From systems for defeasible argumentation to dialogical systems of argumentation

Nonmonotonic reasoning is a reasoning in which temporary conclusions can be drawn on the basis of incomplete information but which might be withdrawn when more information becomes available. Systems for defeasible argumentation capture this kind of reasoning in terms of interactions between conflictual arguments. Nonmonotonic reasoning is explained in terms of defeasibility since arguments can be defeated by stronger counterarguments. This appendix gives a general account of the main approaches to defeasible argumentation. Our guideline (mainly for the first two sections) will be a recently published survey\(^1\) by H. Prakken and G. A. Vreeswijk devoted entirely to the subject. In section 1, we present the basic semantic concepts of systems for defeasible argumentation. Section 2 focuses on the proof theoretical way in which individual arguments can be exchanged in a defeasible framework. In section 3, further analysis of the interaction between individual arguments reveals that the dynamical aspect can better be captured by combining argument-based systems with speech act-based dialogues. The embedding of systems for defeasible systems in formal dialogue systems gives rise to a new approach which can be called defeasible dialogical systems of argumentation. Section 5 illustrates how this new approach can be used effectively to formalise fundamental notions such as the notion of legal justification.

1. Basic concepts of systems for defeasible argumentation.

1.1. The notion of argument.

The basic notion of a system for defeasible argumentation is that of argument. The system does not however define the notion but rather presupposes it. That is why systems for defeasible argumentation need a logical language which specifies the formation rules of arguments. The notion of logical consequence of this underlying language is monotonic in the sense that new premises cannot invalidate an argument as such but can only give rise to a counterargument. Nonmonotonicity arises when the added new premises give rise to a stronger counterargument which defeats the original one. The kind of

---

logical language to be adopted depends on the structure of arguments to be captured. But most systems allow the use of two kinds of inference in the formation of arguments: monotonic rules and nonmonotonic rules. Argumentation systems are set up relatively to an input theory. The propositions expressed by the statements contained in the input can then be regarded as the premises of the system. And the rules of inference specify how new arguments can be constructed from the premises. As a result, an argument can be defined either as a sequence of rules grounded in the premises or, more generally, as a tree of such inferences since an argument can be formed by the combination of several independent subarguments. An argument is defeasible if it contains at least one defeasible rule otherwise it is a strict argument. Some systems impose no condition on the structure of arguments. They merely demand that the conclusion be implied by the premises. In such systems, an argument is just a premises-conclusion pair. The connection between the premises and the conclusion is often called a reason.

1.2. Relation between arguments.

As mentioned above, an argument stays an argument if more premises are added. If adding new premises cannot invalidate an argument, the latter can however be disputed by another argument. It is thus important to know when arguments are in conflict. We will present briefly the ways in which two arguments can be in conflict with each other.

(i) Rebutting an argument. This type of conflict occurs when arguments have contradictory conclusions as in the following example.

Example 1 (called the Tweety example).

A  Tweety flies because it is a bird
B  Tweety does not fly because it is a penguin

(ii) Assumption attack. We have an attacking assumption when one argument makes a non-provability assumption (as in default logic) and another argument proves what was assumed unprovable by the first argument. Adding new information may provide evidence to the contrary of a previously acceptable default conclusion which has now to be withdrawn. For example, “Tweety flies because it is a bird and it is not provable that Tweety is a penguin” is attacked by an argument the conclusion of which is “Tweety is a penguin”.
(iii) Undercutting an inference. What is common between the first two kinds of conflict is that arguments attack statements. The third type of conflict is an attack not on a statement but on a rule of inference. Only non-deductive rules can be subject to an inference attack. Argumentation systems can thus allow the use of non-deductive reasoning such as inductive, abductive or analogical arguments. Here is an example of a non-deductive argument given by Prakken and Vreeswijk: “raven_{101} is black since the observed ravens raven_1 … raven_{100} were black”, this inductive reasoning is undercut by the argument “I saw raven_{102} which was white”.

Since arguments can be in conflict with each other, argumentation systems use the binary relation of defeat aimed at comparing them. The defeat relation defined over a set of all arguments Args forms an argumentative framework (Args, defeat). But how can we decide when an argument defeats another? In Example 1, we have no reason to believe that Tweety flies or Tweety does not fly. In other words, we cannot decide, on the basis of the given two arguments, whether A defeats B or B defeats A. This example shows that the conflict between arguments cannot be resolved if we rely only on the defeat relation. But there is a well-established tradition in nonmonotonic logic which consists in giving priority to specific arguments over the rest. By applying what is known in nonmonotonic logic as the specificity rule to the Tweety example, we may prefer A because it is more specific than B. The assumption here is that the more specific a piece of information is, the more reliable it is. But it is wrong to conclude that the notion of specificity is a general principle of common sense reasoning. To decide between competing arguments, several argumentation systems introduce a priority relation designed to order the set of all arguments. Specificity appears to be just one of the many criteria which might or might not be used. Some researchers (like Prakken and Sartor [96]) suggest the priority criteria should be defined by using the information from the context and not by introducing general context-independent principles since it is hard to find such general principles of defeat applicable to all domains. This opens the way for the priorities to be themselves the subject of defeasible reasoning since information about the criteria could be incomplete or inconsistent.

1.3. The status of arguments.

An important limitation of the notion of defeat is the following. Since the defeat relation can only determine the strength of two individual arguments, the ultimate status of an argument, which depends on the interaction between all the available arguments, cannot be evaluated.
Example 1 (continued). To the arguments A and B of example 1, let’s add the following third argument

C the observation that Tweety is a penguin is not reliable

We have concluded earlier that B defeats A (thanks to the specificity criterion). And it is clear that C defeats B. According to the defeat relation, A has the status of a defeated argument. But if we take into account the interaction between all the three arguments, the ultimate status of A cannot be regarded as a defeated argument. By defeating its defeater B, C reinstates the status of A. Generally, the definition of status divides the set of all arguments into three classes:

(i) arguments with which a dispute can be won (or justified arguments),
(ii) arguments with which a dispute can be lost (or overruled arguments) and
(iii) arguments which left the dispute undecided (or defensible arguments).

There are two forms in which the concepts of an argumentation system can be developed depending on whether the focus is on sets or individual arguments. The declarative form is described as the semantics of argumentation where the units are sets of arguments rather than set of propositions defined over notions such as justified and acceptable rather over truth values. In the dialectical form, the aim is to establish the status not of sets but of individual arguments. The dialectical form can then be considered as the proof theory of an argumentation system. We begin by presenting the declarative version since it is more developed than the dialectical form.

2. The notion of justified argument.

Argumentation systems are not concerned with the truth of a proposition but with the justification for accepting a proposition as true. And a proposition is justified if there is an argument which defends it in case of attack. So for a conclusion to be justified, it is not enough to have an argument for it but it must emerge undefeated from an iterative justificatory process in which the same argument can be successively defeated and reinstated. In the above example, C is justified since there is no argument which defeats it. Is C the only justified argument? No; since C is not the only argument in the set, all arguments which are reinstated by C are also justified. It follows that \{A, C\} is the set of justified arguments of the Tweety example. But it is not always easy to determine the set
of justified arguments. Consider the following example, known as the even defeat cycle example, where two arguments defeat each other.

Example 2 (even defeat cycle).

D The platypus is a mammal because it feeds its babies on milk
E The platypus is not a mammal because it lays eggs

If we assume that D is justified then E is not justified. And similarly E can be justified only if D is not justified. Two possible statuses can then be assigned to D and E: one in which D is justified but E is not justified and one in which E is justified but D is not justified. In sharp contrast, a self-defeating argument T, i.e. an argument T which defeats itself, has no status assignment since both assumptions (T is justified or T is not justified) lead to a contradiction of the form T is justified iff T is not justified. The consequence is that a self-defeating argument might prevent a set of arguments to which it belongs from having a status assignment. Two solutions to this problem have been proposed: (i) either to exclude them by definition (this is Pollock’s solution) or (ii) to use the empty set as the argument which defeats it (but no other argument can defeat it). Concerning the problem of the even defeat cycle example, two approaches to its solution have been developed.

2.1. The unique status assignment approach.

The aim of this approach is to ensure that there is exactly one possible status assignment to every argument. We have shown that reinstatement is an adequate notion that can be used to capture the relation between supporting arguments. There are two ways of reinstating a defeated argument either directly or indirectly depending on whether the notion of justified argument is defined by a fixed-point definition or a recursive definition. Because of lack of space, we cannot discuss both kinds of reinstatement. We focus only on the fixed point approach because it is more elegant and straightforward than the recursive definition.
Fixed point definition.

The direct reinstatement approach can be captured by Dung’s notion of acceptability. In his landmark article, Dung shows that the notion of acceptability provides a unified account to many nonmonotonic logics.

Definition. An argument $A$ is acceptable with respect to a set $S$ of arguments iff each argument defeating $A$ is defeated by an argument in $S$.

Acceptability appears to be a fundamental notion since it is used to capture the relative justification of an argument with respect to a set of arguments. To give an idea of how it works, let’s reformulate the relation between arguments of the Tweety example in terms of acceptability. $A$ is not acceptable with respect to $\{B\}$ because $B$ defeats $A$, but it is acceptable with respect to both $\{B, C\}$ and $\{C\}$ since $C$ defeats its defeater $B$. $C$ is acceptable with respect to the empty set since it is an undefeated argument. Notice that $C$ is trivially acceptable with respect to $\{C\}$. But that does not mean that $C$ reinstates itself because there is no argument which defeats it. There are cases however in which an argument reinstates itself. Consider the even defeat cycle example. Assume $D$ is justified, in terms of acceptability this amounts to saying that $S=\{D\}$. $D$ is acceptable with respect to $S$ since all arguments defeating it (i.e. $E$) are defeated by an argument in $S$ i.e. by $D$ itself. $D$ can be seen as a fixed point of an operator $F$ which, for each set of arguments, returns the set of all arguments which are acceptable to it. The operator $F$ is defined as follows:

$$F(S) = \{ A \in \text{Args} / A \text{ is acceptable with respect to } S \}$$

The operator $F$ of the even defeat cycle example has two fixed points $F(D)=D$ and $F(E)=E$ but they are not its least fixed point which is $\emptyset$. In general, we have $F(\emptyset)=\emptyset$ if no argument is defeated. To prevent an argument from reinstating itself, the set of justified arguments will be defined as the least fixed point of $F$.

Definition 1. An argument is justified if it is a member of the least fixed point of $F$.

Dung gives an interesting reformulation of the fixed point definition which is in fact a technique for constructing the set of justified arguments by iterative application of the operator $F$ to the empty set. The first application of $F$ to $\emptyset$ yields undefeated arguments, and each further application of $F$ adds all

---

arguments which are reinstated by arguments that are already in the set. We apply this technique to get the set of justified arguments of the Tweety example:

\[ F^1 = F(\emptyset) = \{C\}; \quad F^2 = F(F^1) = \{A, C\}; \quad F^3 = F(F^2) = F^2. \]

2.2. The multiple status assignment approach.

Unlike the unique status assignment approach, this approach allows conflictual arguments with equal strength to have the two possible status assignments in each of which an argument is justified at the expense of the other, and consequently defines an argument as justified iff it receives this status in all status assignments. In example 2, there are two status assignments: one in which D is justified and E is not justified, and one in which D is not justified while E is justified. No argument is justified in both assignments, which means that neither D nor E is justified. This kind of reasoning is described as a sceptical attitude. In case of an irresolvable conflict, a sceptic refrains from drawing any conclusion. We obtain the same result as in the first approach. Can we conclude then that the two approaches are both sceptical? Consider the following example in which there are four arguments such that A and B defeat each other, A defeats C, B defeats C and C defeats D. According to the unique status assignment approach, all four arguments are defensible since no argument is undefeated. In the multiple status assignment approach, D is justified since it is contained in all status assignments namely \{A, D\} and \{B, D\}. This example shows that the first approach is more sceptical than the second one. The restrictive nature of the former one can be explained by the fact that it treats the notion of justified argument as a property which, for its transmission to supporting arguments, has to be acquired by a first argument. In the multiple status assignment approach, the three classes produced by a status assignment are defined as follows.

Definition 2. Given a set of arguments S and a relation of defeat on S:

(i) an argument is justified iff every status assignment to S assigns "justified" to it;
(ii) an argument overruled iff no status assignment to S assigns "unjustified" to it;
(iii) an argument is defensible iff it is justified in some and unjustified in some status assignment.

But this definition presupposes that a status assignment is guaranteed to exist which is not the case. The odd defeat cycle example, where there are three
arguments such that each one of them defeats the other, has no status assignment. It follows that sets of arguments which contain an odd defeat cycle might have no status assignment. The way to deal with such difficulties is to adopt a partial status assignment approach in which all arguments are not required to have a status assignment, and to define a justified assignment as a maximal partial status assignment.

Definition 3. A status assignment to a set of arguments $S$ ordered by a binary defeat relation is a partial status assignment of either justified or unjustified status (but not both) satisfying the following conditions:

(i) an argument is justified if all arguments defeating it are unjustified;
(ii) an argument is unjustified if it is defeated by an argument which is justified.
A justified assignment is a maximal (with respect to set inclusion) partial status assignment.

It is interesting to note that this definition is equivalent to Dung’s preferred semantics since conditions (i) and (ii) correspond to the notion of admissibility. A set $S$ is admissible iff it does not attack itself (i.e. conflict free) and attacks every argument that attacks it. In other words, $S$ is an admissible set if it can defend every argument belonging to it. The Tweety example has three admissible sets $\emptyset$, $\{C\}$, $\{A, C\}$. And since a set of arguments can have more than one admissible set, the preferred semantics is defined as the maximal admissible set. This type of reasoning is considered as credulous. In case of an irresolvable conflict, a credulous reasoner is ready to accept any conclusion that he could get. We will end this section by defining the stable semantics.

Definition 4. A conflict-free set $S$ is a stable extension iff every argument that is not in $S$, is defeated by an argument in $S$.

The classical semantics of artificial intelligence appears to be the most aggressive semantics since every argument not belonging to the set is attacked whether or not that argument is hostile to it.

3. From the declarative to the dialectical form

Our chief concern in the declarative form was to characterise the kinds of sets of arguments which can be generated by an argumentative framework. The dialectical form goes deeper by focusing on the status of individual arguments. This can be achieved by formulating an argumentation system in the form of a
dialogue game between two players. The dialogue can be conceived as an alternating series of moves between a proponent whose aim is to establish his main argument and an opponent whose job is to prevent him from succeeding by finding counterarguments capable of defeating the proponent’s arguments. The dialectical form can also be regarded as the proof theory of argumentation systems since it defines the conditions under which an argument is defeasibly provable. There are a variety of ways in which the exchange of arguments can be formalised. We use here the format of Prakken and Sartor to show how the proof theory can be set up. Parakken and Sartor\(^3\) define a dispute on an argument \(A\) as a non-empty sequence of moves. And a move is defined in terms of a player-argument pair, more precisely

1. move\(_i\) = (player\(_i\), \(A_i\)) with \(A_1 = A\) (i > 0),
2. player\(_i\) = P iff i is odd and player\(_i\) = O iff i is even.

These two conditions stipulate that the proponent begins the dialogue and the players take turns. The rest of the proof theory can be defined according to the kind of semantics intended to be captured since different argument extensions give rise to different sets of rules of the game. The following rules reflect the sceptical reasoning of Dung’s abstract system:

3. if player\(_i\) = P (i > 1) then \(A_i\) strictly defeats \(A_{i-1}\)
4. player\(_i\) = O then \(A_i\) defeats \(A_{i-1}\)
5. player\(_i\) = player\(_j\) = P and i \(\neq\) j then \(A_i \neq A_j\)

The first two rules put the burden of proof on the proponent since Dung’s grounded semantics requires that the proponent’s conclusion be justified. As a result, the proponent is not allowed to repeat his attack since the repetition of any attack means that the proponent’s arguments are not strong enough to defeat the counterarguments of his opponent. So the proponent wins the dialogue iff the opponent cannot move. This definition captures the notion of reinstatement since if the proponent’s last argument is undefeated, it reinstates all his previous arguments. And to win the game, the sceptical semantics requires from the proponent win every dialogue. In other words, an argument is defeasibly provable if the proponent has a winning strategy.

Proof theories for Dung’s preferred semantics for both sceptical and credulous reasoning are defined by Prakken and Vreeswijk\(^4\). Actually the real


task to be performed in the sense of dialogic as discussed in the main body of the paper is the connection between the semantic and proof-theoretical level.

4. From the dialectical to the procedural form

It should be noted, as Prakken and Vreeswijk rightly point out, that the players in the dialectical form are not real actors but stand for arguments and counterarguments that are required by the proof theory. The rules of the game are not therefore rules for real discussion between persons but just serve as a proof theory for nonmonotonic logic. This is an important limitation to the dialectical approach since it fails to capture the dynamic nature of a real discussion in two respects (i) argumentation systems are based on a fixed input theory. This assumption leaves no room for the use of arguments generated by the process of exchanging arguments. (ii) the proponent and the opponent are fixed roles since they are given to the players once and for all. On the other hand, formal systems for persuasion dialogues which have been developed by authors such as Hamblin5, MacKenzie6, Walton and Krabbe7 lack the notions of defeat and counterarguments since their underlying logic is monotonic. Some formal systems have been recently proposed which combine the two approaches to dialogue. To show how this combination can be achieved, we give an informal description of the main concepts of Lodder’s Dialaw8.

5. General features of Dialaw

Dialaw is a dialogical model of legal justification. The author argues that legal justification should be modelled procedurally since there are no criteria in terms of which the notion of justification can be defined. Dialaw is set up around three basic concepts.

5.1. The underlying logical systems.

Dialaw uses two kinds of logic. The first underlying logic in which the notion of argument is justified is monotonic. First predicate logic is used as a formal

language for the formation of arguments. The second underlying logic formalises nonmonotonic reasoning as the construction and comparison of conflictual arguments. Lodder adopts as a logic for defeasible argumentation Reason-based logic in which the notion of argument is represented as a reason(condition, conclusion). An argument simply states that the condition is a reason for the conclusion. As we have already explained, reasons can be attacked in three ways.

(i) A reason can be rebutted by giving another reason for a contradictory conclusion. The two following arguments rebut each other:

Chabot must be punished because he helps someone to commit suicide  
Chabot cannot be punished because he acts in force majeure

(ii) An assumption of a reason can be attacked by attacking its condition. The following argument “Chabot does not help someone to commit suicide because he has an alibi” is an attack assumption on “Chabot must be punished because he helps someone to commit suicide”.

(iii) Finally, a reason can be undercut by attacking its underlying rule since in Dialaw arguments can be supported by rules. And a general rule can be defeated if there is an exception rule which excludes it from being applied. Here is an example of undercutting an inference which has been successfully used in the Netherlands. Euthanasia is a punishable act because it is considered as a murder. But doctors can be faced with a situation in which they have on the one hand the duty to preserve life and on the one hand the duty to respect a person’s explicit wish to die. In the Netherlands, such a dilemma is considered as a force majeure which constitutes a legal ground for not prosecuting. So the general rule which says that “if someone helps a person to commit suicide then he must be punished” can be undercut by the exception rule “if a doctor helps a patient to commit suicide and the patient explicitly expresses his wish to die then he acts in force majeure”.

5.2. Moves of the dialogue.

A move is determined by the type of illocutionary act and its propositional content i.e the proposition the act is about. Four illocutionary acts can be used by the players: claim, question, acceptance and withdrawal. The latter three illocutionary acts are always reactions. Acceptance is a crucial move in Dialaw since an argument cannot be justified if it is not accepted by the opponent.
Claim is a more complicated move. By claiming a proposition, the proponent puts the main part of his argument to the test. The proposition is justified if the opponent accepts it in the next move. But the proposition can also be questioned in which case the proponent is required to support his original claim by making another claim in which he introduces a reason for it. This means that the burden of proof is put on the proponent. There are no preference criteria in Dialaw. Instead, reasons can be weighed in case of a conflict. To win a conflictual argument, the proponent has to provide a reason which explains why his arguments are stronger than the opponent’s arguments. This can be done by claiming an outweigh move. So the arguments of the proponent must not only defeat arguments of the opponent but, using Prakken and Sartor terms, must be strictly defeating.

A third possible response available to the opponent is to use a claim move as a reaction to the proponent’s claim. Such claims can be called counterclaims. Two counterclaims are allowed. The first is to deny the proposition claimed by the proponent. The second is to claim that the proponent’s claim is illegal; this move is more specific to legal reasoning because of the importance of the formal aspects of the legal procedure. The opponent can declare that “Chabot helps someone to commit suicide” is an illegal claim on the ground for example that a hidden camera is used to prove that Chabot is guilty. Counterclaims are very special moves since they induce a shift of the burden of proof. It is interesting to note that the negation (of a proposition in the first case, of a legal fact in the second) is interpreted in Dialaw as a change of role of the players. As a result, exactly as in dialogic we distinguish between the roles of attack and defence from O and P, a distinction must be introduced between the label of the parties according to the legal procedure (plaintiff/defendant) and the procedural role they have with respect to the burden of proof (proponent/opponent).

5.3. The commitment store.

This is another essential element of Dialaw since no fixed premises are imposed on the parties. The commitment store is a record in which the moves of the players are stored. As a result, the commitment store varies according to the type of illocutionary act used by the players. The commitment begins when a proposition is claimed or accepted and ends when it is withdrawn. The propositions of the commitment store can be divided in two classes:

(i) the class of undisputed propositions: these are propositions upon which both players agree; and
(ii) the class of disputed propositions: these are propositions to which only one player is committed.

Among the variety of rules which indicate how to play the game, those which put constraints on the players deserve to be mentioned. The author admits that the addition of moves is intended to impose, what he calls, a “minimal rationality” on the players of Dialaw. These special rules are needed for the following reasons:

(i) to make sure that the set of propositions to which each player is committed is conflict-free. Players are not allowed to claim contradictory arguments or arguments which exclude each other. It is forbidden to claim for example that a motorist is guilty of murder and of death by negligence.

(ii) to prevent possible deadlocks, two kinds of move are not allowed:

1. To reclaim the same proposition (to prevent arguments from justifying themselves),
2. To deny a proposition except if it is an immediate reaction to a claim of its opposite. Such a move prevents the opponent from using the negation as a means for forming counterarguments. As a reaction to the argument “Chabot must be punished because he is not a murderer” the opponent cannot counterclaim “Chabot should not be punished because he is not a murderer”.

(iii) to make it possible to justify arguments by using deductive reasoning. In particular, the opponent is forced to accept the conclusion of a previously disputed reason if he accepts both the rule on which the reason is based and the fact that the condition of the rule is satisfied. Here is a simple example to illustrate how deductive reasoning can be used. In the first move, the proponent claims that “Chabot must be punished because he helps someone to commit suicide”. The opponent questions this reason. As a supporting argument, the proponent claims the following underlying specific rule: “if Chabot helps someone to commit suicide then he must be punished”. If the opponent accepts the specific rule (assuming that the opponent does not challenge this claim; the assumption here is that there is a general rule underlying the specific one which is applicable to Chabot’s case) and that Chabot helps someone to commit suicide then he must also accepts that Chabot is punishable.

One of Dialaw’s main drawbacks, which is a consequence of the author stance on the notion of justification, is that a dialogue is not guaranteed to end. This is just an outline of some interesting aspects of Dialaw. A full account of this remarkable dialogical model is out of the scope of this paper. On the whole the
The author has brilliantly succeeded in combining in one system three different domains: formal dialogue, defeasible argumentation and legal reasoning.

Future theoretical research on the mentioned subjects can be found in the related literature. As far as we are concerned, we wish only to make the following concluding observations. Outside artificial intelligence, legal reasoning seems to be the only practical domain which has benefited greatly from the results of systems for defeasible argumentation. Moreover, the introduction of new concepts in the development of formal particular systems is usually based on the study of isolated examples invented to serve the relevant purpose. Concrete debates are not exploited despite the fact that epistemology and history of science, among other domains, are full of various types of dialogues. My project of is precisely to reverse the present trend by proposing the reconstruction of a concrete historical scientific debate following some challenging suggestions of Marcelo Dascal⁹ concerning the role of debate in epistemology. Such an approach may give new impetus to the development of formal systems. And by focusing on the study of scientific and philosophical debates, this line of research wants to put more emphasis on the fundamental role of a critical discussion as a powerful instrument in the discovery of new and insightful ideas.¹⁰

---


¹⁰ See the article « Essai ». 