The Carneades Argumentation Framework

Using Presumptions and Exceptions to Model Critical Questions

Thomas F. Gordon [a,1], and Douglas Walton [b]
[a] *Fraunhofer FOKUS, Berlin*
[b] *Dept. of Philosophy, University of Winnipeg, Manitoba, Canada*

Abstract. We present a formal, mathematical model of argument structure and evaluation, called the Carneades Argumentation Framework, which applies proof standards [5] to determine the defensibility of arguments and the acceptability of statements on an issue-by-issue basis. Carneades uses three kinds of premises (ordinary premises, presumptions and exceptions) and information about the dialectical status of statements (undisputed, at issue, accepted or rejected) to model critical questions in such a way as to allow the burden of proof to be allocated to the proponent or the respondent, as appropriate.

Keywords. Argumentation Framework, Argument Evaluation, Argumentation Schemes, Critical Questions

1. Introduction

The work in this paper flows from previous attempts to solve a key problem common to AI and argumentation theory concerning the using of the device of critical questions to evaluate an argument. Critical questions were first introduced by Arthur Hastings [11] as part of his analysis of presumptive argumentation schemes. The critical questions attached to an argumentation scheme enumerate ways of challenging arguments created using the scheme. The current method of evaluating an argument that fits a scheme like that for argument from expert opinion is by a shifting of the burden of proof from one side to the other in a dialog [23]. When the respondent asks one of the critical questions matching the scheme, the burden of proof shifts back to the proponent's side, defeating or undercutting the argument until the critical question has been answered successfully. At this has been the general approach of argumentation theory. Recently, however, it was observed [2] that critical questions differ with respect to their impact on the burden of proof. These observations led to two theories about the shifting of the burden of proof when critical questions are asked. According to one theory, when any critical question is asked, the burden shifts to the proponent's side to answer the question and, if no answer is given, the argument fails. According to the other theory, merely asking a critical question

[1] Correspondence to: Dr Thomas F. Gordon, Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, Berlin, Germany. E-mail: thomas.gordon@fokus.fraunhofer.de

is not enough to shift the burden of proof back to the proponent. On this theory, to make the argument fail, the question needs to be supported by further argument. Some critical questions fit one theory better, while others fit the other theory better. This duality has posed a recurring problem for the project of formalizing schemes.

In this paper, we put forward a new model for evaluating defeasible arguments that solves this problem, continuing work we began in 2005 [7,20]. The current paper presents a formal, mathematical model of argument evaluation which applies proof standards [5] to determine the defensibility of arguments and the acceptability of statements on an issue-by-issue basis. The formal model is called the Carneades Argumentation Framework, in honor of the Greek skeptic philosopher who emphasized the importance of plausible reasoning [4, vol. 1, p. 33–34].

The formal model has been implemented, using a functional programming language. The implementation, also called Carneades, is being developed to support a range of argumentation tasks, including argument construction ("invention"), evaluation and visualization, primarily for use in the legal domain.

Arguments in Carneades are identified, analyzed and evaluated not only by fitting premise conclusions structures that can be identified using argumentation schemes. Arguments also have a dialectical aspect, in that they can be seen as having been put forward on one side or the other of an issue during a dialog. The evaluation of arguments in Carneades depends on the stage of the dialog. Whether or not a premise of an argument holds depends on whether it is undisputed, at issue, or decided. One way to raise an issue is to ask a critical question. Also, the proof standard applicable for some issue may depend on the stage of the dialog. In a deliberation dialog, for example, a weak burden of proof would seem appropriate during brainstorming, in an early phase of the dialog. The Carneades Argumentation Framework is designed to be used in a layered model of dialectical argument [15] for various kinds of dialogs, where higher layers are responsible for modeling such things as speech acts, argumentation protocols and argument strategies.

The rest of the paper is structured as follows. Section 2 defines the structure of arguments and illustrates this structure with examples from related work by Toulmin, Pollock and others. Section 3 formally defines how arguments are evaluated in terms of the acceptability of statements, the defensibility of arguments, and the satisfiability of proof standards. The paper closes in Section 4 with a presentation of conclusions and ideas for future work.

2. Argument Structure

We begin by defining the structure of arguments. Unlike Dung's model [3], in which the internal structure of arguments is irrelevant for the purpose of determining their defensibility, our model makes use of and depends on the more conventional conception of argument in the argumentation theory literature, in which arguments are a kind of conditional linking a set of premises to a conclusion. Intuitively, the premises and the conclusion of arguments are statements about the world, which may be accepted as being true or false. In [7] the internal structure of statements was defined in such a way as to enable the domain of discourse to be modeled in a way compatible with emerging standards of the Semantic Web [1]. These details, however, need not concern us here. For the purpose of evaluating arguments, the internal structure of statements is not important. We only require the ability to compare two statements to determine whether or not they are equal.

Definition 1 (Statements) *Let* $\langle \mathrm{statement}, = \rangle$ *be a structure, where* $\mathrm{statement}$ *denotes the set of declarative sentences in some language and* $=$ *is an equality relation, modeled as a function of type* $\mathrm{statement} \times \mathrm{statement} \rightarrow \mathrm{boolean}$.

Next, to support defeasible argumentation and allow the burden of proof to be distributed, we distinguish three kinds of premises.

Definition 2 (Premises) *Let* $\mathrm{premise}$ *denote the set of premises. There are three kinds of premises:*

1. *If s is a* $\mathrm{statement}$*, then* $\mathrm{premise}(s)$ *is a premise. These are called* ordinary premises*. As a notational convenience, we will use a statement s alone to denote* $\mathrm{premise}(s)$*, when the context makes it clear that the statement is being used as a premise.*
2. *If s is a* $\mathrm{statement}$*, then $\bullet s$, called a* presumption*, is a premise.*
3. *If s is a* $\mathrm{statement}$*, then $\circ s$, called an* exception*, is a premise.*
4. *Nothing else is a premise.*

Now we are ready to define the structure of arguments.

Definition 3 (Arguments) *An* $\mathrm{argument}$ *is a tuple $\langle c, d, p \rangle$, where c is a* $\mathrm{statement}$*,* $d \in \{\mathrm{pro}, \mathrm{con}\}$ *and* $p \in \mathcal{P}(\mathrm{premise})$*. If a is an argument $\langle c, d, p \rangle$, then* $\mathrm{conclusion}(a) = c$*,* $\mathrm{direction}(a) = d$ *and* $\mathrm{premises}(a) = p$*. Where convenient, pro arguments will be notated as $p_1, \ldots, p_n \rightarrow c$ and con arguments as $p_1, \ldots, p_n \multimap c$.*

This approach, with two kinds of arguments, pro and con, is somewhat different than the argument diagramming model developed by Walton in [21] and implemented in Araucaria. There counterarguments are modelled as arguments pro some statement which has been asserted to be in conflict with the conclusion of the other argument, called a *refutation*. Our approach, with its two kinds of arguments, is not uncommon in the literature on defeasible argument [14,17,10,8].

We assume arguments are asserted by the participants of a dialog. We have specified and implemented a simple communication language and argumentation protocol to test Carneades, but that is a subject for another paper. For our purposes here, it is sufficient to note that argument moves, i.e. speech acts, are modelled as functions which map a state of the dialog to another state. (Again, this is a purely functional model, so states are not modified.) A dialog state is a tuple $\langle t, h, G \rangle$, where t is a statement, the *thesis* of the dialog, h is a sequence of moves, representing the history of the dialog, and G is an *argument graph*.[1]

It is these argument graphs which concern us here. An argument graph plays a role comparable to a set of formulas in logic. Whereas in logic the truth of a formula is defined in terms of a (consequence) relation between sets of formulas, here we will define the *acceptability* of statements in argument graphs. An argument graph is not merely a set of arguments. Rather, as its name suggests, it is a finite graph. There are two kinds of

[1] In prior work [6,8], Gordon has referred to argument graphs as *dialectical graphs*.

nodes, statement nodes and argument nodes. The edges of the graph link up the premises and conclusions of the arguments. Each statement is represented by at most one node in the graph.

To illustrate argument graphs, suppose we have the following (construed) arguments from the domain of contract law:

a1. $\mathrm{agreement}, \circ\,\mathrm{minor} \rightarrow \mathrm{contract}$

a2. $\mathrm{oral}, \bullet\,\mathrm{estate} \multimap \mathrm{contract}$

a3. $\mathrm{email} \rightarrow \mathrm{oral}$

a4. $\mathrm{deed} \multimap \mathrm{agreement}$

a5. $\bullet\,\mathrm{deed} \rightarrow \mathrm{estate}$

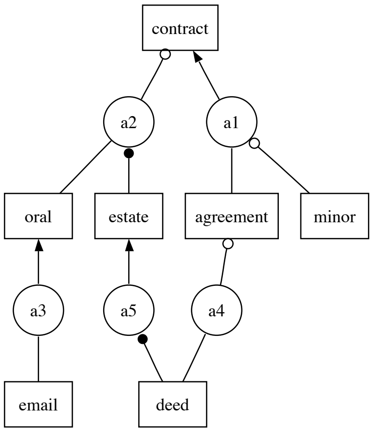

Figure 1. Argument Graph

The argument graph induced by these arguments is shown in Figure 1. In this figure, statements are displayed as boxes and arguments as circles. Different arrowhead shapes are used to distinguish pro and con arguments as well as the three kinds of premises. Pro arguments are indicated using ordinary arrowheads; con arguments with open-dot arrowheads. Ordinary premises are represented as edges with no arrowheads, presumptions with closed-dot arrowheads and exceptions with open-dot arrowheads. (The direction of the edge is implicit in the case of ordinary premises; the direction is always from the premise to the argument.) Notice that the premise type cannot be adequately represented using statement labels, since argument graphs are not restricted to trees. A statement may be used in multiple arguments and as a different type of premise in each argument. The above example illustrates this point. The fourth and the fifth arguments each use the statment g in a premise. In the fourth argument it is used in an ordinary premise but in the fifth it is used in a presumption. Walton has called this use of shared premises a *divergent argument structure* [21, p. 91].

Although argument graphs are not restricted to trees, they are not completely general; we do not allow cycles. This restriction assures the decidability of the defensibility and acceptabilty properties of arguments and statements, respectively.

Definition 4 (Argument Graphs) *An* argument-graph *is a labeled, finite, directed, acyclic, bipartite graph, consisting of* argument *nodes and* statement *nodes. The edges link the argument nodes to the statements in the premises and conclusion of each argument.*

This completes the formal definition of the structure of arguments and argument graphs. Let us now discuss briefly the expressiveness of this model, beginning by comparing our approach with Toulmin's model [16]. Recall that arguments in Toulmin's model consist of a single premise, called the *datum*; a conclusion, called the *claim*; a kind of rule, called the *warrant*, which supports the inference from the premise to the conclusion of the argument; an additional piece of data, called *backing*, which provides support for the warrant; an exception, called a *rebuttal*; and, finally, a *qualifier* stating the probative value of the inference (e.g. presumably, or necessarily). Of these, the datum and conclusion are handled in a straightforward way in our model. The set of premises of an argument generalizes the single datum in Toulmin's system. Claims are modeled comparably, as conclusions. Rebuttals are modeled with con arguments. The probative weight of an argument is handled as part of our model of proof standards, as will be explained shortly.

This leaves our interpretation of warrants and backing to be explained. Our model does not directly allow arguments about other arguments. (The conclusion of an argument must be a statement.) Rather, the approach we prefer is to add a presumption for the warrant to the premises of an argument. If an argument does not have such a presumption, the argument graph can first be extended to add one. We leave it up to the argumentation protocol of the procedural model to regulate under what conditions such *hidden premises* may be *revealed*. In effect, the datum and warrant are modelled as minor and major premises, much as in the classical theory of syllogism. Backing, in turn, can be modelled as a premise of an argument supporting the warrant.

For example, here is a version of Toulmin's standard example about British citizenship.

Datum. Harry was born in Bermuda.

Claim. Harry is a British subject.

Warrant. A man born in Bermuda will generally be a British subject.

Backing. Civil Code §123 provides that persons born in Bermuda are generally British subjects.

Exception. Harry has become an American citizen.

The argument can be reconstructed in our framework as illustrated if Figure 2.

This approach generalizes Toulmin's model, by supporting arguments pro and contra both warrants and backing, using the same argumentation framework as for arguments about any other kind of claim. Indeed, Toulmin appears to have overlooked the possibility of arguing against warrants or making an issue out of backing claims.

Our model of argument is rich enough to handle Pollock's concepts of rebuttal, premise defeat and undercutting defeaters [14]. Rebuttals can be modeled as arguments in the opposite direction for the same conclusion. (If an argument a_1 is pro some statement s, then some argument a_2 con s is a rebuttal of a_1, and vice versa.) Premise defeat can be modeled with arguments con an ordinary premise or presumption, or pro an exception.

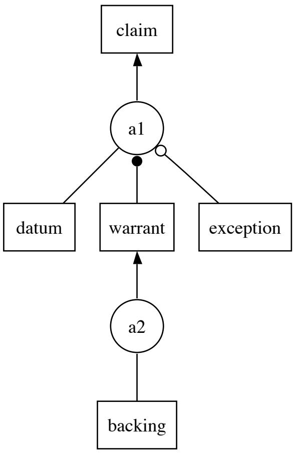

Figure 2. Reconstruction of Toulmin Diagrams

Undercutting defeaters are a bit trickier. The idea of an undercutting defeater is to argue against the argument itself, or the rule or warrant which was applied to create the argument. We model undercutting defeaters by revealing and then attacking premises, similar to the way we handled warrants in the reconstruction of Toulmin's system. Consider Pollock's example of things which look red but turn out to be illuminated by a red light:

Red. The object is red.

Looks Red. The object looks red.

Applicable. The general rule "Things which look red are red." applies to this object.

Illuminated. The object is illuminated by a red light.

An argument graph for this example is shown in Figure 3. Rather than undercutting argument a_1 (the object is red because it looks red) directly, with an argument contra a_1, we undercut the argument by first revealing a presumption (about the general rule being applicable in this case) and then assert an argument contra this presumption. Notice by the way that another presumption is still implicit in this example, namely a presumption for the "warrant" about things which look red being red.

Walton [21] distinguishes two kinds of arguments, called *convergent* and *linked* arguments. Convergent arguments provide multiple reasons for a conclusion, each of which alone can be sufficient to accept the conclusion. Convergent arguments are handled in our approach by multiple arguments for the same conclusion. Linked arguments, on the other hand, consist of two or more premises which all must hold for the argument to provide significant support for its conclusion. Linked arguments are handled in our approach by defining arguments to consist of a set of premises, rather than a single premise, and defining arguments to be defensible only if all of their premises hold. (The concept of argument defensibilty is formally defined below.)

Presumptions and exceptions are a refinement of Walton's concept of *critical questions* [22]. Critical questions enumerate specific ways to defeat arguments matching some argumentation scheme. But so long as an issue has not been raised by actually asking

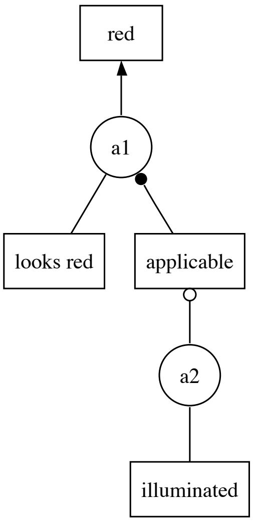

Figure 3. Undercutting Defeater Example

some critical question, we would like to be able to express which answer to presume. The distinction between presumptions and exceptions here provides this ability.

Consider the scheme for arguments from expert opinion [19]:

Major Premise. Source E is an expert in the subject domain S containing proposition A.
Minor Premise. E asserts that proposition A in domain S is true.
Conclusion. A may plausibly be taken as true.

The scheme includes six critical questions:

CQ1. How credible is E as an expert source?
CQ2. Is E an expert in the field that A is in?
CQ3. Does E's testimony imply A?
CQ4. Is E reliable?
CQ5. Is A consistent with the testimony of other experts?
CQ6. Is A supported by evidence?

When the scheme for arguments from expert opinion is instantiated to create a specific argument, the critical questions can be represented, in our model, as presumptions and exceptions. Whether a presumption or exception is appropriate depends on the burden of proof. If the respondent, the person who poses the critical question, should have the burden of proof, then the critical question should be modeled as an exception. If, on the other hand, the proponent, the party who used the schema to construct the argument, should have the burden of proof, then the critical question should be modeled as a presumption.[2]

[2] We agree with Verheij [18] that critical questions which are entailed by the premises of the argument schema are redundant and may be omitted. This is arguably the case in the example for the first three critical questions.

Our model does not require that premises for critical questions be made explicit at the time the argument is first made. Rather, they can be *revealed* incrementally during the course of the dialog. The conditions under which a premise may be left implicit or revealed raise procedural issues which need to be addressed in the protocol for the type of dialog. Our contribution here is to provide an argumentation framework which can be used for modeling such protocols.

3. Argument Evaluation

By argument evaluation we mean determining whether a statement is *acceptable* in an argument graph. As we will see soon, this in turn will depend on the *defensibility* of arguments in the graph. Notice that our terminology is somewhat different than Dung's [3], who speaks of the acceptability of arguments, rather than their defensibility. Also, for those readers familiar with our preliminary work on this subject in [7], please notice that the terminology and other details of the current model are different, even though the basic ideas and general approach are quite similar.

The definition of the acceptability of statements is recursive. The acceptability of a statement depends on its *proof standard*. Whether or not a statement's proof standard is *satisfied* depends on the defensibility of the arguments pro and con this statement. The defensibility of an argument depends on whether or not its premises *hold*. Finally, we end up where we began: Whether or not a premise holds can depend on whether or not the premise's statement is acceptable. Since the definitions are recursive, we cannot avoid making forward references to functions which will be defined later.

To evaluate a set of arguments in an argument graph, we require some additional information. Firstly, we need to know the current *status* of each statement in the dialog, i.e. whether it is accepted, rejected, at issue or undisputed. This status information is pragmatic; the status of statements is set by speech acts in the dialog, such as asking a question, asserting an argument or making a decision. Secondly, we assume that a proof standard has been assigned to each statement. We do not address the question of how this is done. Presumably this will depend on domain knowledge and the type of dialog. In the following, let $\{\mathrm{SE}, \mathrm{PE}, \mathrm{DV}, \mathrm{BRD}\}$ be an enumeration of some proof standards. Their meaning will be defined shortly, below. Finally, one of these proof standards, PE, makes use of numerical weights, comparable to conditional probabilities. To use this proof standard, we require a weighing function.

Let us formalize these requirements by postulating an *argument context* as follows.

Definition 5 (Argument Context) *Let $\mathcal{C}$, the argument context, be a tuple* $\langle G, \mathrm{status}, \mathrm{proof\text{-}standard}, \mathrm{weight}\rangle$*, where G is an* $\mathrm{argument\text{-}graph}$*,* status *is a function of type* $\mathrm{statement} \rightarrow \{\mathrm{accepted}, \mathrm{rejected}, \mathrm{undisputed}, \mathrm{issue}\}$*,* $\mathrm{proof\text{-}standard}$ *is a function of type* $\mathrm{statement} \rightarrow \{\mathrm{SE}, \mathrm{PE}, \mathrm{DV}, \mathrm{BRD}\}$ *and* weight *is a function of type* $\mathrm{statement} \times \mathrm{statement} \rightarrow \{0, \ldots, 10\}$

Intuitively, a statement which has been used in a dialog is initially $\mathrm{undisputed}$. Later in the dialog, an issue can be made out of this statement. Presumably after arguments pro and con have been collected for some period of time, a decision will be taken and the statement will be either $\mathrm{accepted}$ or $\mathrm{rejected}$. The details of how this is done need not concern us further here. These are matters which need to be addressed fully when modeling protocols for dialogs.

Definition 6 (Acceptability of Statements) *Let* $\mathrm{acceptable}$ *be a function of type* $\mathrm{statement} \times \mathrm{argument\text{-}graph} \rightarrow \mathrm{boolean}$. *A statement is acceptable in an argument graph if and only if it satisfies its proof standard in the argument graph:* $\mathrm{acceptable}(s, ag) = \mathrm{satisfies}(s, \mathrm{proof\text{-}standard}(s), ag)$.

Definition 7 (Satisfaction of Proof Standards) *A proof standard is a function of type* $\mathrm{statement} \times \mathrm{argument\text{-}graph} \rightarrow \mathrm{boolean}$. *Let f be a proof standard.* $\mathrm{satisfies}(s, f, G) = f(s, G)$

Four proof standards are defined in this paper. We do not claim these particular proof standards are exhaustive. Others can be defined similarly.

SE. A statement meets this standard iff it is supported by at least one defensible pro argument.

PE. A statement meets this standard iff it is supported by at least one defensible pro argument and its strongest defensible pro argument outweighs its strongest defensible con argument, if any. This standard balances arguments using probative weights.

DV. A statement meets this standard iff it is supported by at least one defensible pro argument and none of its con arguments are defensible.

BRD. A statement meeets this standard iff it is supported by at least one defensible pro argument, all of its pro arguments are defensible and none of its con arguments are defensible.

The names of three of these standards are meant to suggest three legal proof standards: scintilla of evidence, preponderance of the evidence and beyond a reasonable doubt. However, we do not claim that the definitions of these standards, above, fully capture their legal meanings. What these standards have in common with their legal counterparts is their relative strength: $\mathrm{BRD} > \mathrm{DV} > \mathrm{PE} > \mathrm{SE}$. If a statement satisfies a proof standard, it will also satisfy all weaker proof standards.

The name of the DV proof standard is an acronym for *dialectical validity*, a term used by Freeman and Farley [5]. They defined five proof standards. In addition to the four we have defined here, they included a fifth, called *beyond a doubt*, which was defined to be an even stronger standard than *beyond a reasonable doubt*.

The preponderance of evidence (PE) standard compares the weight of arguments. The weight of an argument is defined to be the same as the weight of its *weakest premise*, i.e., to be precise, the same as the weight of the premise with the lowest weight. Recall we assume a weighing function, weight, as part of the context to provide this information. The weight of a premise p for a conclusion c is $\mathrm{weight}(p, c)$. Other proof standards which aggregate and compare weights are conceivable. For example, one could sum the weights of the arguments pro and con and compare these sums.

We have defined weights to be natural numbers in the range of 0 to 10. We originally considered using real numbers in the range of 0.0 to 1.0, as in probability theory. However, on the assumption that the weights will be estimated by human users, we prefer to use a simpler ordinal scale, since we are skeptical that users can estimate such weights with a greater degree of accuracy.

All of the proof standards defined above depend on a determination of the *defensiblity* of arguments. Defensibility is defined next.

Definition 8 (Defensibility of Arguments) *Let* defensible *be a function of type* $\text{argument} \times \text{argument-graph} \rightarrow \text{boolean}$. *An argument* α *is defensible in an argument graph G if and only if all of its premises* hold *in the argument graph:* $\text{defensible}(\alpha, G) = \text{all}(\lambda p.\, \text{holds}(p, G))(\text{premises}\, \alpha)$.[3]

Finally, we come to the last definition required for evaluating arguments, for the holds predicate. This is where the status of a statement in the argument context and the distinction between ordinary premises, presumptions and exceptions come into play. Accepted presumptions and ordinary premises hold. Rejected presumptions and ordinary premises do not hold. Undisputed presumptions hold. Undisputed ordinary premises do not hold. An exception, $\circ s$, holds only if $\text{premise}(s)$ does not hold.

Definition 9 (Holding of Premises) *Let* holds *be a function of type* $\text{premise} \times \text{argument-graph} \rightarrow \text{boolean}$. *Let* $\sigma = \text{status}(s)$. *Whether or not a premise holds depends on its type (ordinary, presumption, or exception). Thus, there are the following three cases:*

If p is an ordinary premise, $\text{premise}(s)$*, then*

$$\text{holds}(p, G) = \begin{cases} true & \textit{if}\ \sigma = \text{accepted} \\ false & \textit{if}\ \sigma = \text{rejected} \\ acceptable(s, G) & \textit{if}\ \sigma = \text{issue} \\ false & \textit{if}\ \sigma = \text{undisputed} \end{cases}$$

If p is a presumption, $\bullet s$*, then*

$$\text{holds}(p, G) = \begin{cases} true & \textit{if}\ \sigma = \text{accepted} \\ false & \textit{if}\ \sigma = \text{rejected} \\ acceptable(s, G) & \textit{if}\ \sigma = \text{issue} \\ true & \textit{if}\ \sigma = \text{undisputed} \end{cases}$$

Finally, if p is an exception, $\circ s$*, then*

$$\text{holds}(p, G) = \neg\, \text{holds}(\text{premise}(s), G)$$

The important thing to notice is that whether or not a premise holds depends in this model not only on the arguments which have been asserted, but also on the kind of premise (ordinary, presumption, or exception) and the status of the premise's statement in the argument graph (undisputed, at issue, accepted, or rejected). We assume that the status of a statement progresses in the course of the dialog:

[3] Here 'all' is a higher-order function, not a quantifier, applied to an anonymous function, represented with λ, as in lambda calculus.

1. Initially, statements used in arguments are undisputed. Whether or not a premise which uses this statement holds at this stage of the dialog depends on the kind of premise. Ordinary premises do not hold; presumptions do hold. *This is the only semantic difference between ordinary premises and presumptions in our model.* An exception holds at this stage only if it would not hold if it were an *ordinary premise*. Notice that exceptions are not the dual of presumptions. As undisputed presumptions hold, an undisputed exception would not hold if we had defined exceptions to hold only if they would not hold if they were presumptions. But this is not the semantics we want. Rather, both undisputed exceptions and undisputed presumptions hold.
2. At some point a participant may make an issue out of a statement. Now ordinary premises and presumptions which use this statement hold only if they are acceptable, i.e. only if the statement meets its proof standard, given the arguments which have been asserted. Exceptions at issue hold only if the statement is not acceptable. We presume that arguments will be exchanged in a dialog for some period of time, and that during this phase the acceptability of statements at issue will be in flux.
3. Finally, at some point a decision will be made to either accept or reject some statement at issue. The model does not constrain the discretion of users to decide as they please. Unacceptable statements may be accepted and acceptable statements may be rejected. This remains transparent however. Any interested person can check whether the decisions are justified given the arguments made and the applicable proof standards. Anway, after a decision has been made, it is respected by the model: Accepted statements hold and rejected statements do not hold, no matter what arguments have been made or what proof standards apply.

4. Conclusion

The Carneades Argumentation Framework is a formal, mathematical model of argument evaluation which applies proof standards to determine the defensibility of arguments and the acceptability of statements on an issue-by-issue basis. The main original contribution of Carneades is its use of three kinds of premises (ordinary premises, presumptions and exceptions) and information about the dialectical status of statements (undisputed, at issue, accepted or rejected) to model critical questions in such a way as to allow the burden of proof to be allocated to the proponent or the respondent, as appropriate. Both of these elements are required for this purpose: presumptions hold without supporting arguments only so long as they have not been put at issue by actually asking the critical question.

The Carneades Argumentation Framework is a semantic model of argumentation, not a calculus. However, since this semantic model is formulated in terms of computable functions, the lambda calculus may be used a formal system for deriving inferences and functional programming languages can be used to implement the model in software. Indeed, the formal model has been fully implemented in this way, using the Scheme programming language [12][4] As a semantic model, the question of the soundness or completeness of Carneades does not arise. Rather, the relevant question concerns the valid-

[4] An earlier version of Carneades was implemented in Standard ML [13].

ity of the semantic model. Do these semantics sufficient for providing the kind of argumentation support required by our intended application scenarios? This question cannot be answered by purely formal means, but rather requires experiments with realistic test cases. We have already tested Carneades on a number of examples from the Artificial Intelligence and Law literature, thus far yielding intuitively acceptable results. In a companion paper, also in this volume [9], Carneades has been used to model and evaluate the arguments in the majority and dissenting opinions of Pierson vs. Post, an important benchmark in the Artificial Intelligence and Law field. More work is required to validate the models of the various proof standards, in particular the model of prepondernance of the evidence, which uses weights. For this purpose, we plan to reconstruct examples of reasoning with evidence. Although our test cases thus far have all been legal, Carneades is intended to be a general model of argumentation, not restricted to some application domain. Outside the legal context, we plan to evaluate the suitability of Carneades for practical reasoning in the context of deliberation dialogs.

When completed, Carneades will support a range of argumentation use cases, including argument construction, evaluation and visualization. Although the focus of this paper was argument evaluation, it contains some hints about the direction we are heading to support argument visualization. One of our next tasks will be to refine the diagramming method used here to illustrate the argumentation framework.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.

[2] F. Bex, H. Prakken, C. Reed, and D. Walton. Towards a formal account of reasoning with evidence: Argumentation schemes and generalizations. *Artificial Intelligence and Law*, 11(2-3), 2003.

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

[4] P. Edwards. *The Encyclopedia of Philosophy*, volume 1. Macmillian and Free Press, 1972.

[5] K. Freeman and A. M. Farley. A model of argumentation and its application to legal reasoning. *Artificial Intelligence and Law*, 4(3-4):163–197, 1996.

[6] T. F. Gordon. The Pleadings Game — an exercise in computational dialectics. *Artificial Intelligence and Law*, 2(4):239–292, 1994.

[7] T. F. Gordon. A computational model of argument for legal reasoning support systems. In P. E. Dunne and T. Bench-Capon, editors, *Argumentation in Artificial Intelligence and Law*, IAAIL Workshop Series, pages 53–64. Wolf Legal Publishers, 2005.

[8] T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, pages 10–18. Melbourne, Australia, 1997.

[9] T. F. Gordon and D. Walton. Pierson vs. Post revisted — a reconstruction using the Carneades Argumentation Framework. In P. E. Dunne, editor, *Proceedings of the First International Conference on Computational Models of Argument (COMMA 06)*, Liverpool, 2006. IOS Press.

[10] J. Hage. A theory of legal reasoning and a logic to match. *Artificial Intelligence and Law*, 4(3-4):199–273, 1996.

[11] A. C. Hastings. *A Reformulation of the Modes of Reasoning in Argumentation*. PhD thesis, Northwestern University, Evanston, Illinois, 1963.

[12] R. Kelsey, W. Clinger, and J. Rees. Revised (5) report on the algorithmic language Scheme. *Higher-Order and Symbolic Computation*, 11(1):7–105, August 1998.

[13] R. Milner, M. Tofte, R. Harper, and D. MacQueen. *The Definition of Standard ML, (Revised).* MIT Press, 1997.

[14] J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[15] H. Prakken. From logic to dialectic in legal argument. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, pages 165–174. Maryland, 1995.

[16] S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.

[17] B. Verheij. *Rules, Reasons, Arguments. Formal Studies of Argumentation and Defeat*. Ph.d., Universiteit Maastricht, 1996.

[18] B. Verheij. *Virtual Arguments*. TMC Asser Press, The Hague, 2005.

[19] D. Walton. *Argumentation Methods for Artificial Intelligence in Law*. Springer, 2005.

[20] D. Walton and T. F. Gordon. Critical questions in computational models of legal argument. In P. E. Dunne and T. Bench-Capon, editors, *Argumentation in Artificial Intelligence and Law*, IAAIL Workshop Series, pages 103–111, Nijmegen, The Netherlands, 2005. Wolf Legal Publishers.

[21] D. N. Walton. *Argument Structure : a Pragmatic Theory*. Toronto studies in philosophy. University of Toronto Press, Toronto ; Buffalo, 1996. Douglas Walton. ill. ; 24 cm.

[22] D. N. Walton. *Argumentation Schemes for Presumptive Reasoning*. Erlbaum, 1996.

[23] D. N. Walton. *Appeal to Expert Opinion*. Penn State Press, University Park, 1997.