DIALECTICAL ARGUMENTATION AS A HEURISTIC FOR COURTROOM DECISION MAKING

Bart Verheij

Department of Metajuridica, Universiteit Maastricht
P.O. Box 616, 6200 MD Maastricht, The Netherlands
bart.verheij@metajur.unimaas.nl, http://www.metajur.unimaas.nl/~bart/

1 Introduction

In his 1994 article 'On the Artificiality of Artificial Intelligence', Crombag claims that artificial intelligence (of the symbolic kind) is of little psychological relevance. According to Crombag, our understanding of the mind (and of the legal mind in particular) is hardly enhanced by artificial intelligence research. The main reason for his opinion is that there is no indication that the brain is like a digital computer in structure or in behavior.

I agree with Crombag that digital computers and brains are very different in structure. However, no one - not even the early far too optimistic advocates of *good old fashioned* artificial intelligence of the 60s - has ever had a different opinion.

I disagree with Crombag that digital computers and brains are different in behavior. Of course a desktop computer (often used for text and message processing only) is very different in behavior from a human brain, but that is irrelevant. In an important sense, digital computers and brains can be similar in behavior. The key point is that digital computers can be programmed, and that there is virtually no limit to the kinds of programs that can run on digital computers. From the perspective of information processing, digital computers are good approximations of implemented universal Turing machines, which essentially means that they can be programmed to mimic any kind of information processing.

And this really means *any* kind. Not only logical reasoning and playing chess, but also interpreting visual data and finding analogies are kinds of information processing. Even legal reasoning can be fruitfully viewed as a kind of information processing. Thagard's (1996) introduction to cognitive science gives many examples of what he calls the computational-representational view of the mind.

This does not mean that if a digital computer mimics certain brain behavior, it mimics this behavior at all levels of inspection. Of course not: a digital computer is not the same as a human brain. If one probes deep enough, the differences become obvious. It remains a miraculous fact that *at a sufficiently high level* digital computers can mimic any kind of information processing behavior. Even though, internally, the digital computer Deep Blue and Kasparov's brain found their chess moves in very different ways, at the level of playing grandmaster chess Deep Blue's and Kasparov's behavior were much alike.

The only limitations on the information processing capabilities of digital computers are computational speed and memory, but these bounds are still pushed back at a high pace.

Crombag (1994) paraphrases the well-known connectionist challenge to artificial intelligence of the symbolic kind, advocating neural network models. The challenge comes down to the demand of designing a model with a slightly deeper level of detail than in symbolic artificial intelligence. Instead of only looking at the whole brain as an information processor, the role of individual brain cells and their interconnections is taken into account.

Interestingly, a more recent challenge to traditional artificial intelligence does not consider the level of individual brains as too high (as in the connectionist challenge), but as too low. The higher level of interaction of brains with their environment and with other brains is given due attention.

The best models of brain behavior must of course take all intertwining levels of looking at the brain's behavior into account: the low level of brain cells[1], the intermediate level of brains, and the high level of communicating brains interacting with their environment.

The psychological relevance of artificial intelligence is however not to be found in the claim that digital computers can behave like brains. The importance of the rise of artificial intelligence is that it has added a powerful tool to the research methodologies of psychology: models of brain behavior can now be successfully implemented on digital computers.

This new methodology has at least three advantages. First, the goal of implementation of a brain behavior model on a digital computer sets very high standards to the design of the model itself. It has to

[1] Some think that the brain cell level is not sufficiently detailed. Even the level of the quantum physics of the brain has been defended as essential, though unconvincingly.

be unambiguous, precise and detailed. The popularity of formal modeling is a consequence of this. Second, computer models have objectively determinable behavior that can be the subject of empirical study. As a result, the testability of brain behavior models has increased. Third, it is easy to do experiments with computer models. Parameters of a model can be adjusted and the behavioral consequences can be studied. This leads to useful 'pre-empirical' testing of brain behavior models. Many unwanted or unexpected types of behavior predicted by the model can already be spotted at the phase of modeling, and not as late as at the phase of empirical testing.

In the present paper, two approaches to the modeling of legal decision making are discussed and compared: the anchored narratives and the dialectical argumentation approaches. The former (by Crombag, Van Koppen and Wagenaar) originates from the psychology of law, the latter from artificial intelligence and law.

The next section introduces the modeling of legal decision making. In section 3, the theory of anchored narratives is summarized, and in the sections 4 and 5 that of dialectical argumentation. The paper ends with a comparison of the two theories in section 6. As an appendix, a formal version of the theory of dialectical argumentation is presented.

2 Modeling legal decision making

Trying criminal cases is hard. The problem faced by a judge in court can be phrased in a deceptively simple way though, as follows: in order to come to a verdict, a judge has to apply the rules of law to the facts of the case. In a naïve and often criticized model of legal decision making (reminding of the *bouche de la loi* view on judges), the verdict is determined by applying the rules of law that match the case facts. This naïve model of legal decision making can be referred to as the subsumption model.

A problem with the subsumption model is that neither the rules of law nor the case facts are available to the legal decision maker in a sufficiently well-structured form to make the processes of matching and applying a trivial matter. First, there is the problem of determining what the rules of law and the case facts are. Neither the rules nor the facts are presented to the judge in a precise and unambiguous way. A judge has to interpret the available information about the rules of law and the case facts.

Second, even if the rules of law and the case facts would be determined, the processes of matching and applying can be problematic. It can for instance be undetermined whether some case fact falls under a particular rule's condition. Additional classificatory rules are then needed. In general, it can be the case that applying the rules of law leads to conflicting verdicts about the case at hand, or to no verdict at all. In the latter situation, it is to the judge's discretion to fill the gap, in the former, he has to resolve the conflict. In both cases, the judge should provide some further justification of his opinion.

As a result of these difficulties, many consider legal decision making as a kind of gradual theory construction. By selecting and interpreting the available material (legal codes, police reports, court pleadings, etc.), the decision maker constructs a preliminary theory about the applicable rules, the proven facts and the appropriate verdict. He then performs a number of checks on the preliminary theory. Are there no inconsistencies? Is the verdict justified? If necessary, the theory is adapted to solve the weaknesses found.

In several disciplines, models of legal decision making have been designed along these lines of gradual theory construction. E.g., in the psychology of law, Crombag, Van Koppen and Wagenaar have proposed the anchored narratives approach towards the modeling of legal decision making (*Dubieuze zaken*, 1992, 1994, *Anchored narratives*, 1993). Legal decisions are seen as structured stories, anchored in common knowledge.

When Crombag, Van Koppen and Wagenaar contrast their theory with logical inference theories, they allude to the possibility of extended logical systems that are better suited as models of legal decision making than the subsumption model (*Anchored narratives*, p. 22). Especially in the field of artificial intelligence and law, such extended logical systems have indeed been designed (cf., e.g., Hage, 1997, Prakken, 1997, Verheij, Hage and Van Maanen, 1999). Among the topics addressed in such extended legal logics are exceptions, inconsistencies, gaps, contingent validity and rule properties. Below, the focus is on a theory of dialectical argumentation, characterized by the exchange of arguments and counterarguments, as it is being designed during the development of the experimental argument assistance system ArguMed (Verheij, 1999).

3 Anchored narratives

In their books *Dubieuze zaken* (1992, 1994, p. 61f.) and *Anchored narratives* (1993, p. 33f.), Crombag, Van Koppen and Wagenaar present the theory of anchored narratives as a model of legal decision making. The starting point of the theory is that proof in a criminal trial is essentially telling a good story. In the following, the theory of anchored narratives is summarized. More information about the theory and many examples illustrating it can be found in the books mentioned.

In the theory of anchored narratives, judges make two judgments in criminal cases. First, they determine whether the stories of the parties before him (i.e., the prosecution and the defense) are plausible. Here the *quality* (or *goodness*) of the stories is at issue. Second, judges decide whether the available evidence is sufficiently supported by facts. This is where the *anchoring* of stories is examined.

Crombag, Van Koppen and Wagenaar consider their theory to be part of the tradition of narrative theory in cognitive psychology. In this tradition, stories (or narratives) provide the context that gives meaning to the elements of the story. This can already be illustrated by the following mini-story:

> Peter shot a gun. George was hurt.

When one is told this mini-story, one is inclined to assume that George was hurt by Peter shooting the gun. This is however not an explicit part of the story, and can be false.

In their discussion of the *quality of stories*, Crombag, Van Koppen and Wagenaar focus on story grammars. They picked the story grammar as proposed by Bennett and Feldman, and extended by Pennington and Hastie. According to Bennett and Feldman, a good story has a central action, to which all elements of the story are related. A good story does not have loose ends. Moreover, in a good story the setting of the action unambiguously explains why the central action occurred as it did. If not, there are elements missing from the story, or there are contradictions. Note that the constraint of a central action can explain why one is inclined to assume that George was hurt by Peter shooting the gun: by the assumption, the two elements of the mini-story are connected.

Crombag, Van Koppen and Wagenaar describe how Pennington and Hastie have extended the theory by Bennett and Feldman. Pennington and Hastie distinguish three types of factors that can explain actions: physical conditions, psychological conditions and goals. Again a general setting connects the elements. In the context of criminal cases, a good story must contain the accused's motive, and show that the accused had the opportunity to commit the crime.

An experiment by Pennington and Hastie has shown that a good story on a criminal case (i.e., a story that contains all the elements prescribed by the story grammar) does not guarantee a unique outcome. It turned out that by different selections and evaluations of the evidence test persons reached outcomes ranging from first-degree murder, through second-degree murder and manslaughter, to self-defense. In another experiment, Pennington and Hastie showed the influence of story order on verdicts. It turned out that if a party's story was told to the test person in story order instead of in a random order, such as the witness order, the test person more easily followed that party's story in the verdict. It turned out that if the prosecution's story was told in story order, while the defense's story was told in a random order, the accused was convicted in 78% of the cases. If on the other hand the prosecution's story was told in random order and the defense's in story order, the accused was convicted in 31% of the cases. Crombag, Van Koppen and Wagenaar conclude that telling the story well is half the work.

Crombag, Van Koppen and Wagenaar claim that *story anchoring* is needed in order to justify why a story is assumed to be true. For instance, the statement of a policeman that he saw that Peter fired a gun at George, can support that Peter indeed fired a gun at George. By itself, the evidence consisting of the policeman's statement does not prove that Peter fired a gun at George. If the policeman's statement is considered as proof, this is the result of the acceptance of the general rule that policemen tell the truth. Rules need not be universally true; there can be exceptions. No one believes that policemen always tell the truth, but many hold the belief that policemen tell the truth most of the time. According to Crombag, Van Koppen and Wagenaar, there are common-sense generally true rules that underlie the acceptance or rejection of a piece of evidence as proof. They refer to such rules as anchors.

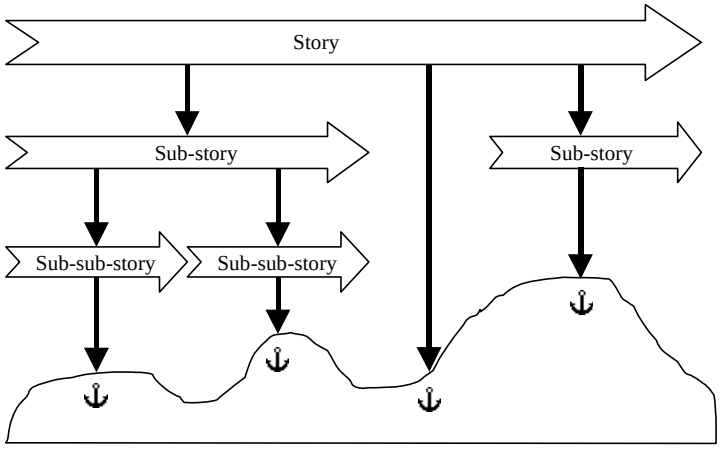
Crombag, Van Koppen and Wagenaar note that different legal systems can not only use different rules as anchors, but even opposites. They give the example of the assessment of confessions. Under English law, a conviction can be based only on the accused's confession, while in Dutch law, additional evidence is required. This suggests that the English use the anchoring rule that confessions are usually true, and the Dutch the opposite rule that confessions are often untrue.

Since the rules used as anchors can have exceptions, it can be necessary to show that a particular exception does not occur. Crombag, Van Koppen and Wagenaar discuss the example of the truthfulness of witnesses (*Anchored narratives*, p. 38). Even if one assumes that witnesses normally tell the truth, the rule is not a safe anchor when the witness has a good reason to lie. Additional evidence is required, for instance, the testimony of a second witness. Even if both witnesses are unreliable since they have good reasons to lie, it can be argued that when their testimonies coincide the combined statements suffice as proof. The anchor would than be that lying witnesses do normally not tell the same lies. There is however again an exception: if the two testimonies are not independent, for instance since the witnesses have conferred, the anchoring is again not safe. In section 5 below, we return to this discussion of the truthfulness of witnesses in the context of dialectical argumentation.

In the theory of anchored narratives, stories are hierarchically structured. The main story can consist of substories that on their turn contain sub-substories, and so on. The idea is that each substory is a further specification of the story or one of its parts. In each substory, a rule is used as an anchor to connect one or more pieces of evidence to the decision of the story or to a part of the decision. A difficulty arises from the fact that the rules used as anchors often remain implicit. Making the naïvely adopted rule explicit can lead us to reject it (*Anchored narratives*, p. 38).

If one goes to a deeper level in the story hierarchy, the anchors will become more and more specific, and as such *safer*. For instance, at a high level, the anchoring rule could be that witnesses normally tell the truth, while at a deeper level it could be replaced by the rule that witnesses that have no good reason for lying normally tell the truth.

The following figure (adapted from *Dubieuze zaken*, p. 72, *Anchored narratives*, p. 39) illustrates the theory of anchored narratives.



The use of rules as anchors gives the theory of anchored narratives a deductive element. A decision follows from the evidence on the basis of a general rule. According to Crombag, Van Koppen and Wagenaar, anchoring is not equal to subsuming under a rule, since rules can have exceptions (*Anchored narratives*, p. 58).

Crombag, Van Koppen and Wagenaar use their theory of anchored narratives in order to explain what they call *dubious cases* (or *dubious convictions*). In their terminology, a criminal conviction is dubious if the District Court's verdict was reversed by the Court of Appeals because of a different evaluation of the evidence, or if the defense attorney remained strongly convinced of his client's innocence, even after (repeated) conviction (*Anchored narratives*, p. 11). Thirty-five of such dubious cases were obtained from criminal lawyers, or were selected from among the cases in which one of the authors served as an expert witness. Crombag, Van Koppen and Wagenaar claim that their set of cases supports the theory of anchored narratives, since the anomalies that occur in the cases can only be explained by their theory.

4 Dialectical arguments

Another approach towards legal decision making does not focus on stories, but on the dialectical nature of argumentation: argumentation does not only involve support by reasons, but also attack by counterarguments.

In a standard view, arguments express how a conclusion is supported by a set of premises. In this view, arguments can be thought of as reason/conclusion-structures, or as formal derivations. An example argument is the following:

(1) Peter shot George since witness A states that Peter shot George. Therefore it should be investigated whether Peter murdered George.

In this argument, the statement that witness A states that Peter shot George, is a reason for the conclusion that Peter shot George, which on its turn is a reason for the conclusion that it should be investigated whether Peter murdered George. The argument does not explicitly state why the testimony is considered a reason for the occurrence of the shooting incident, nor why the shooting incident is a reason for the murder investigation. One could for instance state that witnesses' testimonies are often truthful, and that shooting incidents are to be investigated, respectively. Below we return to such backings of argument steps (that are logically comparable to the anchors of Crombag, Van Koppen and Wagenaar).

In the standard view, there is no room for the idea of counterarguments. It can for instance be the case that one statement in an argument attacks another statement. An example is the following:

(2) At first sight, Peter shot George since witness A states that Peter shot George. However, since witness A is unreliable, A's testimony does not support that Peter shot George.

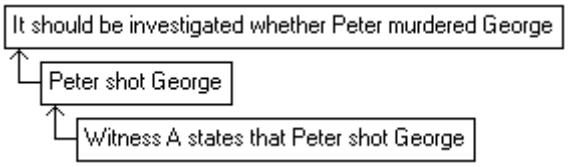
In the argument, the statement that witness A is unreliable, attacks the connection between the reason that witness A states that Peter shot George, and the conclusion that Peter shot George.[2]

Arguments that not only contain supporting reasons, but also attacking reasons, are *dialectical arguments*. The argument (2) is an example. Dialectical arguments are a generalization of the arguments in the standard view, in which arguments only contain supporting reasons.

Dialectical arguments arise naturally if one studies defeasible argumentation, which is currently the topic of much research. In defeasible argumentation, it can happen that an argument that first justifies its conclusion, in a later stage is no longer justifying since it is attacked and defeated by a counterargument. In the following, a theory of dialectical arguments is outlined. An earlier version of the theory has been used in the development of the argument assistance system ArguMed (Verheij, 1999).[3]

Dialectical arguments are structured sets of statements. (Here we restrict ourselves to finite dialectical arguments.) Each dialectical argument is constructed from an initial statement, by consecutively adducing statements supporting or attacking previous statements.

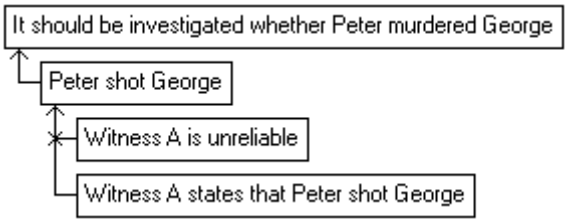
In the following figure, a dialectical argument is graphically represented. It is constructed from the initial statement that it should be investigated whether Peter murdered George, by first adducing the statement that Peter shot George as a reason for the initial statement, and second adducing the statement that witness A states that Peter shot George as a reason for the statement that Peter shot George.



The dialectical argument above is not dialectical proper: it does not contain attacking statements. The following figure represents a proper dialectical argument:

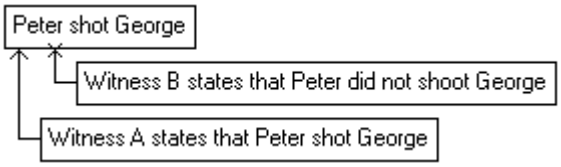
[2] In Pollock's (1987) terminology, a reason that attacks the connection between a reason and a conclusion, is an *undercutting defeater*.

[3] An argument assistance system is a computer program that serves as an aid to draft and generate arguments. Argument-assistance systems should be distinguished from the more common automated reasoning systems. The latter automatically perform reasoning on the basis of the information in their 'knowledge base', while the former merely assist the user's reasoning process. For more information, the reader is referred to my web site on automated argument system at www.metajur.unimaas.nl/~bart/aaa/. ArguMed and its predecessor Argue! can be downloaded.

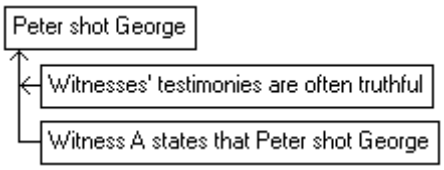


Here the statement that witness A is unreliable, attacks the connection between A's testimony and the shooting incident.

Statements can also attack other statements, as in the dialectical argument represented in the following figure. Here B's testimony that Peter did not shoot George is a reason attacking the statement that Peter shot George.

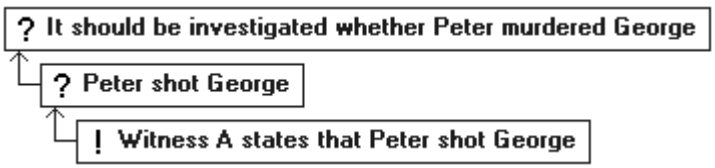


The connection between a reason and a conclusion cannot only be attacked but also be supported. An example is given in the following figure. The statement that witnesses' testimonies are often truthful, is adduced as a reason that supports the connection between A's testimony and the shooting incident.



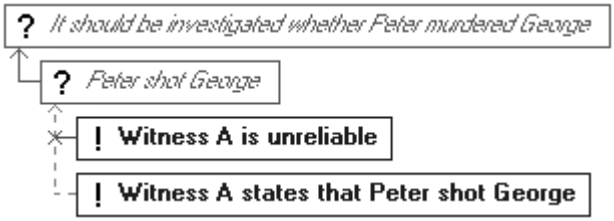
The statement that witnesses' testimonies are often truthful serves as a backing of the supporting argument step (cf. Toulmin's (1958) backings and Crombag, Van Koppen and Wagenaar's anchors).

Dialectical arguments can be evaluated with respect to a set of (defeasible) assumptions. The following figure shows an evaluated dialectical argument.



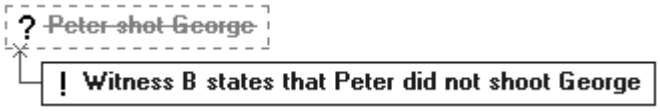
The argument is evaluated with respect to the assumption that witness A states that Peter shot George. It then follows that Peter shot George, and that it should be investigated whether Peter murdered George. Assumptions are preceded by an exclamation mark, all other statements by a question mark. Statements that are not assumptions, are referred to as *issues*. In the argument above, all three statements are evaluated as justified, which is indicated by the use of a dark, bold font. The statement that witness A states that Peter shot George, is justified since it is an assumption and there is no statement attacking it. The other two statements are justified since there are reasons justifying them.

Since statements can be attacked, the statements in an evaluated dialectical argument are not necessarily all justified. An example is shown in the following figure.



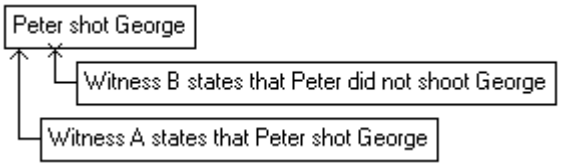
The statement that witness A is unreliable is a second assumption. It has the effect that A's testimony is no longer a reason justifying that Peter shot George. As a result, the statement that Peter shot George is not justified: it is not an assumption, and there is no supporting reason justifying it. The statement is also not defeated since there is no attacking statement defeating it. The light, italic font indicates that the statement that Peter shot George, is neither justified nor defeated.

The following figure shows an evaluated dialectical argument in which a defeated statement occurs.



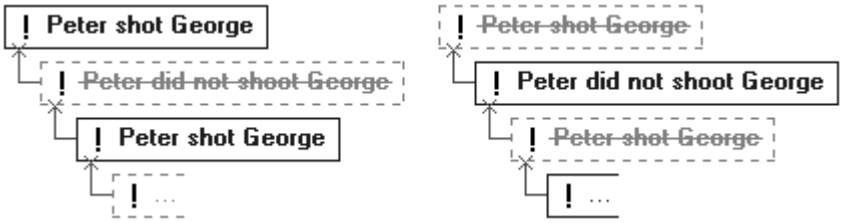
The statement that witness B states that Peter did not shoot George, attacks the statement that Peter shot George. Since it is an assumption, it is justified, and therefore it defeats the statement that Peter shot George. The bold struck-through font indicates that the statement is defeated. If the statement that Peter shot George, would have been an assumption, instead of an issue, it would still be defeated. All assumptions are *defeasible*: if there is an attacking statement defeating an assumption, it is defeated.

Not all dialectical arguments can be evaluated with respect to any set of assumptions. An example is the argument in which there are opposing testimonies with regards to the shooting incident:



If the two witness statements are assumptions, the dialectical argument cannot be evaluated. The statement that Peter shot George, would then have to be justified since there is a statement justifying it, and defeated since there is a statement defeating it.

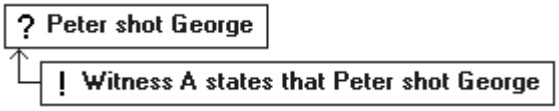
If a dialectical argument can be evaluated with respect to any set of assumptions, the evaluation is not necessarily unique. Multiple evaluations arise for instance if two statements attack each other. The figure below shows what the two evaluations that arise if the statements that Peter shot George and its negation attack each other.



In the appendix, a formal version of the theory of dialectical arguments is briefly presented. The formal version uses two connectives $\rightsquigarrow$ and $\bowtie$ in order to express the support and attack relation between statements. A sentence of the form $\varphi \rightsquigarrow \psi$ (where φ and ψ are sentences) expresses that the statement (expressed by the sentence) φ supports the statement ψ, and the sentence $\varphi \bowtie \psi$ expresses that φ attacks ψ. Using this logical language, the evaluated dialectical argument in the following figure can be formalized as the following one-step *Modus ponens* derivation:

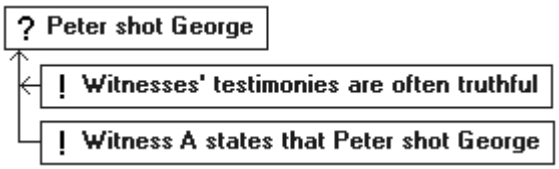
$$\frac{p \rightsquigarrow q \quad p}{q}$$

Here p abbreviates 'Witness A states that Peter shot George' and q 'Peter shot George'. In the *Modus ponens* derivation, the implicit assumption $p \rightsquigarrow q$ underlying the argument step occurs as an explicit premise. One could read $p \rightsquigarrow q$ for instance as 'If witness A states that Peter shot George, then Peter shot George'.



The sentences $\varphi \rightharpoonup \psi$ and $\varphi \rightharpoondown \psi$ can be thought of as the *warrants* of supporting and attacking argument steps (cf. Toulmin, 1958, Verheij, 1999).

The warrant of an argument step can itself be supported, as in the evaluated dialectical argument below.



If r abbreviates the sentence " Witnesses' testimonies are often truthful ", then the argument can be formalized as a derivation consisting of two chained instances of *Modus ponens*:

$$\dfrac{\dfrac{r \rightharpoonup (p \rightharpoonup q) \quad r}{p \rightharpoonup q} \qquad p}{q}$$

In the derivation, the conditional sentences $p \rightharpoonup q$ and $r \rightharpoonup (p \rightharpoonup q)$ occur as premises, making implicit assumptions of the argument explicit. Statements supporting the warrant of a supporting argument step are similar to Toulmin's (1958) backings and Crombag, Wagenaar and Van Koppen's anchors.

Note that *Modus ponens* derivations only correspond to dialectical arguments that do not contain attacking statements, and that do not contain statements supported by more than one reason. As a result, the notion of a dialectical argument is a genuine generalization of that of a *Modus ponens* derivation.

5 Dialectical argumentation

The dialectical arguments discussed in the previous section can be taken as the starting point for a theory of *dialectical argumentation*. Here dialectical argumentation is considered as a process involving several kinds of events, or *argument moves*:

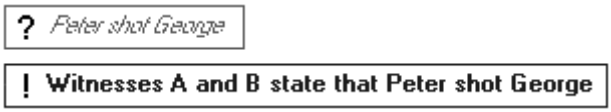
- *Statements are made*, either by raising an issue, by making an assumption, or by changing an issue into an assumption or vice versa.
- *Statements are supported*, by adducing a new reason for a statement, by drawing a new conclusion from a statement, or by turning a statement into a reason for another statement
- *Statements are attacked*, by adducing a counterargument to a statement, by adding a new statement to which an earlier statement is a counterargument, or by turning a statement into a counterargument to another.[4]

An important characteristic of dialectical argumentation, is that the statuses of the statements can change during the process. A statement can for instance at one stage be justified, while it is defeated at the next.

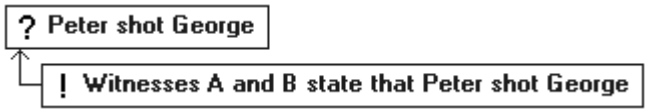
In the following an example line of argumentation is discussed, based on the truthfulness of witnesses. The discussion by Crombag, Van Koppen and Wagenaar (*Anchored narratives*, p. 38) is used (see section 3 above).

A line of argumentation can for instance start with two statements, viz. that Peter shot George and that witnesses A and B state that Peter shot George. Since the former statement is an issue (as is indicated by the question mark), it is not justified (and also not defeated). The latter statement is justified since it is an assumption (as indicated by the exclamation mark) that is not attacked by another statement.

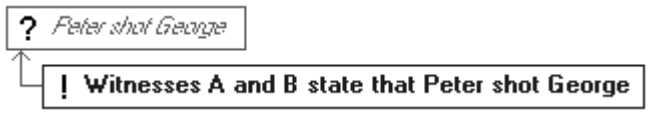
[4] Here the possibility of *withdrawing* statements is not considered. Withdrawal occurs for instance in two types of situations: 1. The statement was an assumption, that one no longer wants to assume. 2. The statement is defeated by a counterargument. In the present context of dialectical arguments, the former type of situation can be dealt with by turning an assumption into an issue, and the latter by raising the counterargument as a statement attacking the defeated statement. In neither of these withdrawal situations, a new type of argument move is needed.



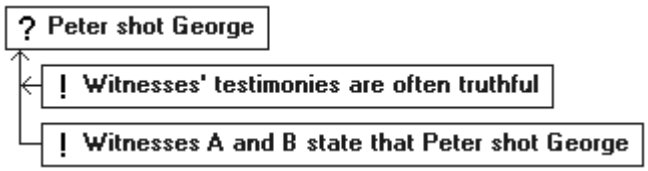
The double testimony is turned into a reason for the issue whether Peter shot George. As a result, the issue is now justified, since there is a justifying reason for it.



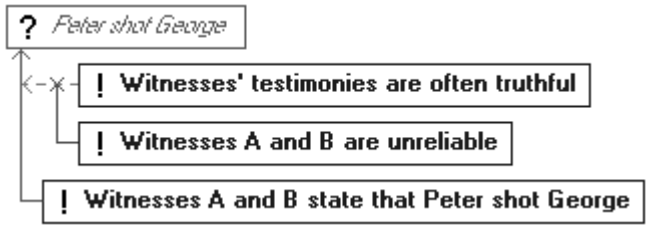
On second thoughts, the connection between the reason and the conclusion is turned into an issue (as indicated by the change in color[5]): do A and B's testimonies really justify that Peter shot George? The reason is no longer justifying, and the conclusion that Peter shot George no longer justified:



The line of argumentation continues by making explicit why A and B's testimonies would justify that Peter shot George. A backing of the supporting step is adduced, viz. that witnesses' testimonies are often truthful. As a result, it is again justified that Peter shot George.



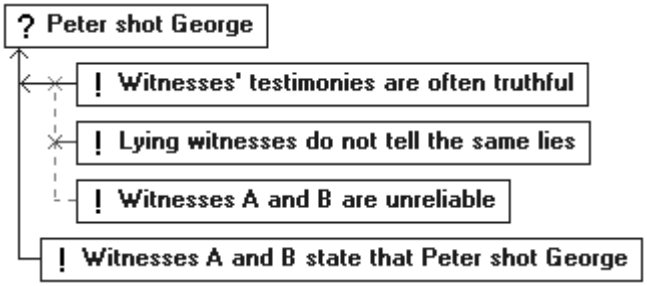
Some further investigation on the case shows that the witnesses A and B are unreliable, for instance because they have good reason to lie. As a result, the general rule that witnesses' testimonies are often truthful, does in this case not work as a backing. It does not justify that A and B's testimonies justify that Peter shot George. The issue whether Peter shot George is again unsettled: it is neither justified nor defeated. The result is the following evaluated dialectical argument:



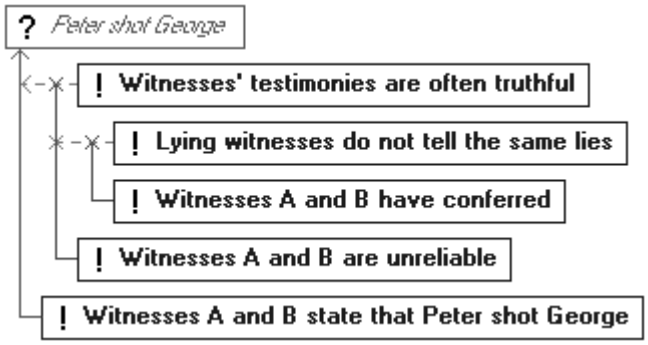
Note that the supporting connection between the backing that witnesses' testimonies are often truthful, and the warrant of the argument step that Peter shot George because of A and B's testimonies, is dotted. This indicates that the backing does not justify the warrant.

An expert states that the counterargument that the witnesses are unreliable is no good: lying witnesses do normally not tell the same lies. As a result, it is attacked and defeated that A and B's unreliability works as a counterargument. In the following figure, this is indicated by the use of a dotted connection between the attacking and the attack statement. The regular truthfulness of witnesses again works as a backing, and it is once more justified that Peter shot George.

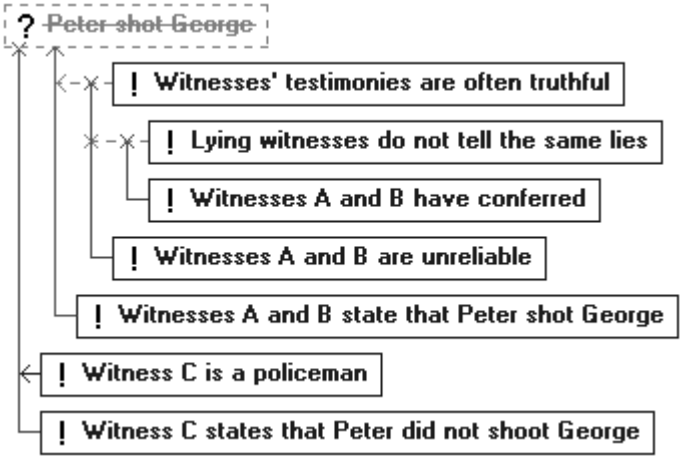
[5] In the black and white version of the paper, the change in color is unfortunately hardly visible. The color version can be obtained from the web at www.metajur.unimaas.nl/~bart/papers/.



It turns out however that the expert's opinion that lying witnesses do normally not tell the same lies, is irrelevant since the witnesses have conferred. Then it is no surprise that they tell the same stories, and if they are unreliable, they might well be equal lies. A and B's unreliability again works as a counterargument, and blocks that the regular truthfulness of witnesses can serve as a backing of the initial argument that Peter shot George since A and B testified that he did.



Finally, the issue whether Peter shot George is settled by C's testimony that Peter did not shoot George. It follows that it is defeated that Peter shot George. The following argument also shows the reason why C's testimony can settle the issue:



The evaluated dialectical argument contains for instance the following information:

- The prima facie reason that witnesses A and B state that Peter shot George, does not justify that Peter shot George.
- Since witnesses testimonies are often truthful, A and B's testimonies would normally justify that Peter shot George. This is not the case since the witnesses are unreliable.
- It could be argued that the unreliability is irrelevant, since lying witnesses do not tell the same lies. However, that counterargument is ineffective since the witnesses A and B have conferred.
- The conclusion that Peter shot George is even defeated, since witness C states that Peter did not shoot George. As a backing, it is adduced that witness C is a policeman.

During the line of argumentation sketched above, the status of the issue whether Peter shot George, changed at each stage, while it was finally settled as defeated.

The theory of dialectical argumentation sketched above can be used to formulate heuristics for courtroom decision making. Such heuristics should prescribe when a judge can stop his line of argumentation and when he needs to continue it in order to reach a better decision.

When can a line of argumentation stop? Four types of questions need to be answered positively:[6]

1. *Is any (justified) assumption sufficiently obvious?*
   If not, the assumption should be turned into an issue, and requires support of its own. In the situation of a criminal court, statements taken literally from a testimony or from a police report, often can serve as sufficiently obvious assumptions. Notice that their obviousness does not imply that they are justified: since assumptions are defeasible, they are not immune to counterarguments. Other examples of sufficiently obvious assumptions include generally agreed upon facts and rules. In the final dialectical argument above, the statement that witnesses A and B are unreliable is an example of a statement that is not sufficiently obvious and requires further support. It should therefore be turned into an issue.
2. *For any justifying or defeating statement, is it clear why it is at all supporting or attacking?*
   If not, backing of the argument step is required. A reference to a general rule phrased in a legal code or precedent can serve well as a backing of an argument step. An example of an argument step that needs further backing can be found in the final dialectical argument above: it is worthwhile to make explicit that policemen's testimonies are often truthful.
3. *For any statement that is not justified, have all statements that can support it been adduced as reasons?*
   If not, the additional reasons should be adduced. Even prima facie reasons that are considered to be not justifying, should be adduced, since it is informative to make explicit why it is non-justifying. It can for instance turn out that there is no backing or that there is an undercutting exception.
4. *For any statement that is not defeated, have all statements that can attack it been adduced as counterarguments?*
   If not, the additional counterarguments should be adduced. Also non-defeating counterarguments should be adduced in order to make explicit why it is not defeating. For instance, there could be a counter-counterargument or no backing.

Each of the actions prescribed when a question is answered negatively, can lead to changes in the evaluation. For instance, if an assumption is turned into an issue, and there is no justifying reason for it, it will no longer be justified. If a statement is not defeated, while a new counterargument is adduced, the statement can become defeated.

The main sources of information that can or even should be used in answering the four types of questions, are the law (as, e.g., written down in legal codes, treaties, and precedents), the case materials (e.g., testimonies by witnesses and experts, police reports and court pleadings) and the decision maker's own knowledge and experience.

If all four questions are answered positively, the line of argumentation can stop, and the statements justified in it can be regarded as a good decision from the point of view of the theory of dialectical argumentation. It should be noted that the resulting evaluation of the decision is not an absolute notion. Additional or deviating insights or information can change the answers to the four types of questions, and can require a continuation of the line of argumentation. For instance, new information can show that there is an unthought-of reason or exception.

As a result, the 'dialectical structure' of a legal decision is not only a tool for the legal decision maker, but also for his challengers, i.e., the prosecution, the defense, and a court of appeal. All can answer the questions for themselves, and thus find clues for undermining or strengthening the argumentation.

6 Comparison of the approaches

The theories of anchored narratives and of dialectical argumentation outlined above are closely related in several respects. In both, the problems of exceptions, conflicts and justification are addressed, and a

[6] A precondition for stopping a line of argumentation is here that the dialectical argument or arguments can be uniquely evaluated with respect to the assumptions. For present purposes, this precondition is not further discussed.

solution is sought for the problem that legal decision making is not a straightforward application of the rules of law to the case facts, but is a gradual process of 'theory construction'.

In the theory of anchored narratives (section 3), these problems are addressed by considering legal decisions as nested stories. If the story of a verdict is problematic, for instance since there is an exception or a conflict, a substory is needed, in which the problem is addressed. Such a substory will have to be more specific in either of two senses: an explanation has to be given for an assumption that was taken for granted in the original story, or a more specific rule is used as an anchor.

In the theory of dialectical argumentation (sections 4 and 5), the focus is on the dialectical arguments that underlie a legal decision. Further argumentation is required if the current dialectical arguments are problematic. It can for instance be necessary that additional supporting or attacking statements are adduced.

There are strong similarities between the two theories. Both adopt a process model of legal decision making. In the theory of anchored narratives, stories are refined by including substories, until a story of sufficient quality and anchoring is found. In the theory of dialectical argumentation, dialectical arguments are constructed, changed or extended, e.g., by making new statements or adducing new reasons and counterarguments.

In both theories, some kind of justification can be required in order to show why some conclusion is supported by a reason. The theory of anchored narratives speaks of anchors, the theory of dialectical argumentation of warrants and backings.

In both, exceptions and conflicts are cues to enhance the preliminarily constructed theory about a good decision in the case at hand. In the theory of anchored narratives, exceptions and conflicts are resolved by the use of more specific anchoring rules. For instance, because of the possible unreliability of a witness, the anchoring rule that witnesses are often truthful can be replaced by the rule that witnesses that do not benefit from their own testimony are often truthful. In the theory of dialectical argumentation, exceptions and conflicts give rise to counterarguments. The exception that a witness is unreliable can for instance be adduced as a counterargument by which the testimony by a witness is no longer assumed to be truthful.

There are also differences between the two theories. The first is that the theory of anchored narratives is story-based, while the theory of dialectical argumentation is argument-based. A second difference follows: the theory of anchored narratives is formally less explicit than the theory of dialectical argumentation. Stories are phrased in natural language, while arguments have a formal, logical structure. Third, the theory of anchored narratives is empirically backed by real, dubious cases. On the other hand, the theory of dialectical argumentation is supported by computational models, such as the argument assistance system ArguMed (Verheij, 1999). Finally, the theory of anchored narratives also discusses 'non-logical' elements of legal decision making. For instance, in the discussion of the quality of stories, particular elements of a good story are prescribed, e.g., the issues of the accused's identity, the *actus reus*, and *mens rea*. The theory of dialectical argumentation is wholly logical.

7 Conclusion

It can be concluded that the theories of anchored narratives and of dialectical argumentation, though arising from different disciplines, viz. legal psychology and artificial intelligence & law, respectively, show an interesting convergence in the problems addressed and the solutions proposed. Still the two approaches are very different. The anchored narratives approach could for instance be strengthened by describing its elements more explicitly, while the dialectical argumentation approach could gain interest by empirical testing.

We have found at least one topic, viz. the heuristics of legal decision making, in which the fields of legal psychology and artificial intelligence & law can benefit from each other's findings. As a result, Crombag's (1994) pessimism with regards to the psychological relevance of artificial intelligence for understanding the legal mind is unwarranted. Legal psychology can benefit from the higher standards of modeling following from the need of computer implementation, while artificial intelligence & law can learn from the stronger empirical orientation of legal psychology.

Appendix: a formal version of the theory of dialectical argumentation

The basis of the theory of dialectical argumentation (as described in the sections 4 and 5) is a *logical language* with two two-place connectives $\rightarrowtail$ and $\rightarrowtriangle$. The former is used to express that a statement supports another, the latter that a statement attacks another. Example sentences are $p \rightarrowtail q$, $p \rightarrowtriangle q$, $p \rightarrowtriangle (q \rightarrowtail r)$ and $p \rightarrowtail ((p \rightarrowtail q) \rightarrowtriangle (r \rightarrowtail s))$. Here p, q, r and s abbreviate logically elementary sentences.

A *(dialectical) theory* is a pair of sets of sentences (Σ, Δ), such that Δ is a subset of Σ and Σ is 'closed under disconnection', i.e., if $\varphi \rightarrowtail \psi$ or $\varphi \rightarrowtriangle \psi$ is in Σ, then φ and ψ are also in Σ. The set Σ represents the set of *statements* of the dialectical theory, the set Δ the set of (defeasible) *assumptions*. The elements of $\Sigma \setminus \Delta$ are the *issues* of the theory.

A *dialectical argument* is a set of sentences α, that is recursively defined by the following construction rules:

1. If φ is a sentence, then the singleton set $\{\varphi\}$ is a dialectical argument.
2. If α is a dialectical argument with ψ a sentence in α and φ any sentence, then $\alpha \cup \{\varphi \rightarrowtail \psi, \psi\}$ and $\alpha \cup \{\varphi \rightarrowtriangle \psi, \psi\}$ are dialectical arguments.
3. If $\alpha_0, \alpha_1, ..., \alpha_i, ...$ (where i ranges over the natural numbers) is a sequence of dialectical arguments such that, for all i, α_i is a subset of α_{i+1}, then the union of the α_i is a dialectical argument.

Note that for finite theories (i.e., with a finite set of statements) the third construction rule is not needed: any infinite sequence of dialectical arguments as it occurs in that construction rule is constant from some point onwards.

By their definition, dialectical arguments can be thought of as having a tree-like structure (cf. the figures in section 4). There are two types of 'links' between statements, viz. supporting and attacking links, expressed by sentences $\varphi \rightarrowtail \psi$ and $\varphi \rightarrowtriangle \psi$ respectively. Since the links are themselves expressed as sentences, there can be statements supporting or attacking them. If a dialectical argument can be constructed as in the definition, starting from the singleton set $\{\varphi\}$, then φ is a final conclusion of the argument. Dialectical arguments can have more than one final conclusion.

If φ is (a sentence expressing) a statement of a dialectical theory (Σ, Δ), then the *dialectical argument concerning* φ *with respect to* (Σ, Δ) is the dialectical argument with final conclusion φ, that is maximal with respect to set inclusion among the subsets of Σ.

Let T be a set of sentences and φ a sentence. Then φ *is supported by* T if there is a finite sequence of sentences $\varphi_0, ..., \varphi_n$ (for some natural number $n \geq 0$), such that φ_n is equal to φ, and each sentence φ_i in the sequence is either in T, or has predecessors φ_j and $\varphi_j \rightarrowtail \varphi_i$. The sentence φ *is attacked by* T if there is a finite sequence of sentences $\varphi_0, ..., \varphi_{n-1}, \varphi_n$ (for some natural number $n \geq 1$), such that φ_n is equal to $\varphi_{n-1} \rightarrowtriangle \varphi$, and each sentence φ_i in the sequence is either in T, or has predecessors φ_j and $\varphi_j \rightarrowtail \varphi_i$.

A *dialectical interpretation* or *extension* of a dialectical theory (Σ, Δ) is a quadruple (Σ, Δ, J, D), where J and D are subsets of Σ, such that the following hold:

0. J and D are disjoint, i.e., have no sentences in common.
1. Δ is a subset of $J \cup D$.
2. $J = \{\varphi \mid \varphi$ is supported by $\Delta \cap J\}$.
3. $D = \{\varphi \mid \varphi$ is attacked by $\Delta \cap J\}$.

The sentences in J are said to be *(dialectically) justified* in the interpretation, the sentences in D *(dialectically) defeated*.

Any dialectical interpretation of a theory (Σ, Δ) - if existing - gives rise to an evaluation of the dialectical arguments concerning the sentences in Σ, as informally discussed in section 4.

If there is a dialectical interpretation of a dialectical theory (Σ, Δ), the theory is *dialectically interpretable*. Not all dialectical theories are dialectically interpretable, and not all dialectical theories are uniquely dialectically interpretable.

A *partial dialectical interpretation* or *stage* of a dialectical theory (Σ, Δ) is a quadruple (Σ, Δ, J, D), where J and D are subsets of Σ, such that the following hold:

0. J and D are disjoint, i.e., have no sentences in common.
2. $J = \{\varphi \mid \varphi$ is supported by $\Delta \cap J\}$.
3. $D = \{\varphi \mid \varphi$ is attacked by $\Delta \cap J\}$.

The set $(J \cup D) \cap \Delta$ is the *scope* of the stage, the set $J \cup D$ its *extent*.

A set of sentences C is an *argument* if C is consistent, i.e., if there is no sentence that is both supported and attacked by C. An argument C is *incompatible* with an argument C' if the union of C and C' is not an argument. An argument C *attacks* an argument C' if C attacks a sentence in C'.

For any dialectical theory (Σ, Δ), an argument C is a (Σ, Δ)-*argument* if C is a subset of Δ. A (Σ, Δ)-argument C is *dialectically justifying* with respect to the dialectical theory (Σ, Δ) if it attacks any (Σ, Δ)-argument C' that is incompatible with C.

A sentence is *(dialectically) justifiable* with respect to a dialectical theory (Σ, Δ) if it is supported by a dialectically justifying argument. A sentence is *(dialectically) defeasible* with respect to a dialectical theory (Σ, Δ) if it is attacked by a dialectically justifying argument.

A stage (Σ, Δ, J, D) of a dialectical theory (Σ, Δ) is *dialectically justified* if $\Delta \cap J$ is a dialectically justifying (Σ, Δ)-argument with respect to (Σ, Δ).

A stage (Σ, Δ, J, D) of a dialectical theory (Σ, Δ) is *maximal* if it has maximal scope. A stage is *preferred* if it has maximal scope among the dialectically justified stages of the theory.

Any dialectical interpretation of a theory is maximal and preferred, but not vice versa. A preferred stage is not always maximal, and a maximal stage not always preferred. If a theory (Σ, Δ) is dialectically interpretable, any sentence in Δ is dialectically justifiable or defeasible, but not vice versa.

The union lemma. If C and C' are compatible dialectically justifying (Σ, Δ)-arguments, then $C \cup C'$ is a dialectically justifying (Σ, Δ)-argument.

The separation lemma. If C and C' are incompatible dialectically justifying (Σ, Δ)-arguments, then there is a sentence φ in Δ, such that C supports φ and C' attacks φ.

The importance of the separation lemma stems from the fact that the separating sentence can be chosen from among the assumptions of the theory, i.e., in Δ. Together the two lemmas show that dialectically justifying arguments are the local building blocks of dialectical interpretations, and are the key ingredients in the proof of the theorem below.

A theory has more than one preferred stage if and only if there is a sentence φ that is both dialectically justifiable and defeasible with respect to the theory. If a theory has different preferred stages, then there is an assumption of the theory, that is defeated in one preferred stage and defeated in the other.

A sentence φ is *dialectically justifiable in the context* C with respect to a dialectical theory (Σ, Δ) if there is an argument C' containing C that dialectically justifies φ with respect to (Σ, Δ). A sentence φ is *dialectically defeasible in the context* C if there is an argument C' containing C that dialectically defeats φ.

A stage (Σ, Δ, J, D) of a dialectical theory (Σ, Δ) is *disambiguating* if there is no sentence φ that is both dialectically justifiable and dialectically defeasible in the context $\Delta \cap J$.

Theorem

a. A theory has no dialectical interpretation if and only if, for any disambiguating stage (Σ, Δ, J, D), there is a sentence φ in Δ that is neither dialectically justifiable nor defeasible in the context $\Delta \cap J$.
b. A theory has a dialectical interpretation if and only if, for some disambiguating stage (Σ, Δ, J, D), all sentences φ in Δ are dialectically justifiable or defeasible in the context $\Delta \cap J$.
c. A theory has two or more dialectical interpretations if and only if there are (at least) two incompatible disambiguating stages (Σ, Δ, J, D) and (Σ, Δ, J', D'), such that all sentences φ in Δ are dialectically justifiable or defeasible, both in the context $\Delta \cap J$ and in the context $\Delta \cap J'$.

The theory above can be generalized to a language with two connectives $\rightsquigarrow$ and $\times$, the first two-place, the second one-place. A sentence $\times\varphi$ expresses that φ is defeated, and a sentence $\varphi \rightsquigarrow \times\psi$ that φ attacks ψ. A sentence $\varphi \bowtie \psi$ is seen as an abbreviation of the sentence $\varphi \rightsquigarrow \times\psi$. The resulting theory of dialectical justification and defeat is called DEFLOG (as yet unpublished).

The discussion finishes with an overview of the types of argument moves as mentioned at the start of section 5. Let (Σ, Δ) be the initial theory and (Σ', Δ') its successor. Since we do not consider withdrawing statements (cf. note 4), it holds that Σ is a proper subset of Σ.

1. *Making a statement*
   For some φ, $\Sigma' \setminus \Sigma$ is a subset of $\{\varphi\}$ and φ is not in Σ.
2. *Supporting a statement*
   For some φ and ψ, $\Sigma' \setminus \Sigma$ is a subset of $\{\varphi, \varphi \rightsquigarrow \psi, \psi\}$, while at least one of φ, $\varphi \rightsquigarrow \psi$, ψ is not in Σ.
3. *Attacking a statement*
   For some φ and ψ, $\Sigma' \setminus \Sigma$ is a subset of $\{\varphi, \varphi \bowtie \psi, \psi\}$, while at least one of φ, $\varphi \bowtie \psi$, ψ is not in Σ.

Acknowledgments

The author gladly acknowledges the financial support by the Dutch National Programme Information Technology and Law (ITeR) for the research reported in this paper (project number 01437112).

References

(*Anchored narratives*, 1993)
Wagenaar, W.A., Koppen, P.J. van, and Crombag, H.F.M. (1993). *Anchored Narratives. The Psychology of Criminal Evidence.* St Martin's Press, New York (New York).

(Crombag, 1994)
Crombag, H.F.M. (1994). On the Artificiality of Artificial Intelligence. *Artificial Intelligence and Law*, Vol. 2, pp.39-49.

(*Dubieuze zaken*, 1992, 1994)
Crombag, H.F.M., Koppen, P.J. van, and Wagenaar, W.A. (1992, 1994). *Dubieuze zaken. De psychologie van strafrechtelijk bewijs. (Dubious cases. The psychology of criminal evidence.)* Uitgeverij Contact, Amsterdam.

(Hage, 1997)
Hage, J.C. (1997). *Reasoning with Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Kluwer Academic Publishers, Dordrecht.

(Prakken, 1997)
Prakken, H. (1997). *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law.* Kluwer Academic Publishers, Dordrecht.

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

(Thagard, 1996)
Thagard, P. (1996). *Mind. Introduction to Cognitive Science.* The MIT Press, Cambridge (Massachusetts).

(Toulmin, 1958)
Toulmin, S.E. (1958). *The uses of argument.* University Press, Cambridge.

(Verheij, 1999)
Verheij, B. (1999). Automated Argument Assistance for Lawyers. *The Seventh International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, pp. 43-52. ACM, New York (New York).

(Verheij, Hage and Van Maanen, 1999)
Verheij, B., Hage, J.C., en Maanen G.E. van (1999). De logica van de onrechtmatige daad. (The logic of tort.) *Nederlands Tijdschrift voor Burgerlijk Recht*, Jg. 16, No. 4, blz. 95-102.