Argumentation (2016) 30:317–340
DOI 10.1007/s10503-015-9387-x

# Some Artificial Intelligence Tools for Argument Evaluation: An Introduction

**Douglas Walton**[1,2]

Published online: 4 November 2015


**Abstract** Even though tools for identifying and analyzing arguments are now in wide use in the field of argumentation studies, so far there is a paucity of resources for evaluating real arguments, aside from using deductive logic or Bayesian rules that apply to inductive arguments. In this paper it is shown that recent developments in artificial intelligence in the area of computational systems for modeling defeasible argumentation reveal a different approach that is currently making interesting progress. It is shown how these systems provide the general outlines for a system of argument evaluation that can be applied to legal arguments as well as everyday conversational arguments to assist a user to evaluate an argument.

**Keywords** Computational systems · Defeasible argumentation · Argument graphs · Bayesian rules · Schemes · Audience · Argument from expert opinion · Carneades Argumentation System

## 1 Introduction

Now in the field of argumentation studies there are useful tools that can be applied to the task of identifying arguments (van Eemeren and Grootendorst 2004), including argumentation schemes, and there are useful tools that can be applied to the task of analyzing arguments, namely argument diagrams, also often called argument maps (Reed et al. 2007). But so far there is no widely accepted calculative tool that can be applied to the task of evaluating everyday defeasible arguments

✉ Douglas Walton
dwalton@uwindsor.ca;
http://www.dougwalton.ca

[1] Center for Research in Reasoning, Argumentation and Rhetoric (CRRAR), 312 Essex Hall, Windsor, ON N9B 3P4, Canada

[2] Department of Philosophy, University of Windsor, Windsor, ON N9B 3P4, Canada

(Schiappa 2002). There is the literature on fallacies (Tindale 2007), but the tools provided there apply only to the more extreme kinds of cases in which an argument is so bad that it can be evaluated as committing a fallacy of a known type.

On the other hand, considerable advances have been made in the field of artificial intelligence on providing formal argumentation systems that can be used to help a person to evaluate arguments (Prakken 2010; Gordon 2010; Verheij 2014). These computational systems of argument evaluation have so far mainly been tested on legal argumentation. These systems are also technical in nature, and are not yet widely known in argumentation studies outside the community of researchers specializing in artificial intelligence and law. But the project of modeling legal argumentation bears many interesting similarities with the broader project of studying argumentation in natural language discourse generally.

Hence it is very useful at this time to try to explain in a relatively non-technical manner how these new tools might be applied to the task of argument evaluation in examples of kinds of cases that would be typical of problems of argument evaluation faced by those working in the area of natural language argumentation studies. That is the aim of this paper.

In Sect. 2, a very brief survey is given of how some argumentation systems currently being developed in artificial intelligence can be applied to the problem of argument evaluation. In Sect. 3 it is shown how argumentation schemes are used as part of the procedure for argument evaluation in these systems. In Sect. 4 a very simple example of an argument is used to illustrate how these features apply to the argument. In Sect. 5 the argument in the example is evaluated using techniques adapted from one of the computational systems. In Sect. 6 a more sophisticated example is introduced, a case used by the ancient Greek sophist Antiphon to illustrate how the prosecutor in a murder trial can construct a plausible argument to provide evidence that the defendant committed the crime. In Sect. 7, argument evaluation tools are applied to the argumentation in this example. Section 8 introduces some more advanced tools, and Sect. 9 presents some conclusions and some qualifications.

## 2 AI Systems for Argument Evaluation

Bayesian methods are widely used in artificial intelligence. The standard Bayesian method of evaluating arguments (Hahn et al. 2013) assigns numerical probability values to the components of an argument and uses Bayesian rules to give as output a numerical probability value for the strength of the argument. These include Bayesian rules defining negation, conjunction, disjunction and conditional probability. This method originated with applying such rules of estimating probabilities of outcomes in games of chance and other statistical settings. Such methods are based on the assignment of a prior probability value which is then transformed into a probability value assigned to the outcome of the operation. A statement is assigned a prior probability value between 0 and 1, and then a formula (Bayes' Rule, explained below) is used to calculate a higher or lower probability value as an outcome. A statement that is a logical tautology is assigned a probability value of 1, and a

statement that is logically inconsistent is assigned a probability value of 0. The conditional probability rule is determined by the negation and conjunction rules. According to the negation rule, the probability of $\sim A$, the negation of statement $A$, is calculated as 1 minus the prior probability of $A$. According to the conjunction rule, the probability of $A$ & $B$ ($A$ and $B$) has the probability of $A$ times the probability of $B$, assuming that $A$ and $B$ are independent of each other. According to the disjunction rule, the probability of $A$ v $B$ ($A$ or $B$) is the probability of $A$ plus the probability of $B$.

The conditional probability rule defines the probability of $B$ given $A$ as the probability of $A$ & $B$ divided by the probability of $A$. This definition can be used to derive a form of the rule for calculating conditional probability widely known as Bayes' rule, where $\Pr(A|B)$ refers to the probability of $A$ given the probability of $B$.

$$\Pr(A|B) = \frac{\Pr(B|A) \times \Pr(A)}{\Pr(B)}$$

The probability of $A$ given $B$ can be calculated from knowing the probability of $B$ given $A$ based on this rule, assuming that the prior probability values of $A$ and $B$ are known. Using this rule, an argument can be evaluated to either increase or decrease the probability of its conclusion based on assignments of probability to its premises (or leave it the same).

The Bayesian rules are widely used in many areas of science. They can also be used in some instances in legal argumentation, for example in cases of presentation of forensic evidence where probability values can be assigned by experts, and judges or juries can then try to decide on the strength of the evidence based on these numerical values. But whether Bayesian calculations could be used to evaluate arguments of the kind a judge or jury generally needs to evaluate in a trial by themselves, is a highly controversial subject in the field of artificial intelligence and law (Bench-Capon 2002). There is a worry that assigning precise probability values to premises and conclusions in such arguments might be based on a false appearance of precision that leads to artificial results and even fallacies, and that confuses juries.

Studies by social scientists have shown that argument evaluations performed on familiar kinds of arguments used in common sense reasoning diverges radically from results of applying Bayesian rules. The most famous example is the conjunction fallacy. One of the most famous examples concerning the conjunction rule is the case of Linda the bank teller (Tversky and Kahneman 1982). They tested the following example concerning judgments of conjunctive probability by posing a hypothetical case and asking people to answer the question about which outcome to choose. Suppose that Linda is a 31-year-old outspoken and very bright bank teller who majored in philosophy. In addition, suppose that it is known that Linda was concerned with issues of social justice when she was a student, and she participated in antinuclear demonstrations. Those to whom the example was described were asked which of two statements is more probable: (1) Linda is a bank teller, or (2) Linda is a bank teller and is active in the feminist movement. Most of the respondents chose answer (2). This poll appears to indicate that the respondents' way of choosing between (1) and (2) violated the Bayesian rules for conjunctive

probability. According to the Bayesian rule, the conjunction of two statements *A* and *B* is less than the probability of either *A* or *B* individually. This outcome might suggest either that those who took the poll were illogical or that the Bayesian rules for probability do not correctly represent the ways we ordinarily arrive at conclusions by logical reasoning.

While it is true that Bayesian methods are much more widely used in AI than computational models of argument, Bayesian methods are not accepted by the mainstream in the computational models of argument community, which is a subfield of artificial intelligence, as a model of argument evaluation. There has been some work on exploring relationships between Bayesian methods and computational models of argument, and there is some interest in trying to incorporate some results from Bayesian networks into computational models of argument, but this work is still in its infancy and remains outside the mainstream line of research based on systems discussed below in this paper, such as Dung Abstract Argumentation Frameworks and structured models of argument such as ASPIC+ and Carneades.

Nevertheless, because the Bayesian rules are so widely used and accepted in many scientific fields, it is very hard to challenge them as a way of rationally evaluating arguments generally. Still the question remains whether or not they can be applied to ordinary arguments such as those used in conversational argumentation in natural language, and legal argumentation, which is also expressed in natural language. Some formal computational systems being developed in artificial intelligence use the Bayesian rules, but there are others that do not. Below, some of the systems that do not need to rely on the Bayesian rules for evaluating arguments are outlined, presenting the reader with some alternatives to Bayesian systems.

The formal computational argumentation system DefLog (Verheij 2003) has an automated argument assistant called ArguMed that helps a user to construct an argument diagram for a given case (Verheij 2003, 320). ArguMed (http://www.ai.rug.nl/~verheij/aaa/argumed3.htm) is available at no cost on the Internet. DefLog is based on two primitive notions, defeasible implication and dialectical negation (Verheij 2003, 323). Dialectical negation represents the defeat of an assumption. In this system there are justified assumptions and defeated assumptions. Such a set has to meet two conditions (Verheij 2007, 197). To qualify as justified, an assumption must not be defeated by an argument having justified statements as premises. In DefLog (Verheij 2007, 187), the notion of one argument $a_1$ attacking another argument $a_2$ is modeled as an undercutting defeater in Pollock's (1995) sense, meaning that $a_1$ defeasibly implies the dialectical negation of $a_2$. It may seem strange to the reader that an argument such as $a_2$ can be negated in the system. But that is only because arguments are modeled as defeasible conditionals in DefLog and such a conditional is treated as a kind of statement.

The formal argumentation system ASPIC+ is based on a logical language *L* consisting of a set of strict and defeasible inference rules used to build arguments from a knowledge base *K*. *K* consists of a set of propositions that can be used as premises that can be combined with the inference rules to generate arguments (Prakken 2010). An example of a strict inference rule would be the deductively valid rule of *modus ponens* of classical logic. An example of a defeasible inference rule would be the argumentation scheme for argument from expert opinion: *E* is an

expert in domain *D*; *E* asserts that proposition *A*; *A* is within domain *D*; therefore *A* can be tentatively accepted subject to critical questioning. Arguments are trees containing nodes representing propositions from *L*, and lines from a set of nodes $\varphi_1$, …, $\varphi_n$ to a node $\psi$ representing an argument from premises $\varphi_1$, …, $\varphi_n$ to a conclusion $\psi$. ASPIC+ (Prakken and Sartor 1997) evaluates argumentation by using abstract argumentation frameworks (Dung 1995).

In an abstract argumentation framework, the proponent starts with an argument he wants to evaluate and when the opponent has his turn, he must provide a defeating counterargument. In such a system each argument can be attacked by other arguments, which can themselves be attacked by additional arguments. The typical result is a graph structure representing a series of attacks and counterattacks in an argumentation sequence of the following sort: $a_1$ attacks $a_2$, $a_2$ attacks $a_3$, $a_3$ attacks $a_2$, and so forth. An argument is refuted if it is attacked by any other argument that is accepted and not refuted, and is accepted only if it survives all attacks against it.

Suppose that a3 = 'We should bring back the death penalty', a2 = 'There is not enough evidence to show that the death penalty is a deterrent.' and a3 = 'Lack of evidence is not enough to prove that the death penalty is not a deterrent'. Let's say that, to begin with all three arguments are accepted, as indicated in Fig. 1, where green (which appears as gray in the printed version) in an argument node indicates acceptance.

But consider what happens next. Since a2 is accepted, and a2 attacks a3, a3 is no longer accepted. This is shown in Fig. 2. The white background indicates that a3 is not accepted.

But now consider what happens when a1 is taken into account. Argument a1 is accepted, and a1 attacks a2, so a2 is no longer accepted. But now, as shown in Fig. 3, a3 is reinstated. It is now accepted once again, since it is no longer attacked by an argument that is accepted.

A three-valued way of talking about arguments is often adopted in abstract argumentation frameworks. An argument that is accepted is said to be 'in', an argument that is rejected is said to be 'out', and an argument that is neither accepted nor rejected is said to be 'neither in nor out'.

Abstract argumentation frameworks can be extended to provide several semantics of acceptance to decide if several arguments can be accepted together. For example, a *complete extension* is a set of arguments that is able to defend itself, including all arguments it defends (van Gijzel and Nilsson 2013, 3).

The term 'graph' has many meanings, but a *graph* is defined in the mathematical field of graph theory as an ordered pair (*V*, *E*), where *E* is a subset of the two-element subsets of *V* (Harary 1972, 9). *V* is as a set of vertices, sometimes called points or nodes. *E* is a set of edges, sometimes called lines or arcs. It is customary to

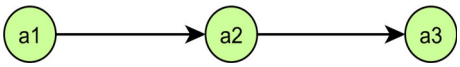


**Fig. 1** First step in an abstract argumentation framework

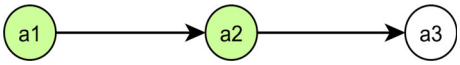


**Fig. 2** Second step in an abstract argumentation framework

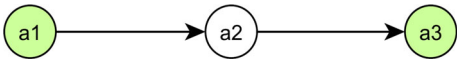


**Fig. 3** Third step in an abstract argumentation framework

represent a graph as a diagram where the nodes are joined by lines. In a directed graph, the edges have a direction associated with them. For example, in a standard argument diagram the nodes are propositions (premises or conclusions) and the lines are arrows, meant to represent inferences joining the propositions together.

The Carneades Argumentation System (CAS) (Gordon 2010)[1] was named after the Greek philosopher who advocated a fallibilistic epistemology (Thorsrud 2002). CAS models arguments as argument graphs consisting of argument nodes connected to statement nodes. Formally, an argument graph is a directed graph $\langle S, A, P, C \rangle$ consisting of four elements. $S$ is a set of statement nodes, $A$ is a set of argument nodes, $P$ is a set of premises, and $C$ is a set of conclusions. Nodes represented as rectangles in a graph represent propositions that function as premises and conclusions of arguments. Circular argument nodes in a CAS graph represent different kinds of arguments corresponding to argumentation schemes. A distinctive feature of CAS is that it distinguishes between pro and con arguments in an argument graph. A pro argument supports a conclusion or another argument. A con argument attacks a conclusion or another argument. In any CAS argument graph, one of the statements is designated at the outset as the main issue (ultimate claim being supported or contested). Newer versions of CAS argument graphs can contain cycles, such as the CAS argument graph in Fig. 4.

In CAS, argument graphs are evaluated by assuming that an audience determines whether the premises of an argument are accepted or not, and then calculates whether the conclusion should be accepted based on premises and on the argumentation scheme that forms the link joining the premises to the conclusion). Conflicts between pro and con arguments are resolved using a variety of proof standards, including preponderance of the evidence and beyond reasonable doubt (Gordon and Walton 2009).

CAS is capable of representing deductive and inductive arguments but can also use argumentation schemes to evaluate instances of defeasible arguments that do not fall into either of these categories, such as argument from expert opinion. The conclusion of a defeasible argument is only presumptively true. CAS has mainly been tested on examples of legal arguments, but may be used to model arguments in any domain. In the beginning of its development, CAS used graphs in its argument diagrams that were acyclic, meaning that they could not contain circles. Figure 8 is an example. However, the more recent versions of the model overcame this

[1] CAS is open source software, available at http://carneades.github.com/.

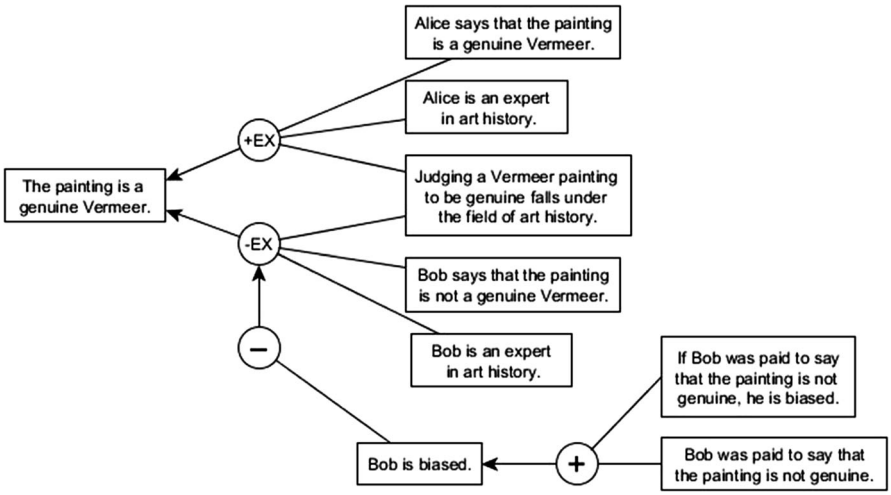


**Fig. 4** The interpretation of the expert opinion example

limitation by mapping CAS argument frameworks onto abstract argument frameworks. It has been shown that the 2007 version of CAS can be simulated using ASPIC+ (van Gijzel and Prakken 2012), but it can be conjectured that it is not an isomorphism, because it has not been shown that ASPIC+ can be simulated using CAS.

## 3 Argumentation Schemes

The argumentation scheme for the argument from expert opinion can succinctly be formulated as follows.

*Major Premise*: *E* is an expert in domain *D*.
*First Minor Premise*: *E* asserts that *A* is true.
*Second Minor Premise*: *A* is within *D*.
*Conclusion*: *A* may tentatively be accepted as true.

This scheme can also be formulated in an expanded conditional version that reveals another element of the inferential structure of the scheme.

*(P1) Conditional Premise*: If *E* is an expert in domain *D*, and *E* asserts that *A* is true, and *A* is within *D*, then *A* may tentatively be accepted as true.
*(P2) Major Premise*: *E* is an expert in domain *D*.
*(P3) First Minor Premise*: *E* asserts that *A* is true.
*(P4) Second Minor Premise*: *A* is within *D*.
*(C) Conclusion*: *A* may tentatively be accepted as true.

The expanded conditional version of the scheme has the following logical structure, where $P_1$, $P_2$ and $P_3$ and $P_4$ are meta-variables for the premises and $C$ is a meta-variable for the conclusion.

If $P_1$, $P_2$, $P_3$ and $P_4$ then $C$
$P_1$, $P_2$, $P_3$ and $P_4$
Therefore $C$

Put in this format, the scheme for argument from expert opinion looks like a substitution instance of *modus ponens* (MP) as an inference, even though it is not a deductive MP argument. It is important to emphasize that this scheme needs to be seen as defeasible in nature when taken as a representation of argument from expert opinion. The reason is that the literature on argument from expert opinion has shown that it is a form of reasoning that can be erroneous in some instances. Exploiting the tendency to take what an expert says as final has been identified with erroneous appeals to authority in which an arguer overlooks required premises of the scheme or overlooks critical questions that need to be raised (Walton 1997). However, whether such erroneous appeals are fallacies is a more complex question (Woods 2013).

To better represent the logical form of argument from expert opinion we need to see it as having a form of argument called defeasible *modus ponens* (DMP) by (Walton 2002). DMP has been adopted as a rule of inference in computational argumentation systems. Verheij (2000, 232) showed that defeasible argumentation schemes should fit a form of argument he called *modus non excipiens*: as a rule, if $P$ then $Q$; $P$; it is not the case that there is an exception to the rule that if $P$ then $Q$; therefore $Q$. Even more generally, many defeasible arguments fit this form. Consider the canonical Tweety example: If Tweety is a bird, Tweety flies; Tweety is a bird; therefore Tweety flies. Current computational argumentation systems such as ASPIC+, DefLog and CAS (the Carneades Argumentation System) use DMP as an inference rule.

Where => is the symbol for the defeasible conditional, DMP is has the following form.

*Major Premise*: $A$ => $B$
*Minor Premise*: $A$
*Conclusion*: $B$ can be tentatively accepted.

The first premise states: If $A$ is true then generally, but subject to exceptions, $B$ can tentatively be accepted as true. Following along these lines, the scheme for argument from expert opinion can now be cast into the following DMP format.

*Conditional Premise*: ($E$ is an expert & $E$ says that $A$ is true & $A$ is in $D$) => $A$.
*First Minor Premise*: $E$ is an expert.
*Second Minor Premise*: $E$ says that $A$.
*Third Minor Premise:* $A$ is in $D$.
*Conclusion*: $A$ can be tentatively accepted.

Note however that this form of the scheme is not identical to DMP because the conditional in the major premise has a conjunctive antecedent. The scheme has the following form, where the minor premises are $P_1$, $P_2$ and $P_3$.

$(P_1 \ \& \ P_2 \ \& \ P_3) \Rightarrow C$
$P_1$
$P_2$
$P_3$
Therefore $C$

But this form of argument is a substitution instance of the DMP form. So we can say that many of the most common defeasible argumentation schemes, including the argument from expert opinion, can be expressed as substitution instances of the DMP form of reasoning.

## 4 The Vermeer Example

The following example, which we will call the expert opinion example, can be used to explain in a simplified way, how argumentation is evaluated in CAS. In a forensic investigation of some potentially valuable fine art, the dispute is about whether a particular painting is a genuine Vermeer. One party to the dispute, the proponent, claims that the painting is a genuine Vermeer by citing some expert opinion evidence. She says that judging a Vermeer painting to be genuine falls under the field of art history, and Alice, an expert in art history, says that the painting is a genuine Vermeer. The other party to the dispute, the respondent, denies that the painting is a genuine Vermeer, and advances an argument to support his contention. He agrees that judging a Vermeer painting to be genuine falls under the field of art history, but cites the opinion of another expert in art history, Bob, who has claimed that the painting is not a genuine Vermeer. Here we have a pair of arguments, each one being an argument from expert opinion, that are deadlocked. The situation is often called the battle of the experts. Finally, there is a third argument to be considered. The proponent alleges that Bob is biased, and supports this allegation by claiming that Bob was paid a large sum of money to say that the painting is not genuine.

The pro-contra argument in this example is represented in the argument diagram shown in Fig. 4. The ultimate conclusion of the argument, the statement that the painting is a genuine Vermeer, is shown at the far left. At the top an argument with three premises is shown. The argument is represented by a circular node containing a plus sign. The plus sign indicates that it is a pro argument. Information about the argumentation scheme is contained within the programming of CAS, but is not shown in the nodes in Fig. 4. Nevertheless the argument at the top fits the argumentation scheme for argument from expert opinion. Just under this top argument, a second argument from expert opinion is shown, but it is a con argument as indicated by the minus sign in its argument node.

So far then, we have a pro argument from expert opinion and a con argument from expert opinion. The two arguments share a common premise, the statement that judging a Vermeer painting to be genuine falls under the field of art history. Since we have both a pro and con argument for the same conclusion at this point in the argument evaluation, it looks like the outcome might be a deadlock. But below these two arguments, there is a third argument to be considered. It is a con argument that is directed against the con argument just above it. The premise of this con argument is supported by a pro argument shown just to the right of it at the bottom of the figure. Since the proponent's side has this additional argument attacking the respondent's argument, it looks like the proponent's argument should ultimately win.

To start the procedure of evaluating the argumentation in this example let's consider the audience. Do they accept the premises of the argument or not? Let's say that the audience accepts all three of the minor premises. They accept that Alice says that the painting is a genuine Vermeer, they accept that Alice is an expert in art history, and they accept the statement that judging a Vermeer painting to be genuine falls under the field of art history. Of course they might not accept these premises. They might bring forward evidence to critically question the claim that Alice is really a certified expert in art history, by disputing Alice's credentials for example. But for the sake of keeping the example simple, let's say that the audience does accept these three minor premises. Placing these assumptions within the form of argument from expert opinion, they accept premises $P_1$, $P_2$ and $P_3$. But do they accept the conditional premise $(P_1 \ \& \ P_2 \ \& \ P_3) \Rightarrow C$? Since this premise represents the scheme for argument from expert opinion, let's say that the audience accepts this form of argumentation. For example, in a legal tribunal, expert opinion testimony is admissible as a form of evidence, even though it is a defeasible form argument that is subject to critical questioning and cross-examination.

If we look back to Fig. 4, we can see that there is a mapping from the logical form of the argumentation scheme for argument from expert opinion to its use as a pro-argument from expert opinion in the top argument shown in Fig. 4. This correspondence is shown below.

$(P_1 \ \& \ P_2 \ \& \ P_3) \Rightarrow C$ [form of the defeasible scheme for argument from expert opinion]
$P_1$ [accepted by the audience]
$P_2$ [accepted by the audience]
$P_3$ [accepted by the audience]
Therefore $C$ can be taken to be accepted by the audience.

This form of argument indicates that since the audience has accepted all four premises of the argument in this instance, because the argument is a substitution instance of DMP, the audience must also accept the conclusion $C$. Audience acceptance of the conclusion would be justified, so long as the argument has not been successfully attacked by a rebuttal, undercutter or premise defeater. The DMP form is shown below.

$(P_1 \ \& \ P_2 \ \& \ P_3) => C$
$P_1 \ \& \ P_2 \ \& \ P_3$
Therefore $C$

This inner defeasible logic is programmed into CAS, but the user can evaluate arguments with it by employing the argument mapping tool to carry out argument evaluations.

## 5 Evaluating the Argument in the Vermeer Example

CAS has developed through four main versions (see https://github.com/carneades). The first version was implemented during 2006–2008. The second version (2011) has a graphical user interface for drawing diagrams to analyze and evaluate argument, and is still available. In this version, an argument is evaluated as justifying its conclusion if the premises of the argument are acceptable (in) and the argument has not been undercut by other arguments that defeat it (Gordon and Walton 2006; Gordon et al. 2007). A more complex method of argument evaluation also available in the second version is the attaching of numerical weights to the argument representing the strength of the argument according to the audience, represented as a fraction between zero and one. In this paper the simpler method of the second version is used, for purposes of exposition, but then later the more complex method is described using a simple example. The third version of CAS is a web-based version for policy discussions, developed in 2010–2015. A fourth version, currently under development, but not yet available, evaluates arguments by two criteria: (1) whether the audience accepts the premises and (2) whether the argument properly instantiates an argumentation scheme. The previous versions cannot evaluate cumulative arguments, where new evidence can alter the acceptability value of an argument upwards or downwards, but the new version has this capability.

Next it is shown how CAS evaluates the argumentation in the Vermeer example by breaking it down into a series of steps. The first step is displayed in Fig. 5.

In Fig. 5 the three premises of the argument at the top are shown with a green background, indicating that both premises are accepted. To simplify the example, let's assume that the argument fits the requirements for the argumentation scheme for expert opinion. Put in other terms, this means that it is a defeasibly valid argument. Given that the premises are accepted and that the argument is valid, CAS automatically shows the conclusion is accepted. Hence in Fig. 6, the ultimate conclusion of the argument is shown in a text box with a green background.

Next let's turn to Fig. 6. In Fig. 6, the second argument from the top, is a con argument from expert opinion. In Fig. 6 all three premises of the con argument from expert opinion are shown as accepted, and the argument node containing the minus sign is shown with a green background as well, indicating that the premises of the argument are acceptable and the argument has not been undercut.

As shown in Fig. 6, the con argument rebuts the prior pro argument by attacking the conclusion of the pro argument. Expressed in Pollock's (1995) terminology, this argument is a rebutter, as opposed to an undercutter. (Pollock 1995) distinguished

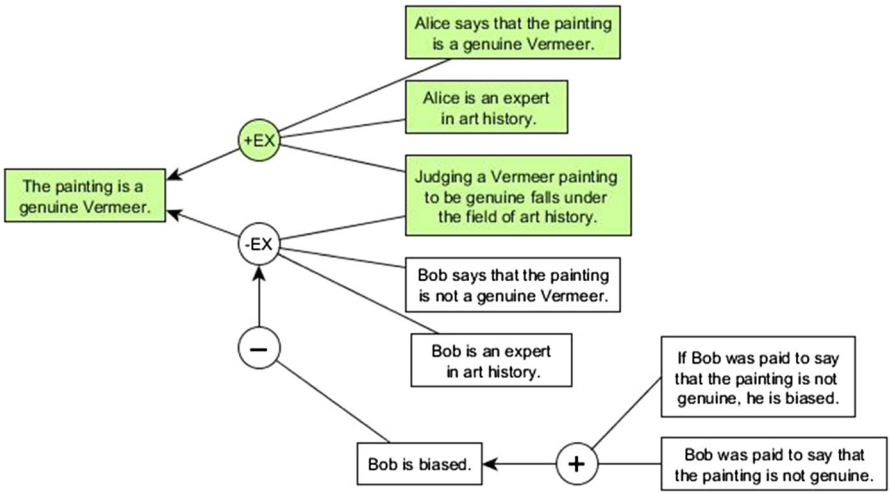


**Fig. 5** First step in evaluating the expert opinion example

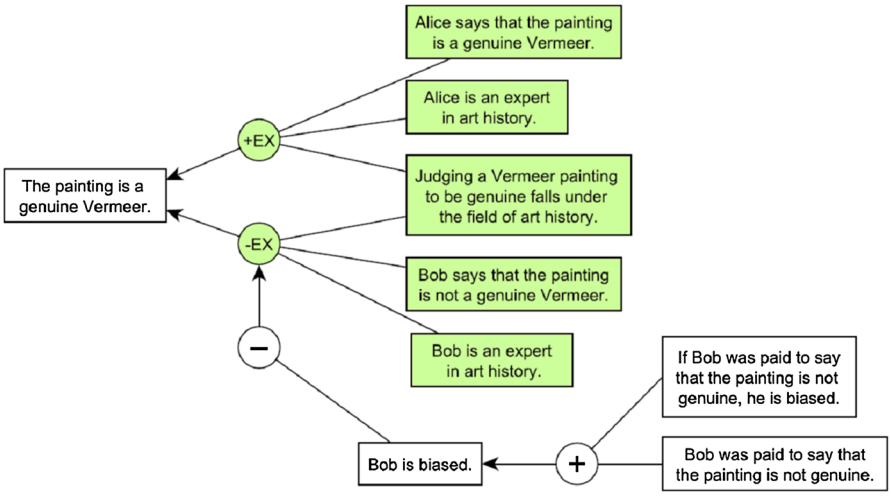


**Fig. 6** Second step in evaluating the expert opinion example

between two kinds of counter-arguments he called rebutting defeaters and undercutting defeaters (often referred to as rebutters and undercutters). A rebutter gives a reason for denying a claim by offering reasons to think it is false (Pollock 1995, 40). An undercutter attacks the inferential link between the claim and the reason supporting it by undermining the reason that supported the claim.

How this rebutter argument is a substitution instance of DMP can be shown as follows, where $P_4$ is the statement that Bob says that the painting is not a genuine Vermeer and $P_5$ is the statement that Bob is an expert in art history.

$(P_3\ \&\ P_4\ \&\ P_5) => \sim C$
$P_3\ \&\ P_4\ \&\ P_5$
Therefore $\sim C$

The situation we have now can be summed up as follows. First there was a pro argument supporting the conclusion that the painting is a genuine Vermeer. Next there was an attacking argument, a con argument directed against that same conclusion. Because there is a con argument against the conclusion, and that con argument is not only valid but also has three of its premises accepted, the pro argument above it is successfully rebutted. This means that the ultimate conclusion of the argument can no longer be accepted. So CAS shows it in a text box with a white background.

Next let's look to Fig. 7 to see what happens once the third argument is taken into account. In Fig. 7, both premises at the bottom right supporting the argument for the conclusion that Bob is biased are accepted, as indicated by each being shown in a text box with a green background. Moreover, the argument node linking these two premises to the conclusion that Bob is biased is defeasibly valid, because it fits the scheme DMP. Hence the conclusion that Bob is biased is automatically calculated by CAS as accepted.

Let's say as well that the con argument with the premise that Bob is biased is taken to be defeasibly valid, because it fits a scheme. The outcome of this situation is that the argument node shown just above this one is now shown with a white background. What has happened here is that the bottom argument about Bob being biased has undercut the con argument from expert opinion just above it. This means that one of the requirements for the argument from expert opinion in this argument has not been met, because it has been shown that the expert is biased, and therefore the argument from Bob's expert opinion has now been undercut. So this argument is

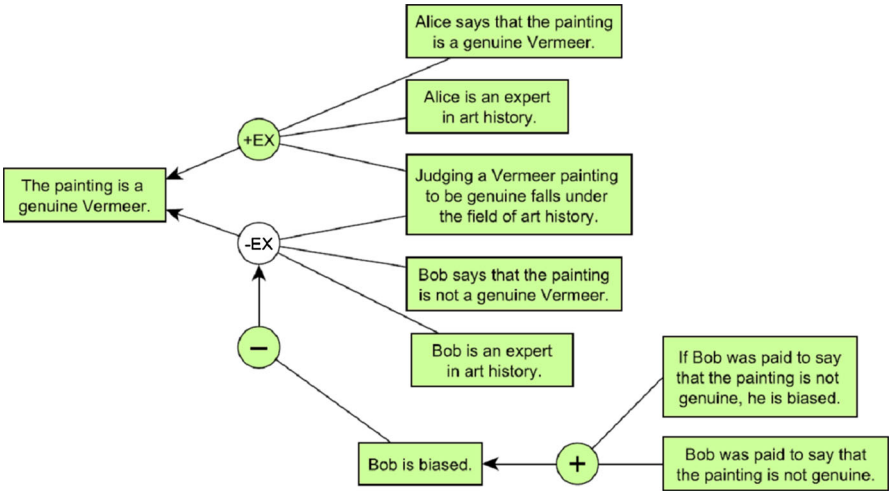


**Fig. 7** Third step in evaluating the expert opinion example

no longer applicable. Note that the counter argument that Bob is biased is not enough to defeat the argument shown above it in Fig. 6. It needs to be supported by evidence to have this effect. In other words, there is a burden of proof on the party who claims that Bob is biased to give some evidence to support her claim before the bias allegation successfully undercuts the argument from Bob's expert opinion.

To sum up, what this argument evaluation has shown, is that the deadlock between the two arguments from expert opinion has now been broken, because the second argument from expert opinion has been attacked and successfully undercut by a third argument. In other words, what has been shown is that the argument from Bob's expert opinion has been nullified, and so now it no longer successfully rebuts the argument from Alice's expert opinion. Hence the ultimate conclusion that the painting is a genuine Vermeer has been proved by the total mass of evidence that has been considered. So the change from the previous step is that the conclusion that the painting is a genuine Vermeer is now shown in a text box with a green background, as contrasted with the outcome in Fig. 6 where that conclusion was shown as not accepted.

## 6 The Antiphon Example

Plausible reasoning was known to be important in the ancient world well before the time of Carneades. The Sophists used eikotic reasoning, also called reasoning from plausibility, using the term *eikos*, meaning "what seems likely". *Eikos* is often translated as plausibility. Eikotic arguments are based on common experience (Tindale 2010, 69–82) and are defeasible, not conclusive. A statement that seems likely to be true to one person may seem likely to be false to another person, and this is especially true in legal cases where there is a conflict of opinions in a trial setting. Although Plato attacked plausible reasoning as unreliable and misleading, as part of his denunciation of the Sophists, other schools of thought, such as the Sophists and later the Skeptics, thought that plausible reasoning is all we have to go by in practical affairs of life where proof beyond all doubt is too high a standard of proof.

According to the analysis given in (Walton et al. 2014, 114), plausible reasoning has ten identifiable characteristics. Six of these are relevant here.

1. Plausible reasoning is based on common knowledge.
2. Plausible reasoning is defeasible.
3. Plausible reasoning is based on the way things generally go in familiar situations.
4. Plausible reasoning can be used to fill in implicit premises in incomplete arguments.
5. Plausible reasoning is commonly based on appearances from perception.
6. Stability is an important characteristic of plausible reasoning.

These six characteristics of plausible reasoning are illustrated in the two examples given below.

Eikotic arguments were used by Sophists to please both sides of a disputed case. The classic example in the ancient world (Gagarin 1994, 50) was the case of the larger and smaller man. In a legal case, one of the disputants in an assault case at trial was larger and stronger than the other. The smaller man argued that it was not possible that he would start the fight because it is obvious that he would get the worst of it. The larger man argued that it is not plausible that he would attack such a smaller and weaker man, because he knew that things would go badly for him if the case went to court. One of the Sophists, Antiphon, even wrote a series of manuals meant to be used as teaching tools to show to his students how to conduct pro-contra argumentation in a trial.

Another of these cases was analyzed as an example of plausible reasoning in (Walton et al. 2014, 90). In this case a slave identified the killer of a man who had been murdered, before himself dying of blows suffered during the assault. The slave had been accompanying the man after both of them had returned from a banquet. Before dying, the slave identified a known enemy of the murdered man as the perpetrator. In court, the prosecutor used the following arguments from plausibility. He argued that professional criminals would not have killed this man, because the victims were found still wearing their cloaks. This argument illustrates characteristics 3 and 4 of plausible reasoning, because generally in familiar situations, professional criminals do things for profit, so since it is an implicit premise that the cloaks would presumably have some value, professional criminals would have taken them. The prosecutor also argued that it is not plausible that someone from the banquet killed him, since he would be identified by his fellow guests. This argument is based on witness testimony, based on perception, illustrating characteristic 5. He also argued that it is not plausible that the man was killed because of a quarrel, because people would not quarrel in the dead of night and in a deserted spot. In this part of his argument, the prosecutor argued by setting up three hypotheses offering different explanations of who committed the murder, and argued that each of them is implausible.

Shifting from implausibility to plausibility, the prosecutor produced additional evidence indicating that the defendant identified by the slave was the murderer (Walton et al. 2014, 91). This factual evidence was that in the past the murdered man had brought several lawsuits against the defendant and the defendant had lost all of them at great personal expense. The prosecutor argued that the defendant bore a grudge against the victim and that for this reason it was natural for him to plot against the victim and to kill him. This argument is an example of stability (characteristic 6) because it involves consistency of actions that build up over a sequence of events. To sum up his case, the prosecutor argued "Who is more likely to have attacked him than an individual who had already suffered great injuries at his hands and could expect to suffer greater ones still?" (Diels and Kranz 1952, 87 B1: 2.1.5).

## 7 Evaluating the Argument in the Antiphon Example

The structure of the sequence of reasoning from the evidence to the ultimate conclusion is displayed in Fig. 8.

A plus (minus) sign in a circular argument node indicates a pro (con) argument. A text box with a dotted perimeter indicates that the proposition contained in it is an implicit premise, i.e. one not explicitly stated in a given text of the case. The expression +WT in the argument node on the left represents argument from witness testimony. The ultimate conclusion of the sequence of argumentation, the statement that it is plausible that D murdered V, is shown at the far left of Fig. 8. The rest of the argument diagram shows how the evidence in the case put forward by the prosecutor is used in his argumentation to support his ultimate conclusion to be proved. We don't know the defendant's argument, but presumably he offered one.

As explained in Sect. 2, CAS evaluates arguments based centrally on two factors: whether the audience accepts the premises of an argument, and whether the argument is defeasibly valid (called "applicable" in CAS). To say an argument is applicable implies that if the premises of the argument are accepted then a presumption is put in place that the conclusion of the argument should also be tentatively accepted, subject to critical questioning or to a counterargument indicating that the conclusion of the argument should not be accepted.

The witness testimony evidence is shown on the left at the top. There is an argumentation scheme for argument from witness testimony (Walton 2008, 60), and also a scheme for argument from motive to action (Walton 2011), but for simplicity we will not go into the details of how the schemes can be applied in this instance. We will assume acceptance of the two premises of the witness testimony argument, along with the two circumstantial findings shown in green (gray in the printed version) at the far right. We will also assume that the defendant has a con argument, shown as based on an acceptable premise at the bottom left of Fig. 9.

But is this argument by itself sufficient to prove the claim that D murdered V? In CAS this issue depends on the standard of proof required. This case is an ancient example so we don't know if any standard of proof was required to persuade the jury. Quite likely it was not. But assuming a reasonably high standard would be required, and assuming the defendant puts up any argument, even a sufficiently weak one to raise some doubt, the witness testimony argument is not sufficient by itself to prove the ultimate conclusion.

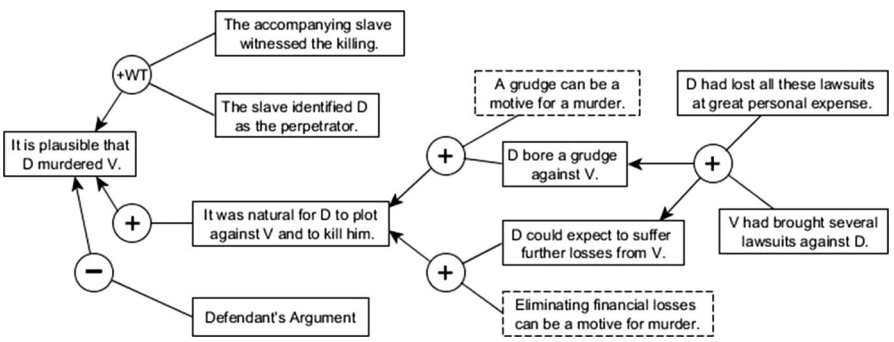


**Fig. 8** Interpretation of the argument in the Antiphon case

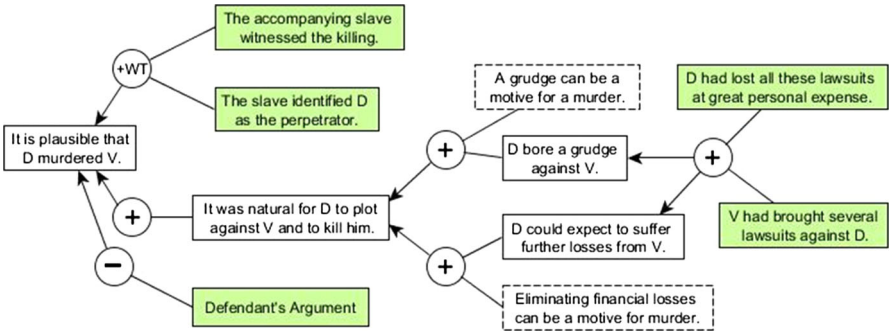


**Fig. 9** Step 1 of evaluating the argument in the Antiphon case

Conflicts between pro and con arguments are resolved using proof standards, such as preponderance of the evidence or clear and convincing evidence (Gordon and Walton 2009). The proof standards are not defined numerically, but using thresholds α and β, as follows (Gordon and Walton 2009, 245): The preponderance of the evidence standard for a proposition *p* is met if and only if there is are last one applicable argument pro *p*, and the maximum weight assigned by the audience to the applicable arguments pro *p* is greater than the maximum weight of the applicable arguments con *p*. The clear and convincing evidence standard is met if and only if (1) the preponderance of the evidence standard is met (2) the maximum weight of the applicable pro arguments exceeds some threshold α, and (3) the difference between the maximum weight of the applicable pro arguments and the maximum weight of the applicable con arguments exceeds some threshold β.

Now let's go on to examine the other evidence in the case. The circumstantial evidence is shown on the right. It is composed of two statements that are used as premises in two arguments that lead to two separate conclusions. One is the statement that D bore a grudge against V. The other is the statement that D could expect to suffer further losses from V. CAS has the capability of using the same premise over again in a different argument. In this instance, it uses the same two premises over again in two different arguments. These are different not only because they have different conclusions, but also because the inferential links represented by their argument nodes represent two different kinds of arguments. Next let's see how to evaluate this argument.

Consider the argument as shown in Fig. 10. Both of the statements shown at the far right are accepted, because both premise statements are parts of the factual evidence in the case. Both these statements are shown in green text boxes, indicating that each of them has been accepted by the audience. Let's also say that both of these arguments are defeasibly valid (applicable).

Once the circumstantial evidence is brought forward, it supports the conclusion that D bore a grudge against V, and it supports the conclusion that D could expect to suffer further losses from V. Hence both of these statements are shown in green boxes in Fig. 10. But what about the two implicit generalizations contained in the boxes with dashed borders?

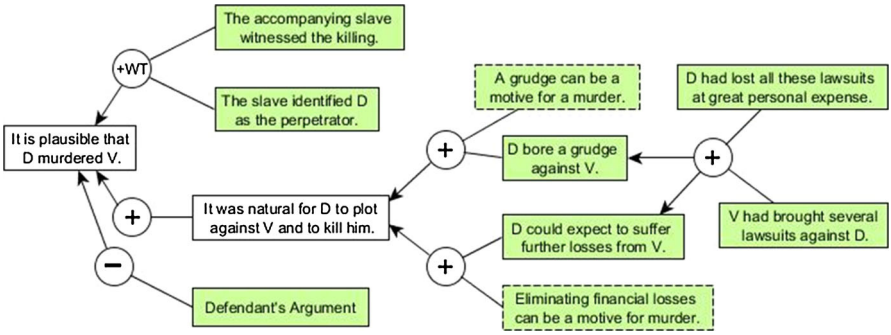


**Fig. 10** Step 2 of evaluating the argument in the Antiphon case

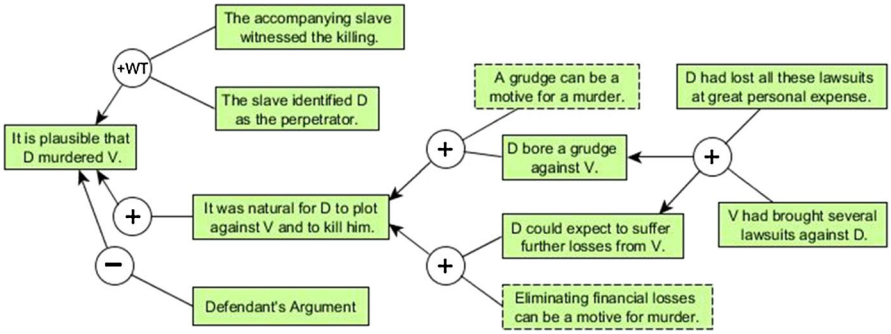


**Fig. 11** Step 3 of evaluating the argument in the Antiphon case

Both of these two propositions would be acceptable to the audience as evidence, and both of these motive arguments are applicable. The resulting evidential situation is shown in Fig. 11. As shown in Fig. 11, CAS automatically shows the proposition that it was natural for D to plot against V and to kill him in a green text box, indicating that this proposition is accepted, based on the argument supporting it. Now the prosecution's two main arguments are strong enough to offset the defendant's argument, assuming it is taken to very weak, and the prosecution's argument is strong enough to meet the beyond reasonable doubt standard.

The outcome of the complete evaluation is that all the arguments in the case, once marshaled together in the way shown in Figs. 8, 9, 10 and 11, provide enough evidence to prove the conclusion that it is plausible that D murdered V. So now CAS will automatically show the ultimate conclusion, the statement that D murdered V, in green. This outcome depends on how the network of argumentation in the case is structured as a directed graph as displayed in these four argument diagrams, and on the definitions of the four standards of proof defined in CAS, as indicated in Sect. 8.

## 8 More Advanced Argument Evaluation Tools

So far, the examples used to illustrate CAS argument evaluations have been kept relatively simple, for purposes of easy exposition. However, it may also be interesting to explain two further tools that CAS offers that can optionally be used to make more sophisticated evaluations. One is the use of proof standards and the other is the numerical weighting of arguments.

We have seen that there is a way of evaluating deadlocks, and it was also mentioned in the Antiphon example that standards of proof can be used for this purpose. But how this works can be more fully explained by defining the four proof standards more precisely (Gordon and Walton 2009). These proof standards are applied using thresholds $\alpha$ and $\beta$ (Gordon and Walton 2009, 245). The four standards of proof are defined as follows. The preponderance of evidence (PE) standard is met if at least one pro argument is accepted (*in*) that weighs more than any *in* con argument. Only arguments pro or con this statement are compared.The *clear and convincing evidence* standard is met if and only if (1) the preponderance of the evidence standard is met (2) the maximum weight of the applicable pro arguments exceeds some threshold $\alpha$, and (3) the difference between the maximum weight of the applicable pro arguments and the maximum weight of the applicable con arguments exceeds some threshold $\beta$. It is left up to the user to input numbers or other comparative values into $\alpha$ and $\beta$. The default in the user interface is the standard of the preponderance of the evidence, but the user is given the option of changing to a different standard of proof as required.

Another feature available in CAS is that of attaching numerical weights to the arguments in an argument graph. The numerical weights represent the strength of the argument, as determined by the audience, represented by a fraction between zero and one. Consider the example shown in Fig. 12. Argument $a_2$, shown at the bottom, has both premises accepted. The audience accepts this argument with strength of 0.4. But there is a counterargument, con argument $a_1$. The sole premise of this argument $p_3$ is not accepted by the audience. So at this stage, the pro argument wins, and so the ultimate conclusion $p_1$ is shown by CAS as accepted. In Figs. 12 and 13,

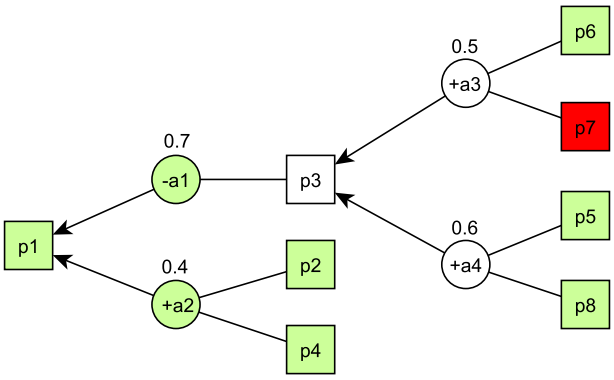


**Fig. 12** First stage of evaluation of example with weights

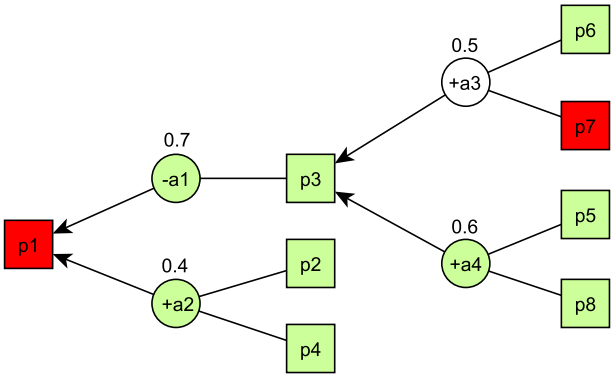


**Fig. 13** Second stage of evaluation of example with weights

green (lighter gray in the printed version), means 'accepted', red (darker gray in the printed version) means 'rejected', and white denotes 'neither accepted nor rejected' (neither in nor out).

But let's take a closer look at the con argument $a_1$. It has two arguments supporting the premise $p_3$, namely $a_3$ and $a_4$. Argument $a_3$ is of no use to support $p_3$, because one of its premises $p_7$ has been rejected by the audience. However, argument $a_4$ has both of its premises accepted by the audience. Take a look at Fig. 13 to see how the evaluation proceeds from this point.

Since argument $a_4$ has both its premises accepted, this argument is applicable, and therefore its conclusion $p_3$ is shown in a green box. Since the premise of the con argument $a_1$ (namely $p_3$) has now been accepted, there is one applicable pro argument and one applicable con argument. What breaks the deadlock is that the con argument $a_1$ (shown as a rebutter) is stronger than the pro-argument $a_2$. Hence the ultimate conclusion $p_1$ is refuted.

It should be noted that the computational argument evaluation systems surveyed in this paper are currently still under development, and rapidly being improved. For example, a new version of CAS will be available shortly that enables a user to evaluate cumulative arguments. This is a type of argument that has already been evaluated as somewhat plausible, but needs to be re-evaluated as new evidence comes in. For example a series of tests may be carried out, and after each test the argument may be re-evaluated as more plausible or less plausible. This feature is especially important for evaluating abductive reasoning used when a hypothesis is conjectured on the basis of some evidence, but needs to be re-evaluated as new experimental evidence that bears on it comes to be known.

Before the advent of this feature, CAS was unable to deal adequately with cumulative arguments. A cumulative argument is one where there is a buildup of evidence that either supports the plausibility of a given hypothesis based on pro arguments or detracts from its possibility based on con arguments. The snake and rope example, the leading example used by the philosopher Carneades to illustrate plausible reasoning, is an instance of cumulative argumentation. In this example (Walton et al. 2014, 12) a man sees what looks like a coil of rope in a dimly lit

room. Based on his perception of how it appears, but also on his inability to view the object clearly in the dim room, he draws the plausible hypothesis that the object is a snake. Reasoning on this hypothesis, he jumps over the object. When he looks back after jumping, he sees that the object remained immobile. At this point he accepts the hypothesis that the object is a rope. But there is a third step in the sequence. He prods the object with a stick and sees it remains immobile. He takes this finding to confirm his hypothesis that the object is a rope (Walton et al. 2014).

The current version of CAS does not support the evaluation of cumulative argumentation, and although a new research project is underway to build a version of CAS that has this capability, the results have not been published yet.

## 9 Conclusions

Application of the Bayesian method to the evaluation of arguments in legal and everyday conversational arguments takes place by assigning probability values to the subjective belief of the arguer. This approach is basically a solitary one, and it confronts the problem of other minds. How can I tell what another agent's beliefs are, since I have no direct access to them? Beliefs, desires and intention are called internal "mental states". In contrast, systems such as CAS are acceptance-based. In the language of argumentation theory, the arguments are based on what one party takes to be the commitments of the other party. The term 'commitment', derived from (Hamblin 1970) is close to, or equivalent to the notion of acceptance (Cohen 1992). CAS evaluates arguments based on input on what the audience of the argument accepts, or does not. The audience and the arguer are two distinct entities in the system, and so this system of argument evaluation is more social than individualistic in approach.

There remains the option, however, that the two approaches could be somehow combined. Pollock (1995, 95) was opposed to Bayesianism, a view in which reasons make their conclusions probable to varying degrees, and a conclusion is justified only if it is made sufficiently probable as evidence accumulates. Bayesianism, on his account, is associated with probabilism, the view that degrees of justification obey the probability calculus. Cohen (1977) argued against probabilism, contending that for some inductive arguments one needs some nonstandard principles of the probability calculus. Pollock (1995, 99) argued that degrees of justification do not work like probabilities. Verheij (2014) has formally modeled Pollock's theory of undercutters, defeaters, and argument strength within probability theory, and put forward a modified version of the Bayesian calculus that retains Bayes' theorem, but rejects the conjunction rule for defeasible reasoning even though he maintains that it holds for deductive (conclusive) reasoning. The conjunction rule, as noted in Sect. 2, states that the probability of $A$ & $B$ equals the probability of $A$ multiplied by the probability of $B$, provided that $A$ and $B$ are independent. Verheij uses the example of witness testimony evidence to argue that this rule does not work for presumptive reasoning. When witness $W_1$ testifies that suspect $S_1$ committed a crime and witness $W_2$ testifies that suspect $S_2$ committed the crime, even though each statement is presumptively supported by the argumentation scheme for witness

testimony, the conclusion '$S_1$ committed the crime and $S_2$ committed the crime' is not presumptively supported.

Given this development, it may become possible to deal with the conjunction fallacy explained in Sect. 2 while retaining a variant of the Bayesian conditional probability rule while excluding the conjunction rule as applied to presumptive reasoning of the kind used in legal argumentation and argumentation in natural language discourse generally. Once this issue is sorted out, there may be some way of combining a Bayesian approach with the general approach of the methods of argument evaluation outlined in this paper. In the meantime, it has been shown that there is a general method of argument evaluation emerging from AI that can be applied to evaluating arguments in legal reasoning and in everyday conversational discourse, independently of the Bayesian rules for the standard probability calculus. This independent methodology, it has been argued in this paper, stands on its own, even though it is dependent on the user's ability to analyze a given argument using argument diagramming methods.

Both DefLog and CAS, in addition to being formal computational systems for argument evaluation, have argument diagramming (mapping) tools that can be used to assist a user making an argument diagram to evaluate a given argument. However, these tools are currently under testing and development, and at this stage they are not as easy for beginners to use as many other argument mapping tools that are available, such as Rationale, Araucaria, ArguNet and so forth. See (Scheuer et al. 2010) for an extensive review of argument mapping tools and a comparative description of their features. What is important is not so much the particular tool used to draw an argument diagram as understanding the features of the logical and computational system for argument evaluation associated with the drawing tool.

When working on an example argument to be evaluated, a good practical approach is to start by drawing a rough argument diagram with pencil and paper and then later use a mapping tool, or even a simple graph drawing software tool, to build a more refined version that can be stored, reused, sent to others and later modified. There are drawing tools available that are easy to learn about and use such as Microsoft Visio, yEd, Gliffy, LucidChart, and so forth, and some are free. These tools can be used to draw quite a presentable argument diagram in conjunction with the argument evaluation systems currently being built in artificial intelligence described in this paper. By manually inserting notations for schemes, audience acceptance and so forth, an argument diagram showing how a given argument can be evaluated can be constructed.

But the aim of this paper is not to show how to use any particular argument mapping software. The aim has been to explain enough of the logical and computational structure underlying the use of such tools for argument evaluation to enable a user to gain some idea of how to go about evaluating arguments. For those who have not been able to successfully apply Bayesian rules to examples of natural language arguments or legal arguments they wish to evaluate, the method of argument evaluation outlined in this paper offers an alternative.

Applying the general method of argument evaluation outlined in this paper depends on the ability to interpret natural language texts to reconstruct arguments and other informal logic skills. The methods described in this paper are not

completely automatic or mechanical, and cannot be. The tools presented in this paper are argument assistants. Applying these computational methods requires six basic informal logic skills: (1) a knowledge of argumentation schemes (2) the ability to represent the structure of a given argument using an argument diagram in the form of a directed graph, (3) the ability to apply the schemes to the diagram, (4) the ability to fill in gaps in the diagram created by implicit premises or conclusions, and (5) the ability to use the device of the audience as applied to the given case to judge which premises are accepted by the audience. In addition, (6) the ability to apply differing standards of proof to the argumentation in a given case may be needed in some instances, as illustrated in the Antiphon example.

**Acknowledgments** Thanks are due to the Social Sciences and Humanities Research Council of Canada for Insight Grant 435-2012-0104: The Carneades Argumentation System. I would also like to thank Tom Gordon for many helpful comments, criticisms and corrections that led to significant improvements in the paper.

## References

Bench-Capon, T. 2002. Agreeing to differ: Modelling persuasive dialogue between parties without a consensus about values. *Informal Logic* 22(3): 231–245.

Cohen, L.J. 1977. *The Probable and the Provable*. Oxford: Oxford University Press.

Cohen, L.J. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon Press.

Diels, H., and W. Kranz. 1952. *Die Fragmente der Vorsokratiker*. Berlin: Weidmannsche Verlagsbuchhandlung.

Dung, P. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2): 321–357.

Gagarin, M. 1994. Probability and persuasion: Plato and early Greek rhetoric. In *Persuasion: Greek Rhetoric in Action*, ed. Ian Worthington, 46–68. London: Routledge.

Gordon, T.F. 2010. *The Carneades Argumentation Support System, Dialectics, Dialogue and Argumentation*, ed. C. Reed, and C. W. Tindale, London: College Publications.

Gordon, T.F., and D. Walton. 2006. The Carneades argumentation framework. In *Computational Models of Argument: Proceedings of COMMA 2006*, ed. P.E. Dunne, and T.J.M. Bench-Capon, 195–207. Amsterdam: IOS Press.

Gordon, T.F., H. Prakken, and D. Walton. 2007. The Carneades model of argument and burden of proof. *Artificial Intelligence* 171: 875–896.

Gordon, T.F., and D. Walton. 2009. Proof burdens and standards. In *Argumentation in Artificial Intelligence*, ed. I. Rahwan, and G. Simari, 239–260. Berlin: Springer.

Hahn, U., A. Harris, and M. Oaksford. 2013. Rational argument, rational inference. *Argument and Computation* 4(1): 21–35.

Hamblin, C.L. 1970. *Fallacies*. London: Methuen.

Harary, F. 1972. *Graph Theory*. Menlo Park: Addison-Wesley.

Pollock, J.L. 1995. *Cognitive Carpentry*. Cambridge, Mass.: The MIT Press.

Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument & Computation* 1(2): 93–124.

Prakken, H., and G. Sartor. 1997. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics* 7: 25–75.

Reed, C., D. Walton, and F. Macagno. 2007. Argument diagramming in logic, law and artificial intelligence. *Knowledge Engineering Review* 22: 87–109.

Scheuer, O., F. Loll, N. Pinkwart, and B.M. McLaren. 2010. Computer-supported argumentation: a review of the state of the art. *Computer-Supported Collaborative Learning* 5(1): 43–102.

Schiappa, E. 2002. Sophisticated modernism and the continuing importance of argument evaluation. In *Arguing Communication and Culture: Selected Papers from the 12th NCA/AFA Conference on*

*Argumentation*, ed. G.T. Goodnight, 51–58. Washington, DC: National Communication Association.
Thorsrud, H. 2002. Cicero on his academic predecessors: The Fallibilism of Arcesilaus and Carneades. *Journal of the History of Philosophy* 40: 1–18.
Tindale, C. 2010. *Reason's Dark Champions: Constructive Strategies of Sophistic Argument*. Columbia, South Carolina: The University of South Carolina Press.
Tindale, C. 2007. *Fallacies and Argument Appraisal*. Cambridge: Cambridge University Press.
Tversky, A., and D. Kahneman. 1982. Judgments of and by representativeness. In *Judgment Under Uncertainty: Heuristics and Biases*, ed. D. Kahneman, P. Slovic, and A. Tversky. Cambridge: Cambridge University Press.
Van Eemeren, F., and R. Grootendorst. 2004. *A Systematic Theory of Argumentation*. Cambridge: Cambridge University Press.
Van Gijzel, B., and H. Prakken. 2012. Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. *Argument and Computation* 3(1): 21–47.
Van Gijzel, B., and H. Nilsson. 2013. *Towards a Framework for the Implementation and Verification of Translations Between Argumentation Models*. ACM. http://www.cs.ru.nl/P.Achten/IFL2013/symposium_proceedings_IFL2013/ifl2013_submission_27.pdf.
Verheij, B. 2000. Logic, Context and Valid Inference Or: Can There be a Logic of Law? Available on bart.verheij@metajur.unimaas.nl.
Verheij, B. 2003. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation* 13(3): 319–346.
Verheij, B. 2007. Argumentation support software: Boxes-and-arrows and beyond. *Law, Probability and Risk* 6: 187–208.
Verheij, B. (2014). Arguments and their strength: Revisiting Pollock's anti-probabilistic starting points. *Proceedings of COMMA 2014*, to appear. http://comma2014.arg.dundee.ac.uk/res/pdfs/43-verheij.pdf.
Walton, D. 1997. *Appeal to Expert Opinion*. University Park: Penn State Press.
Walton, D. 2002. Are some *modus ponens* arguments deductively invalid? *Informal Logic* 22(1): 19–46.
Walton, D. 2008. *Witness Testimony Evidence*. Cambridge: Cambridge University Press.
Walton, D. 2011. Teleological argumentation to and from motives. *Law, Probability and Risk* 10(2011): 203–223.
Walton, D. 2015. *Argument Evaluation and Evidence*. Cham: Springer.
Walton, D., C.W. Tindale, and T.F. Gordon. 2014. Applying recent argumentation methods to some ancient examples of plausible reasoning. *Argumentation* 28(1): 85–119.
Woods, J. 2013. *Errors of Reasoning: Naturalizing the Logic of Inference*. London: College Publications.