidence to the contrary.

Thagard (2005) speaks in this connection of a 'dual pathway model' of reasoning with testimonial evidence. He distinguishes a 'default pathway' in which people almost automatically accept a testimony and a 'reflective pathway' in which people build a model of the relevant knowledge and decide whether to believe the testimony by IBE. People shift from the default to the reflective pathway when the content of the testimony is inconsistent with their current beliefs or when there is reason to doubt the credibility of the source. The problem with a 'cause-to-effect' approach as described here is that it forces crime investigators to always take the reflective pathway since it forces to consider all alternative explanations, even if they are not supported by any further evidence. Clearly, this will in many cases induce unnecessary cognitive overload. More precisely, by leaving the evidential generalization

g_4: w said 'p' $\Rightarrow$ p

out of the problem description, this approach fails to capture that witness testimonies are usually true. Statements like g_4 express the empirical regularity that the usual cause of a testimony is the truth of its content and that any other cause is an exceptional cause.

(It should be noted that our criticism directly applies only to the logical models of abduction as described below. It may be argued that Thagard's own connectionist model of IBE does not suffer from this problem since his connections between statements have no direction.)

For these reasons, we propose a mixed approach: while the construction of stories to explain the available evidence is modelled as abductive reasoning with networks of causal generalizations, source-based reasoning about evidence is modelled as modus-ponens-style reasoning with evidential generalizations. Since these two aspects will be clearly separated, our proposal will respect Pearl's above-mentioned c-e constraints. Sources of evidence include, of course, testimonies but also, for instance, memory and perception. As for testimonies, a shift from the default to the reflective pathway will in our approach be modelled as the construction of counterarguments based on exceptions to the evidential generalization triggered by available evidence. The crucial difference with the above purely abductive approach to reasoning with testimonies is that in our argumentation approach, such counterarguments can only be constructed if there is evidence for a possible exceptional cause; in the criticized approach, such an exceptional cause also counts as a possible explanation without further evidence.

4. An example

Let us illustrate our approach with an example discussed in Crombag *et al.* (1994, pp. 460–463), a relatively simple case about an alleged burglary. (We take our analysis of the example from Bex *et al.*, 2006). The prosecution presents the following story:

> On the 18th of November, Andrew King climbs over the fence of the backyard of the Zomerdijk family with the intention to look if there is something interesting for him in the family's house. Through this yard he walks to the door that offers entry into the bedroom of the 5-year-old son of the family. The door is not closed, so King opens it and enters the bedroom to see if there is anything of interest in the house. Because it is dark, King does not see the toy lying on the floor. King hits the toy, causing it to make a sound which causes the dog to give tongue. King hears the dog and runs outside, closing the door behind him. Mr. Zomerdijk hears the toy and the dog. He goes to the bedroom and sees King running away through the closed garden door. He shouts 'there is a burglar, come and help me!' and runs into the garden after King. King, who wants to pretend he is lost, does not run away. In spite of this, Zomerdijk jumps on King and, aided by his brother, who is visiting the Zomerdijk family, molests King.

The prosecution's story is depicted in Fig 1, together with the witness testimonies on which some of the elements of the story are based.

The part of the figure within the large rounded box represents the causal network corresponding to the prosecution's story. The four small grey boxes outside the causal network are pieces of

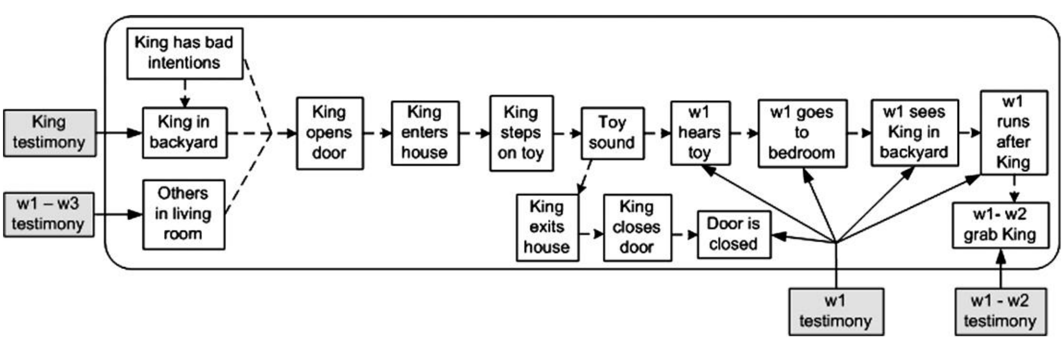

FIG. 1. The prosecution's story in the King case.

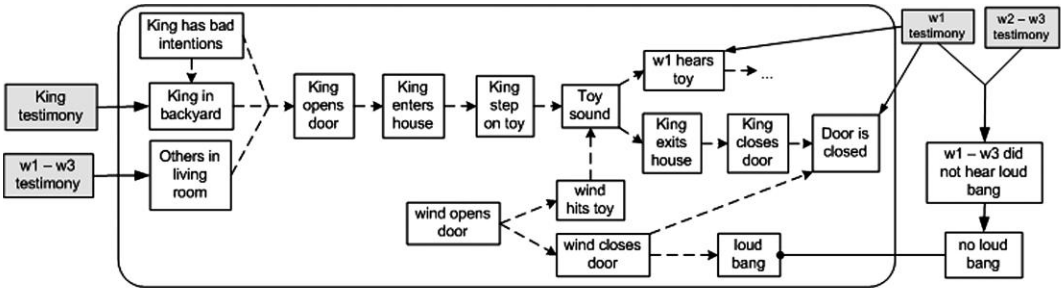

FIG. 2. The prosecution's and defence's stories in the King case.

testimonial evidence. With the evidential generalization 'a witness usually speaks the truth', they can be used to build arguments to support nodes inside the causal network. In this case, the arguments are very simple, directly supporting a node.

King has his own explanation for the fact that the toy made a sound; he claims that the wind may have blown open the door, hit the toy (which caused it to make a sound) and then blown the door shut again. This story predicts a loud bang as the consequence of the fact that the wind blew the door shut again. However, the members of the Zomerdijk family reported that they heard no loud bang.

The two stories can be depicted together in Fig. 2 (where part of the prosecution's story is left implicit for readability).

Besides new causal links within the large rounded box (capturing King's story), also a new argument from witness testimonies is added on the right. This argument in fact concludes to the negation of a network node, namely, that there was no loud bang. Note that so far none of the evidential arguments has a counterargument; below we will add such counterarguments.

Recall that we want to model how the evidence can be explained. In our approach, the explanation task is to explain the positive facts supported by evidential arguments while not contradicting the negative fact supported by evidential arguments. (When evidential arguments have counterarguments, this is restricted to those facts that are supported by arguments that survive the competition). Let us say we are interested in explaining why witnesses w_1 and w_2 grabbed (i.e. molested) King. If we focus on minimal sets of causes, then the following explanations correspond to the stories of the prosecution, respectively, King:

$$H_1 = \{\texttt{King has bad intentions}, \texttt{others in living room}\},$$
$$H_2 = \{\texttt{wind opens door}\}.$$

Hypothesis H_1 explains all the observations that we made, while H_2 does not explain that King climbed into the backyard. Also, H_2 contradicts the observation that there was no loud bang. Intuitively, this clearly makes H_1 the better explanation.

The reader may find some of the causal or evidential generalizations in this example weak or far-fetched. However, this is not a problem for our approach. The very idea of our sense-making system (which it shares with, e.g. Wigmore's charting method) is that it is the user of the system who is responsible for carefully testing the quality of his stories and arguments. The software should support the user in this critical process; it should not itself automatically generate sensible stories and arguments. Ideally, the software should also inform the user about the dangers involved in relying on stories and generalizations (cf. Twining, 1999). How this can be done is part of our research project but falls outside the scope of the present paper.

5. The formalisms

In this section, we outline our combined formal theory.

5.1 *Generalizations*

General knowledge is in our approach expressed with two sets G_{C} and G_{E} of *causal* and *evidential* generalizations. Logically, we formalize both types of generalizations in the same way, with a special conditional connective $\Rightarrow$, which only satisfies the modus-ponens inference rule:

$$g\colon p_1 \& \cdots \& p_n \Rightarrow q.$$

Here, $p_1 \cdots p_n$ and q are predicate logic literals and g, the name of the generalization, is a first-order term. Below, we will indicate whether a generalization is causal or evidential with subscripts, writing $\Rightarrow_{\mathrm{C}}$ and $\Rightarrow_{\mathrm{E}}$.[2]

5.2 *An abductive framework*

We now specify the notion of an *abductive framework*. In AI, many formal and computational accounts of abduction are available. The following simple account captures their essence and suffices for present purposes.

An *abductive framework* is a tuple $A_{\mathrm{C}} = (G_{\mathrm{C}}, O, F, X)$, where the elements of the tuple are defined as follows.

- G_{C}, the *causal theory*, is a set of causal generalizations.
- O, the *observations*, is a set of ground first-order literals. (We will explain below why O does not have to be consistent.)
- $F \subseteq O$, the *explananda*, is a consistent set of first-order literals. They are the observations which have to be explained (as further explained below, the observations not in F do not strictly have to be explained but explaining them does make an explanation better).
- X, the *explanantia*, is the set of all ground literals occurring in the antecedent of some causal generalization in G_{C} and instantiated with some term in $G_{\mathrm{C}} \cup O$.

Now, an *explanation* in terms of A_{C} is a set $H' \subseteq X$ of hypotheses such that for each $f \in F$, it holds that

- $H \vdash f$ and
- $H \nvdash \perp$.

Here $\vdash$ stands for logical implication according to the set of all deductive inference rules extended with modus ponens for $\Rightarrow$. In AI, it is usually also required that a hypothesis explains or is at least consistent with all observations in O but these conditions are unrealistic if the correctness and

[2] As usual in AI, a conditional expressed with $\Rightarrow$ which contains variables is a scheme for all its ground instances.

completeness of the causal theory are not guaranteed (cf. Prakken and Renooij, 2001), as is often the case in legal cases.

When F can be explained in more than one way, the alternative explanations have to be compared. Sometimes, e.g. in Poole (2001), this is done in terms of probability distributions over H. However, in our purely qualitative approach this method is, of course, not applicable. For now, we work with the following simple ordering on explanations. (As not uncommon in AI, we will only consider subset-minimal sets of hypotheses.) For any explanation H, the sets H_p and H_n contain the observations in O that are explained, respectively, contradicted by H.

- If $H_p \subset H'_p$ and $H_n \supseteq H'_n$, then $H < H'$.
- If $H_p \subseteq H'_p$ and $H_n \supset H'_n$, then $H < H'$.
- If $H_p = H'_p$ and $H_n = H'_n$, then $H \approx H'$.
- In all other cases, H and H' are incomparable.

In words, if H' is better than H on the observations explained and not worse on the observations contradicted, or if H' is better on the observations contradicted and not worse on the observations explained, then H' is better than H. If they are equal on both criteria, then they are equally good overall. In all other cases, they are incomparable.

Below, we will see that combining abduction with argumentation allows a refinement of this preference relation. In future research, we want to investigate whether the criteria can be further refined with content-based criteria from, for instance, the legal psychological literature on storytelling, e.g. Pennington and Hastie (1993) and Wagenaar *et al.* (1993).

Let us illustrate these definitions with our King example. The set G_C consists of all causal generalizations corresponding to a link within the large rounded box in Fig. 2. To be brief and for simplicity, we give just five of them and make some of the literals propositional:

g_1: x `has bad intentions` $\Rightarrow_C$ x `climbs into backyard`,
g_2: x `has bad intentions` & x `climbs into backyard` &
`others in living room` $\Rightarrow_C$ x `opens door`,
g_3: x `steps on toy` $\Rightarrow_C$ `toy sound`,
g_4: `toy sound` $\Rightarrow_C$ x `hears toy sound`,
g_5: `wind closes door` $\Rightarrow_C$ `loud bang`.

The set O contains all nodes that are supported by witness statements
$O = \{$
King `climbs into backyard`,
w_1 `is in living room`,
`others in living room`,
w_1 `hears toy sound`,
`door is closed`,
w_1 `goes to bedroom`,
w_1 `sees` *King* `in backyard`,
w_1 `runs after` *King*,
w_1 `and` w_2 `grab` *King*,
$\neg$ `loud bang`
$\}$.

The set F of observations that have to be explained consists of just w_1 `and` w_2 `grab` *King*.

The explanantia X consist of all ground literals that can be formed from any node with an outgoing causal link and instantiated with *King*, w_1 or w_2. To be brief, we do not list the set X.

Now, it is easy to see that the two explanations listed in Section 4 can be formally generated by the following hypotheses:

$$H_1 = \{King\ \texttt{has bad intentions}, \texttt{others in living room}\},$$
$$H_2 = \{\texttt{wind opens door}\}.$$

To make our earlier observations about their relative strength formal, we have

$$\begin{aligned}H_{1p} = \{&King\ \texttt{climbs into backyard}, w_1\ \texttt{hears toy sound},\\ &\texttt{door is closed}, w_1\ \texttt{goes to bedroom}, w_1\ \texttt{sees}\ King\ \texttt{in backyard},\\ &w_1\ \texttt{runs after}\ King, w_1\ \texttt{and}\ w_2\ \texttt{grab}\ King\},\\ H_{2p} = \{&w_1\ \texttt{hears toy sound}, \texttt{door is closed},\\ &w_1\ \texttt{goes to bedroom}, w_1\ \texttt{sees}\ King\ \texttt{in backyard},\\ &w_1\ \texttt{runs after}\ King, w_1\ \texttt{and}\ w_2\ \texttt{grab}\ King\},\end{aligned}$$

so $H_{2p} \subset H_{1p}$. Also, we have

$$H_{1n} = \emptyset,$$
$$H_{2n} = \{\neg\ \texttt{loud bang}\},$$

so $H_{1n} \subset H_{2n}$. Then, we have that $H_1 > H_2$, so our formal preference relation captures our intuitions of Section 4 that H_1 is the better explanation.

5.3 *A logic for defeasible argumentation*

We now specify the logic for reasoning with evidential generalizations. Since such generalizations allow for exceptions, this logic must be non-monotonic. Following our earlier work in Bex *et al.* (2003) and Prakken (2004), its design is very much inspired by Pollock's (1995) OSCAR system. In this system, the inference rules of classical logic are extended with defeasible inference rules that give rise to '*prima facie* reasons'. Reasoning proceeds by constructing, attacking and comparing arguments. Arguments can be constructed by chaining deductive and/or defeasible inference rules into trees. The leaves of such a tree are the input information (in our case the evidence) and the root is its conclusion. A special kind of inference rule is an *undercutter*, which says that in certain circumstances some defeasible inference rule does not apply. Accordingly, defeasible arguments can be defeated in two ways: they can be *rebutted* with an argument for the opposite conclusion and they can be *undercut* with an argument that applies the relevant undercutting reason. Arguments can be directly defeated on their final conclusion or inference step but also indirectly on intermediate ones. If appropriate, rebutting conflicts between arguments can be adjudicated with a suitable preference relation on arguments, possibly defined in terms of a preference relation on its premises.

In our case, the classical inference rules are those of standard first-order logic while the only defeasible inference rule is the modus-ponens rule for the $\Rightarrow$ connective. Undercutters to this defeasible version of modus ponens are formalized as arguments for the conclusion $\neg\ \texttt{valid}(g)$, where g is the name of the generalization to which modus ponens is applied.[3]

[3] The condition *valid*(g) is included to support reasoning about whether a generalization is acceptable at all. Strictly speaking, g is a function expression $g(t_1, \ldots, t_n)$, where $t_1, \ldots, t_n$ are all terms occurring in a generalization. Thus, a

Summarizing, we have the following.

An *evidential framework* is a tuple $A_{\mathrm{E}} = (G_{\mathrm{E}}, E)$, where

- G_{E} is a set of *evidential generalizations*;
- E, a consistent set of literals, contains the *evidence*.

An evidential framework induces an *argumentation framework* in the sense of Dung (1995), i.e. a set *Arg* of arguments ordered by a binary relation of defeat. We say that an argument A *defeats* an argument B if A rebuts or undercuts either B itself or one of its subarguments and (in case of rebutting) B is not preferred over A. The *dialectical status* of the arguments in *Arg* can be defined in various ways (Dung, 1995). For present purposes, their differences do not matter so we confine ourselves to an intuitive sketch. An argument is *justified* if it survives the competition with all its counterarguments, it is *overruled* if it is defeated by a justified argument and it is *defensible* otherwise, i.e. if at least one conflict with a counterargument cannot be resolved. It is important to note that the dialectical status of an argument depends on the interactions between all arguments in *Arg*: although arguments can make themselves justified by defeating all their counterarguments, sometimes they do not, in case they need the help of other arguments, which *reinstate* them by defeating their defeaters. Suppose, for instance, that we generally prefer police witness statements over statements of other witnesses. Then, if A is based on a police statement and B on another witness, we have that A defeats B but B may still be justified by a finding a reinstater. For example, if we find out that the policeman lied, this gives rise to an argument C that defeats A and so reinstates B.

As for the content of G_{E}, we are inspired by the notion of an *argument scheme* from argumentation theory (see, e.g. Walton, 1996). Such schemes express stereotypical patterns of defeasible reasoning. Their defeasibility is captured by a list of critical questions, which all have to be answered before the scheme can be applied. In Bex *et al.* (2003), it was argued that argument schemes can be formalized as defeasible inference rules and that negative answers to critical questions correspond to undercutters of such rules (see also Verheij, 2003; Prakken, 2004). In Bex *et al.* (2003) and Prakken (2004), it was argued that for reasoning about evidence, a combination of the witness testimony scheme (adapted from Walton, 1996) and the perception, memory and temporal-persistence schemes of Pollock (1995) comes close to David Schum's (1994) well-known decomposition of witness credibility into the aspects of veracity, objectivity and observational sensitivity. In this paper, we adopt the approach of Bex *et al.* (2003) but we slightly change its formalization: instead of formalizing the argument schemes as particular defeasible inference rules, we assume that they are present in G_{E} as generalizations and that they are applied with the modus-ponens inference rule for the $\Rightarrow$ connective. Note, finally, that G_{E} may also contain knowledge that is relevant for making inferences from the evidence or for assessing the validity of causal generalizations in G_{C}.

generalization can be invalidated for only some of its ground instances since the general name of a generalization could be $g(x)$ (for variable x), while a statement $\neg\,\mathtt{valid}(g(a))$ asserts its invalidity for a specific individual a. Below, we will leave such subtleties implicit and write names of generalizations as constants. Also, for simplicity we make no distinction between 'full' invalidity (inapplicable in all cases) and 'partial' invalidity (inapplicable in only some cases). In Bex *et al.* (2007), this distinction is made by distinguishing between the validity of and exceptions to a generalization.

5.4 *The combination*

We now describe the combination of the above abductive and evidential frameworks. The main idea is that evidential argumentation influences story-based explanation in two ways. It influences the content of the set G_C of causal generalizations with arguments about the validity of causal generalizations, and it influences the content of the set O of observations with evidential arguments about whether the observations can be derived from the evidential sources. In this paper, we illustrate both types of influences with an example. In Bex *et al.* (2007), the insights drawn from these examples are fully formalized.

To illustrate the first situation, let us assume that the prosecution wants to strengthen its story by adding a causal statement

$$g_6: \quad x \texttt{ was broke} \ \& \ x \texttt{ uses drugs} \Rightarrow_C x \texttt{ has bad intentions}.$$

Suppose the police testified that King uses drugs. Then, in explanation H_1 the assumption *King* `has bad intentions` can be replaced by *King* `was broke`. This arguably makes H_1 more convincing since the new assumption can be evidentially supported by documentary evidence. However, suppose that G_E contains the generalizations

$$\begin{aligned} g_7: &\quad g \texttt{ is based on prejudice} \Rightarrow_E \texttt{invalid}(g), \\ g_8: &\quad g \texttt{ looks at drug use} \ \& \texttt{ suspect } x \texttt{ is from} \\ &\quad \texttt{ethnic minority} \Rightarrow_E g \texttt{ is based on prejudice} \end{aligned}$$

and suppose that E contains

$$\begin{aligned} e_1: &\quad g_6 \texttt{ looks at drug use}, \\ e_2: &\quad \texttt{suspect } King \texttt{ is from ethnic minority}. \end{aligned}$$

Then, on the basis of A_E the following argument can be constructed:

1.	g_6 `looks at drug use`	(from E)
2.	`suspect` *King* `is from ethnic minority`	(from E)
3.	g_6 `is based on prejudice`	(from 1, 2 with g_8)
4.	$\neg$ `valid`(g_6)	(from 4 with g_7).

Suppose for the moment that this argument has no defeaters. Then, clearly our new version of H_1 should not count as an explanation since it uses a generalization that is declared invalid by a justified argument. So the prosecution should either return to its original version of H_1 or find another causal generalization for explaining King's bad intentions. However, things can be more subtle since the just-given argument may have defeaters and may be only defensible. (We leave it to the reader to think of plausible counterarguments.) In that case, it seems that explanation H_1 is still constructible but has lost some of its quality.

Let us now consider disputes about witness credibility, affecting the content of the observations O. Suppose G_E contains the following generalization, formalizing the witness testimony scheme for any combination of n witnesses:[4]

$$g_9: \quad \texttt{witness } w_1 \texttt{ said } `p' \ \& \cdots \& \texttt{ witness } w_n \texttt{ said } `p' \Rightarrow_E p.$$

[4] For accrual of evidence in argumentation, see Prakken (2005).

The argument that gave rise to the observation $\neg$ `loud bang` then looks as follows.

1. `witness` w_1 `said 'no loud bang'` (from E)
2. `witness` w_2 `said 'no loud bang'` (from E)
3. `witness` w_3 `said 'no loud bang'` (from E)
4. $\neg$ `loud bang` (from 1, 2, 3 with g_9).

Without defeaters, this argument is justified so its conclusion can safely be included in O. However, if this observation has no justified arguments, things are different. If it only has overruled arguments, it should not be in O, while if it at best has a defensible argument, it can be included in O but explaining it is less important than explaining a justified observation.

We now briefly illustrate this with the following generalizations in G_{E}:

g_{10} : $\neg$ `witness` w `is truthful` $\Rightarrow_{\mathrm{E}}$ $\neg$ `valid`(g_9),
g_{11} : `witness` w `has lied before` $\Rightarrow_{\mathrm{E}}$ $\neg$ `witness` w `is truthful`.

Clearly, if we add to E that

e_3 : `witness` w_1 `has lied before`,
e_4 : `witness` w_2 `has lied before`,
e_5 : `witness` w_3 `has lied before`

(for simplicity, we leave the sources of these facts implicit), then for all witnesses an argument that they are not truthful can be constructed, so the argument that there was no loud bang is undercut (as well as the versions of g_9 for one and two witnesses). In that case, this conclusion only has overruled arguments and should not be included in O.

However, suppose that the undercutting arguments can be in turn attacked, and that in the end they remain defensible. Then, the conclusion $\neg$ `loud bang` can be added to O but it becomes less important that they are explained and less serious that they are contradicted.

To conclude this section, the examples in this section were deliberately simple for ease of explanation, which might have made them rather simplistic. It should be noted that in Bex *et al.* (2003) and Prakken (2004, 2005), a number of hopefully more realistic examples of argumentation about witness testimonies are discussed.

6. Building sense-making software

As said in Section 1, the ultimate goal of the research project on which this paper reports is to design (proof-of-concept) sense-making software for crime investigation. In this section, we briefly describe the visualization component of this system as far as it has been designed so far and we illustrate it with the King case. The system (called *AVERs*) is implemented as a web front end to an SQL database. A case can be represented visually through multiple views; in this paper, we will focus on the two graphical views, i.e. the evidence view and the story view. The supplementary views (including a node view, a report view and an argument view) are, among other things, useful for reporting purposes. The software design makes extensive use of colours, which naturally cannot be shown in this paper. Therefore, colour indications have in some cases been added to the screenshots by hand. A version with colour pictures has been put online at http://www.cs.uu.nl/research/projects/evidence/publications/lpr07submitted.pdf.

It should also be noted that the visualization methods described here are meant to be used in software and therefore some pictures may be less readable when displayed in print.

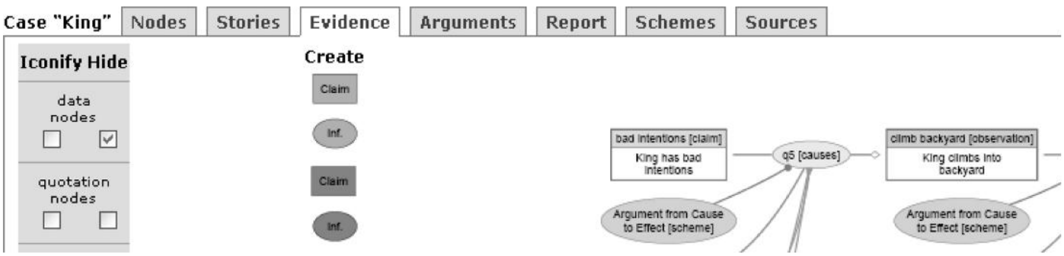

FIG. 3. Screenshot of *AVERs*' evidence tab.

TABLE 1 *Node types*

Type/polarity	Data (box)	Inference and scheme
Positive (green)	*Interpretation node*	*Inference node*
	Claim PRO main thesis	Inference from positive nodes
Negative (red)	*Interpretation node*	*Inference node*
Negative (red)	Claim CON main thesis	Inference from negative nodes
Neutral (blue)	*Quotation node*	*Scheme node*
Neutral (blue)	Quote from source document	

6.1 *The evidence tab: drawing arguments and causal networks*

The evidence tab (see Fig. 3) can be used to draw evidential arguments and causal networks. It consists of a split screen where the upper half displays a global overview of the case (the combined argument and story graph containing nodes and links) and the lower half displays the attributes of a node that is selected by the user in the upper half of the screen. New nodes can be added to the screen by clicking the desired node type. Two nodes can be connected by drawing lines from node to node. If a node is clicked in the upper half of the screen, its attributes can be edited in the lower half of the screen. Thus, a case is built.

Square-boxed nodes represent claims about a case and may be connected by directed links to represent inferential relations between claims. To link claims to the real world, some of them are coupled to external source documents from which text is selected. Such selections are made in much the same way as text in hard copy source documents is marked with highlighting marker pencils.

Nodes can be of three different types and three different polarities (see Table 1 for an overview and Fig. 4 for an example). More precisely, nodes can be of the data, inference or scheme type, and they can either be positive, negative or neutral. *Data nodes*, represented as square boxes, can in turn be of two subtypes: *interpretation nodes*, either positive or negative, represent contestable claims, while *quotation nodes* represent quotes from textual sources. *Inference nodes*, depicted as small ellipses, are justifications for inferential links. This distinction between data nodes and inference nodes is founded on the argument interchange format (Chesñevar *et al.*, 2006). Finally, *scheme nodes* represent argument schemes and justifications for inference nodes that are not supported by other inferences. Scheme nodes are depicted as blue ellipses. Currently, *AVERs* contains the argumentation schemes present in Araucaria (Reed and Rowe, 2004). Moreover, users are able to modify schemes and add their own schemes.

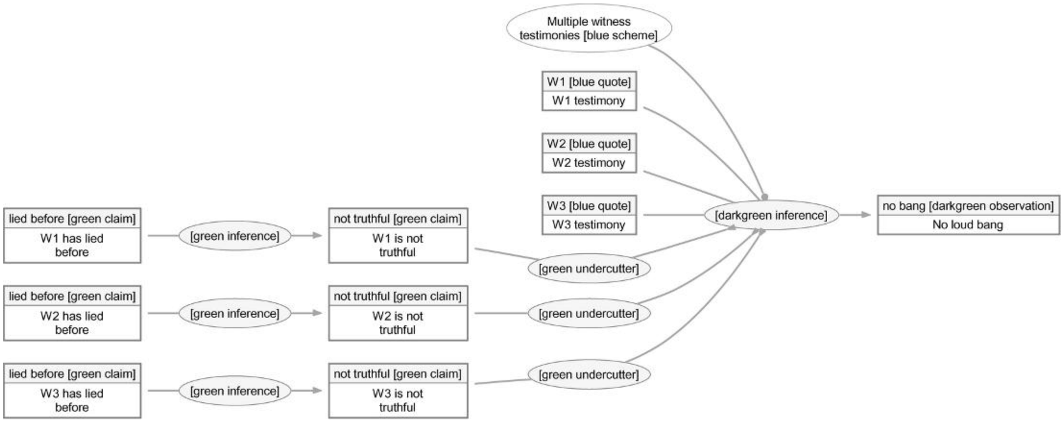

FIG. 4. An argument with counterarguments in *AVERs*.

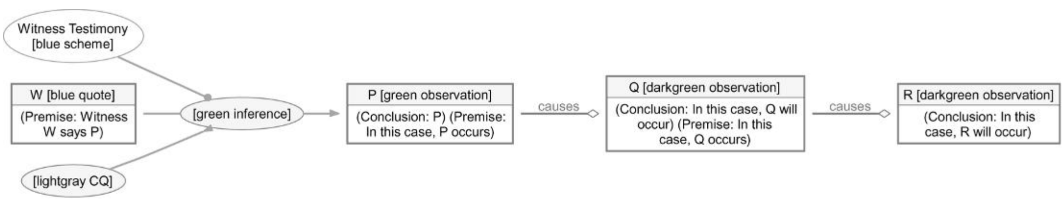

FIG. 5. A simple story in *AVERs*.

Given these node types, inferences can be expressed by connecting two or more nodes with inferential links, based either on causal or on evidential relations between states of affair. Evidential links are black with triangle-shaped arrowheads, while causal links are yellow and have diamond-shaped arrowheads (see Fig. 5). The polarities of linked nodes determine the type of the link between them: two nodes of the same polarity support each other, while two nodes of opposing polarities attack each other (the polarity of a node is indicated by its colour). Nodes can be attacked in two different ways. A *rebuttal* attacks a data node, while an *undercutter* attacks another inference node. Figure 4 contains an argument with conclusion `no loud bang` attacked by three undercutting attacks. A rebutting attack of the same argument could be drawn by creating a node `loud bang` and linking it via a `rebut` ellipse to the node `no loud bang`.

Generally, a case is built as follows (see the abstract example in Fig. 5). Starting from the evidence, for instance, a quote from a witness testimony that P occurred, a user may use the witness testimony scheme in order to infer P. While doing that, the user has to take the critical questions that are attached to this scheme into account. A negative answer to any of these invalidates the instantiation of the scheme and is therefore added as an undercutter, while a positive answer will result in a latent undercutter (as displayed in Fig. 5 with the ellipse containing '[lightgray CQ]'). The user may then use causal knowledge to connect the observed fact P to another fact, say Q, and Q to another fact R. Thus a story is built. To complete a story, the user may also add events or facts that are not yet supported by evidence. If evidence for such a proposition is found, the user may use an evidential argumentation scheme in order to connect it to evidence since schemes can expand nodes not only bottom-up but also top-down. In other words, a certain node can be marked as being (one of) the premise(s) of a scheme, but also as the conclusion. In the first case, the system automatically adds the conclusion of the scheme (together with the inferential link and the corresponding scheme node), while in the latter case it automatically adds the premise(s). Moreover, the system provides functionality to define links as being instantiations of a certain scheme. If two nodes are connected by dragging the mouse, the chosen scheme is added automatically.

Based on inferential connections, *AVERs* is able to evaluate the dialectical status of nodes, i.e. whether they are supported by a justified argument or not, based on algorithms described in Vreeswijk (2006). For example, in Fig. 4 the node `no loud bang` is defeated by three undercutters. In the software, both this node and the inference node on which it is based will therefore be coloured red.

6.2 *The story tab: investigating explanations*

The story tab displays a selection of the nodes and links that were already added by the user using the evidence tab. Put more precisely, it displays a graph that contains all and only those nodes that are part of a causal network, i.e. nodes that are linked to one or more other nodes through causal connections. The nodes that are part of the set O, which are those that are supported by a justified or defensible argument, are displayed in a different colour, and the nodes in F are in addition displayed as encapsulated boxes.

The support the system should provide to the user in the story tab is twofold (currently these features have not yet been implemented). Firstly, it should be able to determine, given a set H of nodes marked by the user as hypotheses and the set of causal generalizations G_c drawn by the user, which of the observations in O are explained by $H \cup G_c$. When applicable, it should also tell which overruled facts are explained. Secondly, and reasoning the other way around, the system should be able to infer the possible explanations for the observed facts O that satisfy certain criteria. For instance, it could be told to infer minimal explanations of F that only consist of initial causes, or it could be told to infer a minimal connected part of the causal network that starts with initial causes and ends with all elements of F.

6.3 *The King case visualized in AVERs*

We now visualize some aspects of the King case in *AVERs*. In doing so, we assume that the two stories were gradually built and refined by the police investigators.

Let us suppose that in this example case, the investigation started with the observation that King was molested by Mr Zomerdijk and his brother in their backyard (see Fig. 6, node w_1 `and` w_2 `grab` *King*). Subsequently, the police might have considered two possible explanations:

- King was trying to burgle the Zomerdijks and was caught by Mr Zomerdijk, who acted in self-defence, or

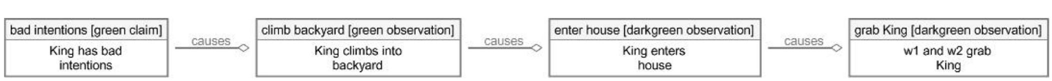

FIG. 6. King case: preliminary story.

- the Zomerdijk family saw King in their backyard because the toy made a sound, and molested him, apparently for no reason.

The police opted for the first story, supported by the testimonies of the Zomerdijks. So as an explanation for the fact that King was grabbed by Mr Zomerdijk in his backyard, they assumed that King had bad intentions, i.e. he wanted to break into the Zomerdijks' house, and climbed into their backyard in order to enter the house. Their preliminary story looks like Fig. 6.

Assume next that based on testimonies of the Zomerdijk family and the hypothesized events, the police refined their story. For example, based on their causal assumption that King entering the house caused him to step on the toy, they added a causal link from *King* `enters house` and predicted the node *King* `steps on toy`, which in turn led to the toy making sound, and so on. The complete prosecution's story may be represented in *AVERs* as displayed in Fig. 7 (for readability purposes, the causal network is split into two parts at node `Toy makes sound`).

We next describe how stories can be anchored in evidence. For this purpose, witness testimonies are used. Consider, e.g. witnesses 1, 2 and 3 who declared that they were in the living room. By selecting text from the original source documents, quotation nodes can be added to the graph. These evidence nodes can now be connected to the fact in the story by selecting the witness testimony scheme for multiple witnesses (see Fig. 8). Applying argumentation schemes to all testimonies results in an anchored story as displayed in Fig. 9 (also split at node `Toy makes sound`). In this figure, all generalizations, inference nodes and scheme nodes are hidden for readability purposes. This illustrates how *AVERs* allows to collapse certain nodes based on their type or status (these nodes

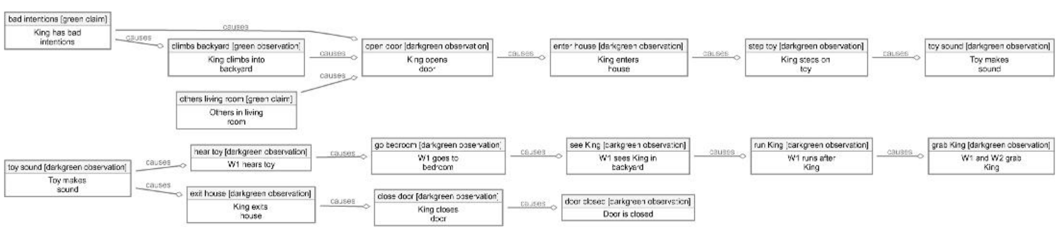

FIG. 7. King case: prosecution's complete story.

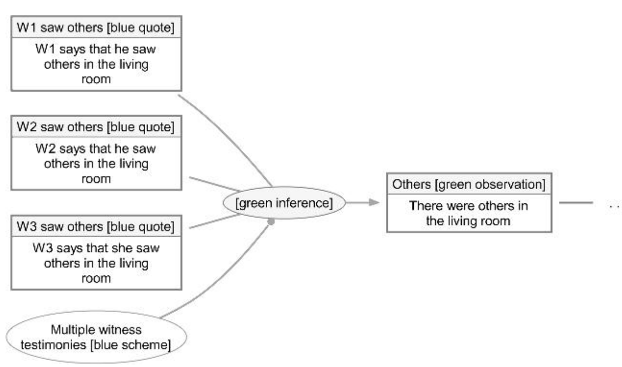

FIG. 8. King case: observation supported by witness testimonies.

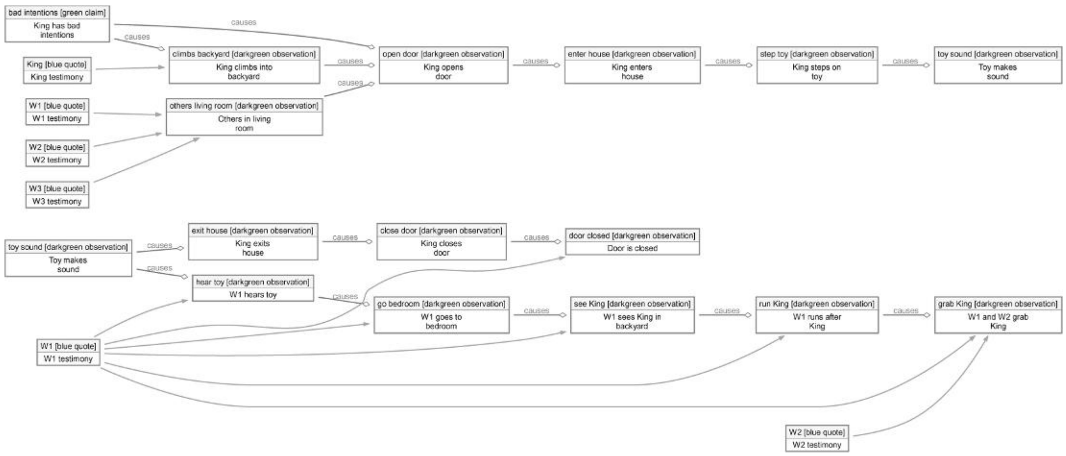

FIG. 9. King case: prosecution's story supported by evidence.

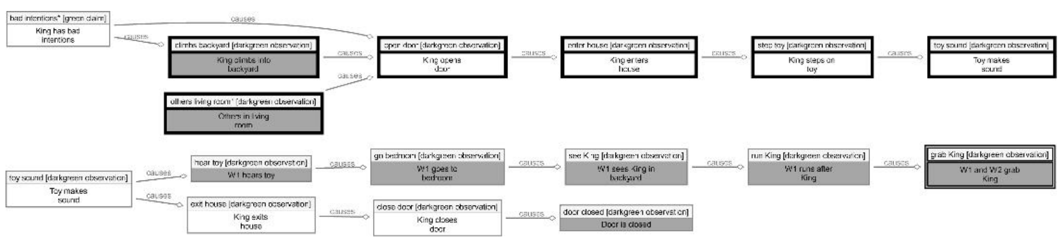

FIG. 10. King case: prosecution's story with highlighted observations.

are depicted in smaller form without text) or to hide them completely, so that the user maintains overview of complex graphs.

Now that the prosecution has completed their story, it has to be checked whether the observations can be explained by it. For this purpose, the story tab must be used. This tab displays the prosecution's story and colours the facts or events that are supported by justified or defensible arguments gray (see Fig. 10). These nodes are part of the set of observations O. This set thus contains *King* `climbs into backyard`, `Others in livingroom`, w_1 `hears toy`, w_1 `goes to bedroom`, W_1 `sees` *King* `in backyard`, w_1 `runs after` *King*, w_1 `and` w_2 `grab` *King* and `Door is closed`. The node w_1 `and` w_2 `grab` *King* is selected as the only member of F (displayed with encapsulated boxes). The hypothesis H should explain as many observations from O as possible but at least explain w_1 `and` w_2 `grab` *King*. The prosecution's hypothesis (H_1) is as follows: King had bad intentions and wanted to burgle the Zomerdijk family. The others were in the living room, so he had the opportunity to do so. H_1 thus contains the nodes *King* `has bad intentions` and `Others in livingroom`. After selecting the nodes that are part of the hypothesis (marked by an asterisk in Fig. 11), the system deduces which observations are explained by it and marks them by adding a black box.

As displayed in Fig. 11, the prosecution's hypothesis explains all observations in O and more. After all, this figure shows how all observations (coloured gray) are surrounded by a black box.

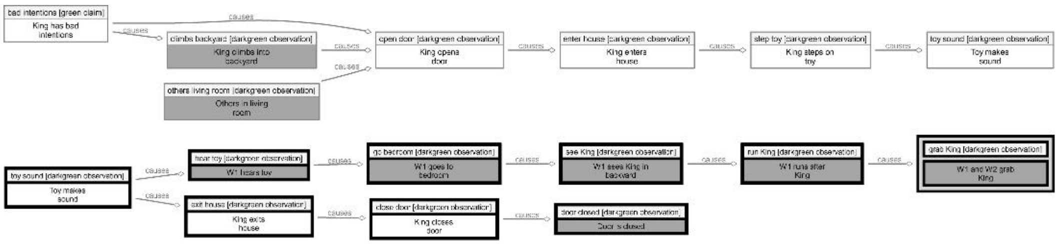

FIG. 11. King case: prosecution's story with explained observations.

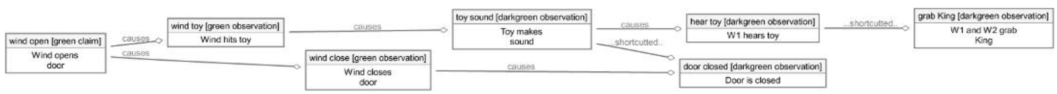

FIG. 12. King case: King's story.

King offers an alternative story, displayed in Fig. 12 (to keep the figure readable, some paths are compressed to single links). However, when taking into account that generally when a door is blown shut, there is a loud bang, the story has to be updated (see Fig. 13, which also contains the relevant argument from witness statements that support King's story). Unfortunately for King, none of the witnesses heard a loud bang, so assuming that all witnesses told the truth there was no loud bang. Figure 13 therefore contains an attack link from `no loud bang` to `loud bang`. If we combine the prosecution's story with King's explanation, we obtain Fig. 14. (Because of space limitations, several causal paths are in this figure compressed to links.) The prosecution's hypothesis still explains all observations since King's alternative did not add any new observations to O. Now, let us determine whether King's alternative is equally good. King's hypothesis (H_2) is that the wind opened the door. H_2 thus only contains `Wind opens door` (marked by an asterisk). The story tab will now display which observations are explained (see Fig. 15). There are two major problems with King's story. First of all, it does not explain why he was in the backyard. This means that King

- must provide an alternative explanation or
- has to admit that he had bad intentions.

Secondly, H_2 explains that a loud bang must have occurred. However, this conflicts with the witness testimonies and is therefore defeated by a justified argument. So, using the figures displayed in the story tab of *AVERs* (see Figs 11 and 15), it is fairly easy to see that H_1 is better than H_2 because it explains more (more observations are surrounded by a black box) and does not explain defeated propositions (no red observations are marked by a black box).

Let us finally discuss how the two invalidity arguments of Section 5.4 can be added to these figures. Figure 16 displays the argument attacking the causal link that was supposedly added by the prosecution to explain why King had bad intentions, while Fig. 4 above in fact contains the argument attacking the presence of the node `no loud bang` in the story part of the graphs.

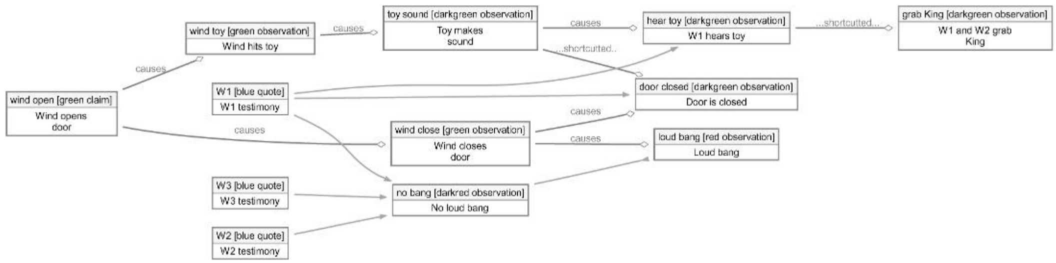

FIG. 13. King case: King's story supported by evidence.

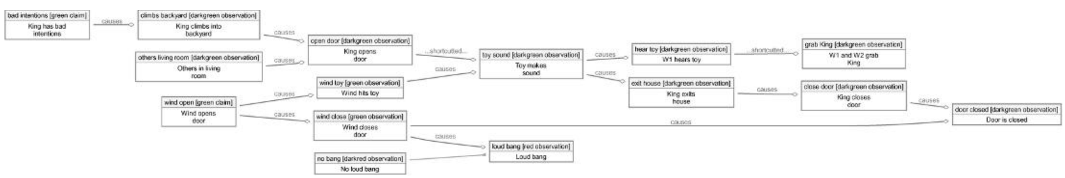

FIG. 14. King case: the stories combined.

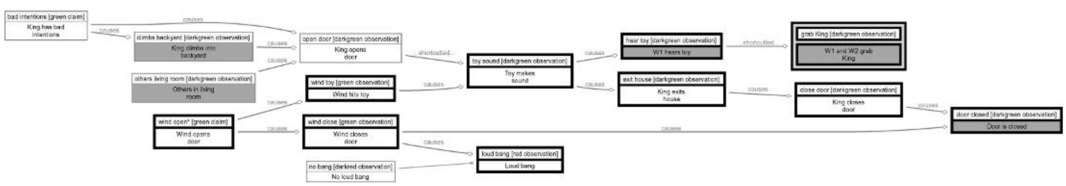

FIG. 15. King case: combined stories with explained observations.

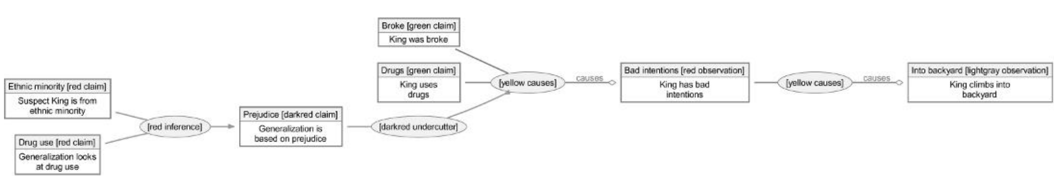

FIG. 16. King case: a causal link attacked.

6.4 *State of implementation*

The *AVERs* system as described in this section has not yet been fully implemented. Functionality that has been implemented includes the ability to create nodes, the ability to create and apply inference schemes and the possibility to compute the dialectical status of individual nodes. In short, the evidence tab as described above has been fully implemented. *AVERs* currently lacks functionality to perform explanatory or abductive reasoning. This aspect, i.e. the story tab, is currently being developed and implemented.

7. Conclusion

This article has reported on a research project on the development of sense-making software for crime investigation. We have discussed how the causal aspects of evidential storytelling can best be modelled and visualized and linked to the available evidence. We claim to have made two main contributions. Firstly, we have argued that in the context of sense-making systems for crime investigation there are reasons to combine two AI approaches, abductive explanatory reasoning and modus-ponens-style default reasoning, which are usually considered as irreconcilable alternatives. We have outlined how this proposal can be formalized. From a legal theoretical perspective, this amounts to an attempt to unify the ANT and the NET. Secondly, we have described the current design of a visualization software tool in which causal networks can be linked to the available evidence with modus-ponens-style argumentation structures. This tool is still under construction. At later stages of the project, it will be empirically tested in user experiments to see whether using such a tool indeed has the benefits it is often claimed to have. Among other things, this should bring clarity on whether a combination of story-based and argument-based reasoning is indeed natural to crime investigators.

Funding

The Netherlands Organisation for Scientific Research (NWO) 634.000.429.

References

HOLMES 2: Home office large major enquiry system (www page). http://www.holmes2.com/holmes2/index.php, 2006.

i2 Analyst's Notebook: Investigative analysis software (www page). http://www.i2.co.uk/Products/Analysts_Notebook/default.asp, 2006.

T.J. ANDERSON AND W. TWINING. *Analysis of Evidence. How to Do Things with Facts Based on Wigmore's Science of Judicial Proof*. Little, Brown and Company, Boston, MA, 1991.

F.J. BEX, H. PRAKKEN, C. REED, AND D.N. WALTON. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, **12**:125–165, 2003.

F.J. BEX, H. PRAKKEN, AND B. VERHEIJ. Anchored narratives in reasoning about evidence. In T. M. van Engers, editor, *Legal Knowledge and Information Systems. JURIX 2006: The Nineteenth Annual Conference*, pages 11–20. IOS Press, Amsterdam etc., 2006.

F.J. BEX, H. PRAKKEN, AND B. VERHEIJ. Formalising argumentative story-based analysis of evidence. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pages 1–10. ACM Press, New York, 2007.

B.G. BUCHANAN AND H.E. SHORTLIFFE (eds.) *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

C. I. CHESÑEVAR, J. MCGINNIS, S. MODGIL, I. RAHWAN, C. A. REED, G. SIMARI, M. SOUTH, G. A.W. VREESWIJK, AND S. WILLMOTT. Towards an argument interchange format. *The Knowledge Engineering Review*, **21**(4): 293–316, 2006.

H.F.M. CROMBAG, P.J. VAN KOPPEN, AND W.A. WAGENAAR. *Dubieuze zaken: De psychologie van strafrechtelijk bewijs*. Uitgeverij Contact, Amsterdam, second edition, 1994 (in Dutch).

P.M. DUNG. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*-person games. *Artificial Intelligence*, **77**:321–357, 1995.

J.R. JOSEPHSON. On the proof dynamics of inference to the best explanation. *Cardozo Law Review*, **22**: 1621–1643, 2001.

R.P. LOUI. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, **2**:100–106, 1987.

P. LUCAS. Symbolic diagnosis and its formalisation. *The Knowledge Engineering Review*, **12**:109–146, 1997.

J. PEARL. Embracing causality in default reasoning. *Artificial Intelligence*, **35**:259–271, 1988.

N. PENNINGTON AND R. HASTIE. The story model for juror decision making. In N. Hastie, editor, *Inside the Juror, The Psychology of Juror Decision Making*. Cambridge University Press, Cambridge, 1993.

J.L. POLLOCK. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.

D.L. POOLE. Logical argumentation, abduction and Bayesian decision theory: a Bayesian approach to logical arguments and its application to legal evidential reasoning. *Cardozo Law Review*, **22**:1733–1745, 2001.

C.J. POOT, R.J. BOKHORST, P.J. VAN KOPPEN, AND E.R. MULLER. *Rechercheportret: Over Dilemma's in de Opsporing*. Kluwer, Alphen aan den Rijn, 2004. In Dutch.

H. PRAKKEN. Analysing reasoning about evidence with formal models of argumentation. *Law, Probability and Risk*, **3**:33–50, 2004.

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

H. PRAKKEN AND S. RENOOIJ. Reconstructing causal reasoning about evidence: a case study. In *B. Verheij, A.R. Lodder, R.P. Loui, A.J. Muntjewerff. Legal Knowledge and Information Systems. JURIX 2001: The Fourteenth Annual Conference*, pages 131–142, Amsterdam etc, 2001. IOS Press.

H. PRAKKEN AND G.A.W. VREESWIJK. Logics for defeasible argumentation. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht/Boston/London, second edition, 2002.

C. A. REED AND G.W.A. ROWE. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of Artificial Intelligence Tools*, **14**(3–4):961–980, 2004.

R. REITER. A logic for default reasoning. *Artificial Intelligence*, **13**:81–132, 1980.

J. VAN DER SCHOOR. Brains voor de recherche. *Justitiële Verkenningen*, **30**:96–99, 2004. In Dutch.

D. SCHUM AND P. TILLERS. Marshaling evidence for adversary litigation. *Cardozo Law Review*, **13**:657–704, 1991.

D.A. SCHUM. *Evidential Foundations of Probabilistic Reasoning*. Northwestern University Press, Evanston. Illinois, 1994.

S.M. WEISS, C.A. KULIKOWSKI, AND S. AMARAL. A model-based method for computer-aided medical decision-making. *Artificial Intelligence*, **11**:145–172, 1978.

P. THAGARD. Testimony, credibility and explanatory coherence. *Erkenntnis*, **63**:295–316, 2005.

W. TWINING. Necessary but dangerous? Generalisations and narrative in argumentation about "facts" in criminal process. In M. Malsch and F. Nijboer, editors, *Complex Cases. Perspectives on the Netherlands Criminal Justice System*, pages 69–98. Thela Thesis, Amsterdam, 1999.

B. VERHEIJ. Dialectical argumentation as a heuristic for courtroom decision-making. In P.J. van Koppen and N.H.M. Roos, editors, *Rationality, Information and Progress in Law and Psychology. Liber Amicorum Hans F. Crombag*, pages 203–226. Metajuridica Publications, Maastricht, 2000.

B. VERHEIJ. Dialectical argumentation with argumentation schemes: an approach to legal logic. *Artificial Intelligence and Law*, **11**:167–195, 2003.

G.A.W. VREESWIJK. Abstract argumentation systems. *Artificial Intelligence*, **90**:225–279, 1997.

G.A.W. VREESWIJK. An algorithm to compute minimally grounded and admissible defence sets in argument systems. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA-06*, pages 109–120. IOS Press, Amsterdam etc, 2006.

W.A. WAGENAAR, P.J. VAN KOPPEN, AND H.F.M. CROMBAG. (1993) *Anchored Narratives. The Psychology of Criminal Evidence.* St. Martins Press, New York, 1993.

D.N. WALTON. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

J.H. WIGMORE'S. *The Principles of Judicial Proof*. Little, Brown and Company, Boston, 2nd edition, 1931.