Analysing reasoning about evidence with formal models of argumentation[*]

Henry Prakken

Institute of Information and Computing Sciences, Utrecht University
PO Box 80 089, 3508 TB Utrecht, The Netherlands
Email: henry@cs.uu.nl

This paper is on the formal modelling of reasoning about evidence. The main purpose is to advocate logical approaches as a worthwhile alternative to approaches rooted in probability theory. In particular, the use of logics for defeasible argumentation is investigated. Such logics model reasoning as the construction and comparison of arguments for and against a conclusion; this makes them very suitable for capturing the adversarial aspects that are so typical for legal evidential reasoning. Also, it will be shown that they facilitate the explicit modelling of different kinds of knowledge, such as the distinction between direct vs. ancillary evidence, and the explicit modelling of different types of evidential arguments, such as appeals to witness or expert opinion, applying generalisations, or temporal projections.

Keywords: evidential reasoning, logic, defeasible reasoning, argumentation, generalisations

1 Introduction

This paper is on the formal modelling of reasoning about evidence. The main purpose is to advocate logical approaches as a worthwhile alternative to approaches rooted in probability theory. In particular, I will discuss the use of logics for defeasible argumentation. Such logics, which are a result of artificial intelligence research on so-called nonmonotonic reasoning, model reasoning as the construction and comparison of arguments for and against a conclusion; this makes them very suitable for capturing the adversarial aspects that are so typical for legal reasoning. Also, I will argue that they facilitate the explicit modelling of different kinds of knowledge, such as the distinction between direct vs. ancillary evidence, and the explicit modelling of different types of evidential arguments, such as appeals to witness or expert opinion, applying generalisations, or temporal projections.

I focus on formal models since my main research interest is in designing computer programs that support or perform evidential reasoning. This choice of standpoint allows me to avoid the debate on the use of formal methods in evidence theory. However, I still hope that my discussion will also be relevant for those outside computer science or artificial intelligence who are interested in the use of such methods.

I am especially interested in two types of computer programs: 'knowledge-based' systems, and 'sense-making' systems. Knowledge-based systems have two main components, a knowledge base and an inference engine. The knowledge base contains knowledge about a certain problem domain, formulated in a language suitable for computer manipulation, and the inference engine reasons with this knowledge in order to solve a certain concrete problem, or at least to suggest alternative solutions to it. Ideally, the reasoning of the inference engine conforms to the meaning of the knowledge representation language, which, again ideally, is defined in terms of logic and/or probability theory. Sense-making systems (see e.g. Kirschner et al., 2002) do not themselves reason to solve a problem. Instead, the goal of such software is to support humans in making sense of a problem. In particular, they provide tools for structuring (usually visualising) the problem and the user's reasoning in solving it. Often they also provide tools for manipulating these structures, e.g. by converting one visualisation into another, by combining pieces of information, or even by performing logical or probabilistic computations on the user's input. In addition, some sense-making systems also support the communication between different people working on the same problem. An example of sense-making software for evidential reasoning is Tiller and

---

[*] This paper is partly based on work done jointly with Chris Reed and Douglas Walton, reported in Prakken et al. (2003).

Schum's MarshalPlan project (Schum & Tillers, 1991), an early, pre-world-wide-web Hypertext application that supports preliminary fact investigations. The main difference between knowledge-based and sense-making systems is that the latter have no knowledge base, i.e., no collection of permanently stored general knowledge about a certain domain. This means that for building sense-making systems, unlike for knowledge-based systems, no laborious and difficult knowledge-acquisition phase is necessary. On the other hand, both types of systems rely on a theory of reasoning, and this is why formal models of reasoning are equally relevant for both types of systems.

What then is a good basis for a formal theory of evidential reasoning? An obvious candidate is probability theory since, after all, almost all evidential reasoning is reasoning with uncertainty. It seems particularly attractive to use so-called probabilistic networks, a result of recent artificial intelligence research on reasoning with uncertainty, since such networks elegantly capture conditional dependencies in a graph structure. The nodes of a probabilistic network stand for statistical variables (e.g. in a one-car accident 'absence or presence of skid marks on the road', 'whether or not the driver speeded', 'the position of the handbrake' 'whether or not the passenger pulled the handbrake'). The links between the nodes express probabilistic dependencies between the values of such variables (for instance, 'speeding causes skid marks with 85% probability'). If these dependencies are quantified as numerical probabilities, and if prior probabilities are assigned to the node values (assigning probability 1 to the node values that represent the available evidence), then the conditional probability concerning certain nodes of interest (e.g. 'the driver speeded' or 'the passenger pulled the handbrake') given a body of evidence (modelled by setting the corresponding node values to 1) can be calculated according to the laws of probability theory, including Bayes' rule. Theoretically, the use of such methods is very attractive, since probability theory is the standard mathematical theory concerning notions of plausibility and uncertainty.

However, applying probabilistic techniques in practice is often problematic. An obvious reason for this is that they require numbers as input, and in the fast majority of legal cases reliable numbers are very hard to obtain, either because there are no reliable statistics, or because legal experts are unable or reluctant to provide numerical estimates. For some purposes this may still not be a problem. For instance, Kadane & Schum (1996) use probabilistic networks to perform sensitivity analysis: by comparing the effect of different estimates of probability distributions, they aim to discover which variables are most relevant to the outcome of a problem. However, things are different when a knowledge-based system has the task of computing an accurate solution to an evidential reasoning problem. Moreover, probabilistic networks have other limitations, which are especially problematic for sense-making systems. Essentially, probabilistic networks compile all available knowledge into a probability distribution over certain variables of interest, and thus they conceal some important distinctions of ordinary evidential reasoning. The first is that legal disputes often consist of an exchange of explicit arguments and counterarguments, stated by opposing sides of the dispute. Moreover, probabilistic networks blur the distinction between directly relevant and ancillary evidence. (Ancillary evidence is evidence that has a bearing on the probative force of directly relevant evidence. A typical example is information on the credibility of a witness.) Schum (2001, p. 1948) remarks that the need for ancillary evidence especially arises when the conditional probabilities cannot be established by statistical relative frequencies. Moreover, even when such statistics are available, one may still need ancillary evidence, since lawyers often try to undermine the use of statistics by their adversary.

These, then, are my main reasons for exploring an alternative formal modelling tool, viz. logical systems for defeasible argumentation. Being logical systems, they can deal naturally with ancillary evidence, by using certain logical representation techniques. And being argumentation systems, they make the sources of conflict explicit in the notion of attack between arguments; moreover, they support the explicit modelling of stereotypical forms of evidential reasoning. The purpose of this paper is to explain and illustrate how argumentation logics can perform these tasks. The system of choice for my analysis will be that of John Pollock (1987, 1995,1998), since this system models epistemological reasoning, and evidential reasoning is essentially of this kind.

The rest of this paper is organised as follows. I will first introduce a small example case in Section 2, to be used as the running example of this paper. This will be followed by an overview of the idea of logical argumentation systems in Section 3. Then I discuss in some detail how reasoning with generalisations (Section 4) and with argumentation schemes (Section 5) can be captured in an adapted version of Pollock's argumentation system. In Section 6 I illustrate my analysis with the example of Section 2, after which I conclude.

2 An example case

As the running example of this paper I will use a case that was earlier analysed by Wigmore (1931) with his well-known charting method. The case (Commonwealth v. Umilian, 1901, Supreme Judicial Court of Massachusetts, 177 Mass. 582) is a murder case in which a farm labourer Umilian (U) was accused of having murdered his colleague Jedrusik (J), whose headless body was found 500 feet from the barn. U had a motive, since J had tried to prevent U to marry the maid working on the farm: he had sent a letter to the priest in charge of the wedding ceremony, writing that J had a wife and children in England. For this reason, the pries refused to marry U and the woman, until he found out that the content of the letter was false. Although the marriage was then performed, U still showed that he was very angry with J, and made threats of vengeance against him. Some time later J was found dead, and evidence suggested that U and J were isolated in the area of the barn around the time of the murder.

3 Logics for defeasible argumentation

Logics for defeasible argumentation, or argumentation systems for short, are an example of a nonmonotonic logic. In this section I will first briefly explain the idea of nonmonotonic logic and why it is relevant for evidential reasoning, and then discuss the main elements of argumentation systems.

3.1 Nonmonotonic logic

Most by now classical systems of nonmonotonic logic, such as default logic, circumscription and autoepistemic logic, were developed around in the late seventies and early eighties (see e.g. Ginsberg 1987 for an overview and reprints of many classical papers). Essentially, nonmonotonic logics are a result of the so-called logicist approach to building intelligent robots. Stripped to the bones, the idea of logicism is to feed a robot with logical formulas that present general common-sense knowledge about the world, and with other logical formulas that express the robot's observations, and to let the robot plan its actions by applying the rules of logic to these formulas. It was soon realised that standard logic is not sufficient for this purpose, since common-sense knowledge has a high rule-of-thumb nature, with lots of conflicting rules and subject to lots of exceptions. A classic example is the rule-of-thumb (or 'default') that birds usually fly. Although this rule has many exceptions (penguins, ostriches, birds with their feet set in concrete, ...) commonsense reasoners tend to 'boldly' apply such rules to a given bird without first verifying that the bird is in no way exceptional. Yet when using standard logic one has to give an explicit list of all exceptions, and a rule can only be applied if all these exceptions are known to be absent. What is needed then to model commonsense reasoning is a theory of 'quick-and-dirty' reasoning, where one applies a rule of thumb if nothing is known about exceptions, but one is prepared to retract a conclusion if further knowledge tells us that there is an exception. Nonmonotonic logics are meant to be such theories of 'quick-and-dirty' reasoning.

The prima facie relevance of nonmonotonic logics for our purposes should be obvious, since a key role in evidential reasoning is played by empirical generalisations. For example, in the Umilian case "If x falsely tries to prevent y's marriage, a revengeful murderous emotion from y towards x tends to be created" is implicit. According to Schum (1994) such generalisations, which lawyers usually leave implicit, are the "glue" which holds evidential arguments together. See also Anderson (1999) and Twining (1999). And in the "anchored narratives" theory of legal evidence proposed by forensic psychologists, e.g. Wagenaar et al. (1993), generalisations are essential as the "anchors" grounding a narrative about what happened in the available evidence. Now virtually all such generalisations allow for exceptions, which makes evidential reasoning nonmonotonic. Although quite some work has already been done on applying nonmonotonic logics to legal reasoning (for a recent overview see Prakken & Sartor, 2002), virtually all this work is on reasoning about the law and largely ignores evidential reasoning. In fact, most current AI & Law applications of nonmonotonic logic are to reasoning about the interpretation of legal concepts and to capturing the defeasibility of legal rules, i.e.,

modelling the fact that legal rules are subject to exceptions on the basis of principles, values or the purpose of the rule (see e.g. Prakken & Sartor, 2002 for an overview). A notable exception is Verheij (2000) who, in discussing the anchored narratives theory, notes the defeasible nature of most anchors, and proposes that the critical testing of anchors can be modelled as defeasible argumentation. This paper aims to further develop Verheij's observations, proposing to formalise evidential reasoning within Pollock's argumentation system, with special attention to the various ways in which generalisations can be attacked and to some other stereotypical patterns of evidential argumentation.

First, however, some objections to the use of nonmonotonic logics have to be discussed. Although nonmonotonic logic is still an active field of AI research, its usefulness has been disputed on several grounds. The first is that the merits of the logicist approach to building intelligent robots have been heavily disputed (see e.g. Brooks 1991). However, we need not be concerned with this dispute, since our aims are different: we do not want to build intelligent robots, but we have the much more modest aim of building software for solving or structuring reasoning problems. A for present purposes more serious objection to the use of nonmonotonic logics concerns the so-called knowledge-acquisition bottleneck, already briefly alluded to in the introduction. This objection in fact concerns any knowledge-based attempt to build a problem-solving program. In many domains it has as yet proven too hard to scale systems up to realistic size: in particular, many problems not only require high-level knowledge about the particular problem area but also "low-level" commonsense knowledge about the world. The problem of representing a sufficient amount of commonsense knowledge to be able to solve nontrivial problems seems as yet unsolved (although some, such as Lenat (1998) are confident that it will be solved soon). In the legal domain, experiences with knowledge-based systems vary. They have proven very successful for processing legislation, especially in public administration. In this context, some main benefits of knowledge-based systems are that they provide easy and complete access to large amounts of legislation and that they allow the user to investigate the legal consequences of her particular problem (see e.g. Van Engers et al., 2001). However, applications to evidential reasoning are still rare. A major reason for this is that for modelling evidential reasoning a very large body of commonsense knowledge is needed but which is very hard to obtain and represent. While legislation is relatively easy to identify and formalise, the factual commonsense knowledge needed in evidential reasoning is extremely diverse, vague and uncertain (think alone of the commonsense generalisations that evidential arguments rely on), so that building reliable systems of realistic size for evidential reasoning seems a formidable task. Nevertheless, this problem only affects knowledge-based systems; as said in the introduction, sense-making systems do not require a knowledge base, and consequently for such systems a logical account of evidential reasoning is not only theoretically but also practically relevant.

3.2. Argumentation systems: general idea

While most major systems of nonmonotonic logic were developed around 1980, the first logical argumentation systems were developed somewhat later (e.g. Loui, 1987; Pollock, 1987). However, unlike most other research in this field, much research on argumentation systems draws its inspiration from earlier philosophical work on epistemology, such as from Pollock (1974) and Rescher (1977). Argumentation systems formalise nonmonotonic reasoning in terms of the dialectical interaction between arguments and counterarguments. They tell us how arguments can be constructed, when arguments are in conflict, how conflicting arguments can be compared, and which arguments survive the competition between all conflicting arguments. I will now briefly explain each of these elements in turn, being somewhat biased towards my favourite approaches (not all systems proposed in the literature fully adhere to my picture; see Prakken & Vreeswijk, 2002 for an overview of the various systems).

As for constructing arguments, the basic idea is the same as in standard logic: one constructs arguments by applying inference rules to a set of premises. What are the allowed inferences of defeasible reasoning? Clearly, they should include those of deductive reasoning, since sometimes commonsense reasoning is deductive: for instance, the statements 'the murder was committed near the barn' and 'the suspect was in the barn around the time of the murder' deductively imply that 'the suspect was near the murder scene around the time of the murder'. But in addition we need an inference rule for applying rules of thumb: from the statements `If P then usually Q' and `P' we should be able to defeasibly infer Q (thereby implicitly assuming that the P case is a usual P case). Most nonmonotonic logics leave it with that: the only inference rule they add to standard logic is a rule for

applying default generalisations. However, as noted in the introduction, Pollock's work shows that a richer theory of defeasible inference rules is possible.

What about conflicting arguments? When an argument is deductive, the only possible attack is on its premises. However, a defeasible argument can be attacked even if all its premises are accepted. Consider, for instance, an argument "The suspect was at the murder scene at the time of the murder since witness John says so" (applying a rule "if a witness says P, then usually P"). One way to attack it is to *rebut* it, i.e., to state an argument with an incompatible conclusion. For instance: "The suspect was not at the murder scene since witness Bob says he was with him in the pub at the time of the murder, and one cannot be in two places at the same time" (applying the same rule). A second way to attack the argument is to *undercut* it, i.e., to argue that in this case the premises do not support its conclusion. For instance: "John is a friend of the suspect, so his testimony is unreliable". Note that both rebutting and undercutting attack have a direct and an indirect version; indirect attack is directed against an intermediate conclusion or inference step of an argument. For instance, indirect rebuttals contradict an intermediate conclusion of an argument.

Rebutting arguments must be compared on their relative strength, to determine which argument *defeats* the other. When an argument *A* is stronger than a rebutting argument *B*, I will say that *A strictly* defeats *B*, when they are equally strong, I will say that they both defeat each other (an undercutter always strictly defeats its target). What are good standards for assessing the strength of arguments? In general it depends on the nature of the problem and the domain. For evidential arguments it will often involve probabilistic assessments. Consider our above rebutting arguments based on witnesses John and Bob: if Bob is an adult and John a small child, one might say that it is more likely that John speaks the truth than Bob.

The notion of defeat only tells us something about the relative strength of two individual conflicting arguments; it does not yet tell us with what arguments a dispute can be won. All available arguments should be classified into three kinds: the *justified* arguments (those that survive the competition with their counterarguments), the *overruled* arguments (those that lose this competition) and the defensible arguments (those that are involved in a tie). The important point is that the dialectical status of an argument depends on its interactions with all other available arguments. An important phenomenon here is *reinstatement*: suppose that argument *B* defeats argument *A* but that *B* is itself defeated by a third argument *C*; in that case *C* reinstates *A*. Consider again our rebutting arguments based on witnesses John and Bob. Even if we would prefer Bob's testimony given that he is an adult and John a child, the argument using Bob's testimony may be undercut by a third argument *C* "Bob's testimony is unreliable since he has a strong reason to hate the suspect". Several ways to define the dialectical status of arguments have been proposed, but in the examples below the outcomes are obvious. Technical details can be found in Prakken & Vreeswijk (2002).

3.3. Pollock's system

I next turn to a particular argumentation system, the one of the American philosopher John Pollock (see e.g. Pollock 1987,1995,1998), based on his earlier philosophical work in epistemology (Pollock, 1974). As noted above, Pollock pays much attention to the nature of the defeasible inference rules for epistemic reasoning (which he calls "prima facie reasons", as opposed to "strict reasons", which are the rules of deductive reasoning) and to the ways in which they can be undercut. Prima facie reasons are general epistemic principles for obtaining beliefs from other beliefs and perceptual inputs, such as memory, statistical reasoning and induction. To capture the relative strength of arguments, Pollock allows the assignment of numerical probabilities to applications of reasons. For instance, if we have a generalisation "96% percent of the Americans watch TV every day" and we have that Davis is an American, the reason called the "statistical syllogism" allows us conclude with 96% strength that Davis watches TV every day. However, as noted above, legal-evidential knowledge usually does not come with numbers attached to it, but will be formulated in qualitative terms, such as in: "By far most/almost all/usually/ Americans watch TV every day". For this reason I will ignore the probabilistic aspects of Pollock's system. In fact, in this paper I will completely ignore issues of strength of arguments and focus only on the representation of evidential information and the construction of evidential arguments and counterarguments with

this information. Of course, modelling the strength of evidential arguments is an extremely important issue, but it has to be left for future occasions.

Now arguments can be constructed by chaining reasons, starting from given input information (INPUT). To represent the inferential dependencies between the propositions in an argument, arguments can be depicted as AND trees, where the nodes are propositions and the links represent applications of reasons to these propositions. Pollock combines sets of such trees into an AND/OR graph and adds the appropriate defeat links between nodes, resulting in an "inference graph".

Let us now look in more detail at Pollock's prima facie reasons and their undercutters (but suppressing for simplicity many technical details). For present purposes, five reasons are especially relevant, which I will below paraphrase together with some of their undercutters. As for notation, if reason R says that P is a prima facie reason for Q, then "S is an undercutting defeater of R" is shorthand for "S is a prima facie reason for "P is not a prima facie reason for Q"" (this presupposes that reasons can somehow be expressed in the object language).

The full picture can be summarised as follows. First *perception* is applied to sense data, yielding specific beliefs, and *memory* is used to record and retrieve these data. Then *induction* infers general rules from them, after which the *statistical syllogism* derives new specific beliefs from these rules. Finally, beliefs thus derived *persist* over time.

R1: Perception*: Having a percept with content φ is a prima facie reason to believe φ*

In legal contexts perception applies to witness testimonies, but also to tangible evidence as presented at trial. Pollock (1987) formulates a general undercutter for perception, which I paraphrase as: *"The present circumstances are such that having a percept with content φ is not a reliable indicator of φ" undercuts R1*. Clearly, this undercutter is just the tip of the iceberg of theories on the reliability of perception.

R2: Memory*: Recalling φ is a prima facie reason to believe φ.*

One undercutter (Pollock, 1987) is: *"φ was originally based on beliefs of which one is false" undercuts R2*.

R3: Statistical syllogism*: 'c is an F' and 'F's are usually G's' is a prima facie reason for 'c is a G'*

This principle drives default reasoning with empirical generalisations. The main undercutter is subproperty defeat (which I give in a weak and a strong qualitative form):
'c is an F&H' and 'it is not the case that F&H's are usually G's' is an undercutter of R3.
'c is an F&H' and 'F&H's are usually not G's' is an undercutter of R3.
This expresses that statistical information about a certain class is overridden by conflicting statistical information about a subclass.

As an (admittedly somewhat contrived) example, consider an imaginary piece of statistical information that 55% of American husbands commit adultery in the first ten years of their marriage, and suppose that after ten years of marriage a woman files for divorce since the statistic would prove her husband's adultery 'on the balance of probabilities'. A way for the husband to undercut his wife's argument is to find a weaker statistical relation for some subclass of American husbands to which he belongs.

R4: Induction*: 'most observed F's were G's' is a prima facie reason for 'F's are usually G's'*

Pollock formulates various undercutters to induction based on *bias of samples*.

R5: Temporal persistence*: "φ is true at T_1" is a prima facie reason for "φ is true at a later time T_2" (provided that φ is temporally projectible).*

Temporal persistence is an important aspect of evidential reasoning. For instance, in civil cases the usual way to prove that one has a legal right (e.g. ownership) is to prove that the right was created (e.g. by sale plus delivery).

The other party must then usually prove later events that terminated the right. The condition that φ is temporally projectible is very important, since many propositions, such as a position of a moving object, do not typically persist in time. The Umilian case illustrates that temporal-persistence arguments are also common in criminal cases: from the statement that a revengeful murderous emotion was created with U when the priest refused to marry him an argument is constructed that he still had that revengeful emotion at the time of the killing.

The general scheme for undercutters of temporal persistence arguments is *"Having reason to believe $\neg\varphi$ at T_2 between T_1 and T_3 is an undercutter of R5"*. (Actually, Pollock restricts this to percepts of $\neg\varphi$). For instance, in the Umilian case an argument may be constructed that U ceased to have the murderous emotion when the marriage took place after all.

This completes the brief overview of Pollock's theory of epistemic defeasible reasoning. I next discuss how evidential reasoning can be reconstructed within this theory. Two notions are especially important: generalisations and argumentation schemes.

4 Generalisations

As noted above, empirical default generalisations are an essential element of many evidential arguments. I first discuss how they can be applied and how they can be derived from sources (4.1). Then I show how arguments attacking the application or derivation of generalisations can be modelled (4.2).

4.1 Obtaining and Applying Generalisations

In Pollock's framework, generalisations are applied with the statistical syllogism, and part of their critical testing can be modelled as the search for undercutters of the syllogism. One subtlety not captured by the above qualitative version of the syllogism is that generalisations often come with different modalities, such as "almost always, probably, usually, sometimes". As said above, this is an issue that I leave for future research.

Something also seems to be missing from Pollock's original account. Pollock assumes that all generalisations are based on the reason from *induction*, and that attacks on generalisations can be expressed as undercutters of this reason. However, the generalisations used in evidential reasoning are often not based on careful empirical testing. In fact, according to Twining (1999) they are often based on folk beliefs, infected with value judgements, prejudice or ideology, and so on. Therefore, the induction scheme must be supplemented with other sources of generalisations, and suitable undercutters for these sources must be formulated. I now briefly sketch how this could be done. In fact, this sketch amounts to an analysis of applying and attacking ancillary evidence.

Anderson (1999) distinguishes five kinds of generalisations according to their sources: scientific, expert-based, general knowledge, experience-based and belief-based generalisations. The first source is captured by the induction scheme and the second source will be captured by the expert testimony scheme (see below). Experience-based and perhaps also belief-based generalisations seem to be based on a commonsense counterpart of scientific induction, briefly discussed by Pollock (1995, pp. 82-3). Essentially, these are generalisations that people somehow base on their daily experiences in a non-methodical way. Furthermore, the general-knowledge source could be formulated as a new prima facie reason:

R6: general knowledge: *'It is general knowledge that φ' is a prima facie reason for φ*

Possible undercutters are that a piece of general knowledge is infected by prejudice or value judgements, etcetera.

A typical argument then looks as follows (ending each line with the reason with and the preceding lines from which the line is inferred, and suppressing classical reasoning steps):

Argument A:

1. It is general knowledge that "If x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created". (INPUT)
2. So (presumably) if x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created (1, R6)
3. J falsely tried to prevent U's marriage (INPUT)
4. So (presumably) a revengeful murderous emotion from U towards J was created (2,3,R3)

If R6 is not regarded as an argumentation scheme but as a generalisation, an extra line 1' between 1 and 2 must be added containing that generalisation, and 2 is then derived by from 1 and 1' by the statistical syllogism.

4.2 Attacking Generalisations

As said above, critically testing generalisations is just as important as obtaining and applying them. In the present account, four ways to attack a generalisation can be modelled.

1 *Attacking that they are from a valid source of generalisations*, e.g. "it is not general knowledge that if x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created".
This attack can be modelled as a rebutting attack on a subargument for the intermediate conclusion that something is general knowledge.

2 *Attacking the defeasible derivation from the source*, for instance: "it is indeed general knowledge that if x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created, but this particular piece of general knowledge is infected by folk belief".
This attack can be modelled as an undercutter of R6.

3 *Attacking application of the generalisation in the given circumstances*. This can be modelled as the application of applying more specific generalisations (e.g. "if x falsely tries to prevent y's marriage but y is known to be a gentle person, then usually not a revengeful murderous emotion from y towards x is created", or the weak form with "not usually") Then the subproperty defeater of the statistical syllogism undercuts the use of the general default.

4 *Attacking the generalisation itself*. Such an attack takes the form of an argument for the negation of the attacked generalisation. An example of such an attack is the combination of the above more specific generalisation with the claim that the additional condition is not unusual, or perhaps even that it is usual, as in "People are usually gentle".

The main difference between attacks of the third and the fourth kind is that the third kind of attack accepts the generalisation as a general rule, but denies its application in the case at hand, while the fourth kind of attack denies the generalisation as a general rule ("it is not the case that usually ...").

It might be argued that case-specific generalisations are less prone to attack than universal generalisations (cf. e.g. Twining, 1999, p. 94). In one respect this is indeed the case. For instance, to refute "if *Jedrusik* falsely tries to prevent *Umilian*'s marriage, then presumably a revengeful murderous emotion from *Umilian's* towards *Jedrusik* is created", one must show the contrary for Jedrusik and Umilian, while to refute "if x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created" it suffices to show the contrary for an arbitrary pair of individuals x and y. So any attack on the first generalisation also attacks the second but not vice versa. However, in another respect case-specific generalisations may be more prone to attack. As we saw, part of the reasoning about generalisations involves showing that they are based on a reliable source. Now a universal generalisation can be based on the behaviour of any (random) sample of individuals while a case-specific generalisation must be based on the past behaviour of the specific individuals involved. In many cases the latter may be harder to obtain than the former and in such cases a case-specific generalisation may be more prone to source-based attacks than its universal counterpart.

Concluding this subsection, we see that an argument-based approach within Pollock's system supports the modelling of several ways to use and attack ancillary evidence.

4.3. A logical digression

Let us expand a bit on the difference between the third and fourth way to attack a generalisation, taking ourselves into deeper logical waters. What exactly is the difference between using an exceptional default and denying that a default is true? A person who accepts that P usually implies Q but maintains that R is an exception, is prepared to accept Q if all s/he knows is P: s/he is prepared to assume that R is false, since P&R is an exceptional P case. However, someone who denies that P usually implies Q, is not prepared to accept Q if all s/he knows is P: s/he is not prepared to assume that R is false, since s/he maintains that P&R is not an exceptional P-case. In the Umilian case, a person denying the truth of the murderous-emotion default will, even if its conditions are satisfied, not accept the conclusion that a revengeful murderous emotion was created if she knows nothing about U's gentleness. Let us express this more formally (where $\Rightarrow$ stands for default implication and $\neg$ for negation):

(p1) $\neg(P \& R \Rightarrow Q) \& \neg(P \Rightarrow \neg R)$, so $\neg(P \Rightarrow Q)$

A version with a stronger antecedent is:

(p2) $\neg(P \& R \Rightarrow Q) \& (P \Rightarrow R)$, so $\neg(P \Rightarrow Q)$

An interesting question is whether we can say that such arguments are based on a general pattern, i.e., on a reason. Here, it is illuminating to consider the contrapositives of the two implications underlying arguments (p1) and (p2):

(p1') If $P \Rightarrow Q \& \neg(P \Rightarrow \neg R)$ then $P \& R \Rightarrow Q$

(p2') If $P \Rightarrow Q \& P \Rightarrow R$ then $P \& R \Rightarrow Q$

In the literature on nonmonotonic logic, (p1') is called the principle of rational monotony and (p2') is called the principle of cautious monotony. Both principles have, with other principles, been proposed as valid inference rules for defeasible conditionals (cf. e.g. Pearl, 1992). Note that the issue here is not whether a given generalisation can be applied to a specific case, but whether it can be logically derived from other information. Mathematical semantics for conditional logics have been developed in which the defeasible conditional is (roughly) statistically interpreted as "Most P's are Q's" or as "Almost all P's are Q's" (for an overview see Pearl, 1992) and in which some or all of these principles are deductively valid. The axioms of such logics could be added to the deductive inference rules of Pollock's system. (It should be noted, however, that these principles are deductively valid only on the strong interpretation of $A \Rightarrow B$ as "almost all A's are B's". Arguably, many evidential generalisations satisfy at best the weaker interpretation "most A's are B's".)

Let us see what happens if we add p1, p2 and their contrapositives to the strict reasons, looking how then the murderous-emotion default can be attacked:

Argument B:
1. It is general knowledge that if *x* falsely tries to prevent *y*'s marriage but *y* is known to be a gentle person, then usually not a revengeful murderous emotion from *y* towards *x* is created (INPUT)
2. If *x* falsely tries to prevent *y*'s marriage but *y* is known to be a gentle person, then usually not a revengeful murderous emotion from *y* towards *x* is created (1, R6)
3. It is general knowledge that it is not the case that if *x* falsely tries to prevent *y*'s marriage then *y* is usually not a gentle person (INPUT)
4. It is not the case that if *x* falsely tries to prevent *y*'s marriage then *y* is usually not a gentle person (3, R6)

5 So (presumably) it is not the case that if x falsely tries to prevent y's marriage, then usually a revengeful murderous emotion from y towards x is created (2,4,p1).

Now B rebuts and is rebut by A's subargument containing lines 1 and 2.

It is instructive to see what happens if the gentleness default is instead regarded as a (weak) exceptional default:

Argument B':
1. It is general knowledge that if x falsely tries to prevent y's marriage but y is known to be a gentle person, then usually not a revengeful murderous emotion from y towards x is created (INPUT)
2. If x falsely tries to prevent y's marriage but y is known to be a gentle person, then usually not a revengeful murderous emotion from y towards x is created (1, R6)
3. U is a gentle person (INPUT)
4. So (presumably) R1 as applied in line 4 of argument A is undercut (2,3, subproperty defeat)

A crucial difference between argument B and B' is that B does not need a premise that U is a gentle person (line 3 of B'). This is since in B, unlike in B', this fact is not regarded as an exception to the general default, so that in B the assumption that the exception is false does not have to be refuted. So it turns out that the difference between the third and fourth way of attacking generalisations is very relevant when it comes to the burden of proof.

5 Argumentation schemes

5.1. Introduction

When looking at evidential reasoning (or indeed at reasoning in general), one sees that arguments often follow stereotypical patterns, such as inferences from witness or expert testimonies, causal arguments, or temporal projections. The same holds for attacks on arguments. For instance, witness testimonies are typically attacked on one of three grounds (cf. Schum, 1994): veracity (does the witness believe what s/he says?), objectivity (did the witness's senses give evidence of what s/he believes?) and observational sensitivity (did what the senses gave evidence of really happen?). Such argumentation patterns or "schemes" with associated patterns of attack are the subject of much research in argumentation theory (cf. e.g. Perelman & Olbrechts-Tyteca, 1996; Walton, 1996,1997). They are formulated as schemes of premises and a conclusion, and a set of "critical questions". Some of these questions just ask whether the premises are true, but others are like pointers to rebuttals or undercutters of a scheme. Consider by way of example Walton's (1997, p.210) analysis of the scheme from expert testimony.

Major Premise: Source E is an expert in subject domain S containing proposition A.
Minor Premise: E asserts that A (in domain S) is true (false).
Conclusion: A may plausibly be taken to be true (false).

One critical question listed by Walton is "is E an expert in domain S"? When the major premise is just stated, this is a challenge of an argument's premise, and when the major premise is itself derived with another argument, the challenge points to a rebuttal of that subargument. Another critical question listed by Walton is "Is A consistent with what other experts assert?". This points to rebuttals of an expert-based argument itself. And yet another of Walton's critical questions is "Is E personally reliable as a source?". This points to undercutting arguments, such as an argument that E is not reliable since some of his past research was funded by the company for which he is called as an expert witness.

Since argumentation schemes look like prima facie reasons and critical questions look like sources of undercutters, Pollock's framework seems very suitable for modelling reasoning with argumentation schemes. I will investigate the modelling of two such schemes, arguments based on expert opinions and arguments based on witness testimonies.

5.2. Argumentation schemes in Pollock's system

When modelling argumentation schemes in Pollock's framework, an important question is whether they must be regarded as additional prima facie reasons or as empirical generalisations: in the latter case applying the schemes boils down to applying the (qualitative) statistical syllogism. Technically, the main difference is that the body of reasons is fixed while generalisations can be inferred from, and attacked on the basis of, other knowledge. Now, is it conceivable that someone wants to argue against the expert and witness testimony schemes in general? Or will all attacks take the form of rebutters or undercutters of applications of these schemes? I tentatively believe the latter, and therefore, I will formulate the two testimony schemes as additional prima facie reasons. However, formulating them instead as generalisations is straightforward: what below are undercutters must then be formulated as the second premise of the subproperty defeaters of the statistical syllogism.

I first discuss the scheme from witness testimony. In fact, several treatments are possible. Because of space limitations, we can discuss only one of them, following the terminology of Schum (1994). Let us first (rather arbitrarily) assume that the veracity of witnesses may be presumed (alternatively, it can be regarded as an additional premise of R7 below). Then the scheme can be formulated as the following reason:

R7: Witness testimonies: *'Witness W says φ' is a prima facie reason for believing φ.*

Let us define the following undercutter for this scheme: *"Witness W is not truthful" is an undercutter of R7.*

Actually, there is no need to formulate lack of a witness's objectivity and observational sensitivity as additional undercutters of this scheme. This is since a witness will always tell about his or her past observations, so φ will in practice always be of the form "I recall that I observed ψ". So reasoning with witness testimonies is in fact a chain of three prima facie reasons: the witness testimony reason, the memory reason, and the perception reason: first the witness scheme is used to infer "I recall I observed ψ", then the memory scheme provides "I observed ψ" and finally the perception scheme yields ψ. Thus lack of objectivity is handled by undercutters of both memory and perception, and defects in observational sensitivity by undercutters of perception.
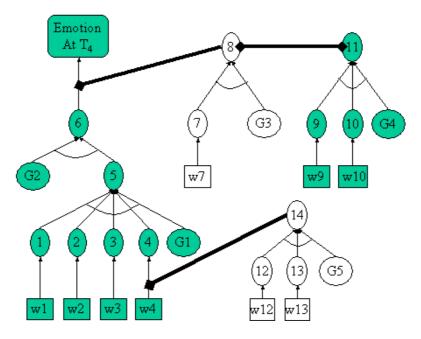
Treating the scheme from expert testimony is simpler:

R8: Expert testimonies*: 'E says φ' and 'E is expert about φ' is a prima facie reason for believing φ.*

Of the critical questions discussed by Walton (1997), two seem to correspond to undercutters, viz. "is *E* an expert in domain *S?*" and "is *E*'s assertion of *A* backed by evidence?". But other accounts may be possible.

6 An example

Let us now illustrate the above analysis with the Umilian case of Section 2, focussing on the "motive" probandum that at the time of the killing the suspect had a revengeful murderous emotion towards the victim. (My interpretation closely follows Wigmore's (1931) own Wigmore chart of the case, but other interpretations also seem possible). Not surprisingly, the case contains several uses of witness testimonies. Wigmore discusses one attack on such a use, viz. an attack on the witness who said that the content of the letter received by the priest was untrue. The attack is that the witness was a discharged employee of U and therefore had a motive to testify in a way that discredits U. This is an undercutter, concluding to "the witness is not truthful".

As earlier mentioned above, the case also contains an application of the temporal persistence scheme: the revengeful emotion created when the priest refuses to marry U is assumed to persist till the time of the murder. This persistence argument is undercut by the argument that the emotion disappeared after the priest still agreed to marry U. This argument is in turn rebut by the argument that the emotion lingered on since U and J remained in daily contact and U's wife also remained there.

The other inferences identified by Wigmore all seem to be based on implicit generalisations, which all seem to be of the general-knowledge, experience-based or belief-based type. I list them below (the second one was already mentioned above, but this time I include a reference to time, to make temporal persistence explicit).

G1: if (1) a priest receives a letter (2) written by x, and (3) the priest refuses to marry y at T because of the letter, and (4) the content of the letter is not true, then (5) x falsely tries to prevent y's marriage at T.

G2: If (5) x falsely tries to prevent y's marriage at T, then (6) a revengeful murderous emotion from y towards x tends to be created at T

G3: if (7) a marriage of x prevented by y still takes place at T, then (8) x will not have a revengeful murderous emotion towards y after T

G4: if (9) x and y remain in daily contact between T_1 and T_2 after a marriage of y that x falsely tried to prevent, and (10) the wife also remains there, then (11) a revengeful murderous emotion from y towards x tends to exist at T_2.

G5: if (12) a witness is a discharged employee of the suspect and (13) the witness says something that discredits the suspect then (14) the witness will tend to be untruthful.

For simplicity, I leave further generalisations that can be used to derive (14) implicit.

The argument graph now looks as in Figure 1 (Where "Emotion at T_4" is the probandum that a revengeful murderous emotion existed in U at the time of the killing, and where wi means that a witness testified to i.). Readers familiar with Wigmore's charting method will note many similarities but also one crucial difference: while in Wigmore's charts (especially as used by the modern evidence scholars) generalisations appear as links, here they appear as nodes. This is since in the present account generalisations are regarded as propositions, so that they can be reasoned about. The thin links in Figure 1 correspond to applications of prima facie inference rules: the eight lines from a square box upwards are applications of the witness testimony scheme, and the inferences that involve a generalisation are applications of the statistical syllogism. The three thick lines express defeat relations between arguments.

The chart reflects two independent sources of doubt with respect to the main argument for the probandum (which is the grey-coloured structure on the left). The first is the undercutting of its witness-based argument "w4, so 4". In the present account, undercutting implies strict defeat, and since the undercutting argument "12,13, G5, so 14" (the white structure at the bottom) is not itself defeated by any argument, the main argument for the probandum is overruled since one of its subarguments is overruled. However, let us for the sake of argument assume that the main argument is reinstated by an argument strictly defeating its undercutter (e.g. with an exception to G5). Then there is still the second source of doubt, viz. the undercutting of the final, temporal

projection step in the main agument by the argument for (8) (the white structure at the top). This undercutter is in turn rebut by the argument for (11) (the grey structure to its right), but without an evaluation of their relative strength this rebutting relation is mutual, so that both rebutting arguments are defensible; and this in turn means that the main argument for the probandum is also defensible, since its status depends on the relative assessment of the arguments for (8) and (11).

What has our case study illustrated? Firstly, it makes the sources of conflict explicit: there is a dispute on whether witness w4 is credible, and there is a dispute on whether the murderous emotion created when the priest refused to marry U persisted to the time of the murder. The various arguments pertaining to these issues can easily be attributed to a side in the dispute (visualised by the use of grey (prosecution) and white (defence) in the figure). Secondly, it illustrates that at least some form of sensitivity analysis is also possible in a logical approach: for instance, a sense-making program could calculate and visualise the change of status of the main probandum given a certain assessment of the rebutting conflict, or given some hypothetical attack on the argument for (14). Of course, the threevalued output of our argument-based approach is much cruder than the fine-grained one of probability theory, which would assign numerical probabilities to the nodes (8), (11) and "emotion at T_4", but one point of this paper is that in many cases a qualitative threevalued assessment is the best result one can obtain.

7 Conclusion

In this paper I have argued that logical models of evidential reasoning based on argumentation are a worthwhile alternative to probabilistic models. The main strong points of an approach based on argumentation logics seem that they require no numerical input, they make ancillary evidence and sources of conflicts explicit instead of compiling them away in probability distributions, and they support an explicit modelling of the stereotypical argument and attack forms used in evidential reasoning.

As for related research and future research issues, as remarked above, there is not much AI work on the formal modelling of legal reasoning about evidence, but there are a few exceptions. In section 3.1 I already remarked that Verheij (2000) earlier advocated the approach of the present paper. Compared to Verheij the main contributions of this paper are a systematic account of the various ways to attack generalisations and a discussion of the embedding of some stereotypical forms of evidential reasoning in Pollock's system. While Verheij and this paper advocate an approach in terms of defeasible argumentation, Keppens & Zeleznikow (2002) instead take their starting point in AI models for automated diagnosis based on abduction, presenting a model of rational evidence-gathering in murder cases. This is an interesting approach, since much evidential reasoning is causal reasoning, which arguably is abductive in nature; Pollock does not include a reason for abductive arguments in his system, so that it remains to be seen whether his system supports a natural modelling of abductive reasoning. Another issue for further research is the modelling of reasoning about the relative strength of arguments. For reasoning about issues of law such models already exist (see Prakken & Sartor, 2002 for an overview) but it is still an open question whether these models naturally apply to reasoning about the facts.

Finally, an important element left out by the above analysis is the dynamics of evidential reasoning. For instance, when a (continental) trier of fact is faced with two rebutting witness-based arguments and has no information on their relative strength, s/he will not conclude that both arguments are defensible, but try to obtain such information, for instance, by asking questions to the witnesses. In AI research on diagnosis (applied by Keppens & Zeleznikow, 2002 to evidential reasoning) strategies for evidence gathering have been modelled, but their modelling in argument-based approaches is still largely an open question.

References

Anderson, T.J. (1999), On generalizations I: a preliminary Exploration. *South Texas Law Review*, Summer 1999, 455-481.

Brooks, R.A. (1991), Intelligence without representation. *Artificial Intelligence* 47:139-159.

Ginsberg, M.L. (1987) (ed.), *Readings in Nonmonotonic Reasoning*. Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Kadane, J.B. & Schum, D.A. (1996), *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. New York: John Wiley & Sons.

Keppens, J. & Zeleznikow, J. (2002), On the role of model-based reasoning in decision support in crime investigation. *Proceedings of the Third International Conference on Law and Technology*, Boston, MA, pp. 77-83.

Kirschner, P.A., Buckingham Shum, S.J. & Carr, C.S. (2002) (eds.): *Visualizing Argumentation. Software Tools for Collaborative and Educational Sense-Making*, London: Springer Verlag, pp. 75-96.

Lenat, D.B. (1998), From *2001* to 2001: Common sense and the mind of HAL. In D.G. Stork (ed.), *HAL's Legacy: 2001's Computer as Dream and Reality*. Cambridge, MA: MIT Press.

Loui, R.P. (1987) Defeat among arguments: a system of defeasible inference. Computational Intelligence 2: 100-106.

Pearl, J. (1992), Epsilon-semantics. *Encyclopedia of Artificial Intelligence*, Vol 1, pp. 468-475.

Perelman, Ch. & Olbrechts-Tyteca, L. (1969), *The New Rhetoric (La Nouvelle Rhetorique),* Notre Dame: University of Notre Dame Press.

Pollock, J.L. (1974), *Knowledge and Justification*. Princeton University Press.

Pollock, J.L. (1987), Defeasible reasoning. *Cognitive Science* 11:481-518.

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

Pollock, J.L. (1998), Perceiving and reasoning about a changing world. *Computational Intelligence* 14:498-562.

Prakken, H. & Reed, C. & Walton, D.N. (2003), Argumentation schemes and generalisations in reasoning about evidence. *Proceedings of the Ninth International Conference of Artificial Intelligence and Law*. New York: ACM Press, pp. 32-41.

Prakken, H. & Sartor, G. (2002), The role of logic in computational models of legal argument: a critical survey. In A. Kakas and F. Sadri (eds.), *Computational Logic: Logic Programming and Beyond. Essays In Honour of Robert A. Kowalski, Part II*. Springer Lecture Notes in Computer Science 2048, Berlin, 342-380.

Prakken, H. & Vreeswijk, G.A.W. (2002), Logical systems for defeasible argumentation. In D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, second edition, Vol 4, pp. 219-318. Kluwer Academic Publishers, Dordrecht etc.

Rescher, N. (1977), *Dialectics: a Controversy-oriented Approach to the Theory of Knowledge*. albany, NY: State University of New York Press.

Schum, D.A. (1994), *The Evidential Foundations of Probabilistic Reasoning*. New York: John Wiley & Sons.

Schum, D.A. (2001), Alternative views of argument construction from a mass of evidence. *Cardozo Law Review* 22:1461-1502.

Schum, D.A. & Tillers, P. (1991), Marshalling evidence for adversary litigation. *Cardozo Law Review* 13:657-704.

Twining, W. (1999), Necessary but dangerous? Generalizations and narrative in argumentation about "facts" in criminal process. In M. Malsch & J.F. Nijboer (eds.): *Complex Cases. Perspectives on the Netherlands Criminal Justice System*, pp. 69-98. Amsterdam: Thela Thesis.

Van Engers, T.M., Gerrits, R., Boekenoogen, M., Glassée, E. & Kordelaar, P. (2001), POWER: using UML/OCL for modelling legislation – an application report. *Proceedings of the Eighth International Conference of Artificial Intelligence and Law*. New York: ACM Press, pp.157-167.

Verheij, B. (2000), Dialectical argumentation as a heuristic for courtroom decision-making. In P.J. van Koppen & N.H.M. Roos (eds.): *Rationality, Information and Progress in Law and Psychology. Liber Amoricum Hans F. Crombag*. Maastricht: Metajuridica Publications, pp. 203-226.

Wagenaar, W.A., Koppen, P.J. van & Crombag, H. (1993), *Anchored Narratives. The Psychology of Criminal Evidence*. NewYork: St. Martin's Press.

Walton, D.N. (1996), *Argumentation Schemes for Presumptive Reasoning*. Mahwah, N. J.: Erlbaum.

Walton, D.N. (1997), *Appeal to Expert Opinion*, University Park: Penn State Press.

Wigmore, J.H. (1931), *The Principles of Judicial Proof*, 2nd ed., Boston: Little, Brown and Company.