Gradual acceptability in argumentation systems

C. Cayrol **M.C. Lagasquie-Schiex**
IRIT-UPS
118 Rte de Narbonne, 31062 Toulouse
ccayrol@irit.fr, lagasq@irit.fr

Abstract

The argumentation is based on exchange and valuation of arguments interacting, then on the definition of accepted arguments (or sets of arguments) w.r.t. the proposed valuation. In this paper, using the argumentation system of [Dun95] and the valuations proposed in [CLS02a], we introduce graduality in the acceptability of arguments.

1 Introduction

As shown by [Dun95], argumentation frameworks provide a unifying and powerful tool for the study of many formal systems developed for common-sense reasoning, as well as for giving meaning to logic programs. Argumentation is based on the exchange and valuation of arguments. It can be applied, among others, in the legal domain, for collective decision support systems or for negotiation support.

The main characteristic of an argumentation system is the presence of interactions (in particular attack relations) between arguments. For instance, if the argument takes the form of a logical proof, we can propose some arguments for a proposition and some others against this proposition (in this case, the attack relation relies upon logical inconsistency).

The argumentation process is composed of a valuation step of the relative strength of the arguments (intrinsic valuation, interaction-based valuation, both, see [KAEF95; Par97; PS97; AC98; Dun95; JV99; BH01; CLS01; CLS02a]). Then, in a second step, the most acceptable arguments w.r.t. this valuation are selected. In this domain, most of works lead to define the acceptability of an argument by its membership of an *acceptable* set ; this is a *collective acceptability*. [Dun95]'s framework follows this approach but it proposes only two possible states for an argument: accepted or not-accepted.

It seems very natural to introduce a graduality in this notion of acceptability in order to distinguish between arguments more or less acceptable, and then to define "acceptability levels". So, we use gradual interaction-based valuation in order to take into account the "quality" of the defeaters, of the defenders, …

In section 2, [Dun95]'s framework is presented: an argumentation system based on a set of arguments and a binary relation on this set. Using the notion of [Dun95]'s acceptability, we identify different levels of collective acceptability. Then, in section 3, we remind two kinds of gradual valuation issued of [CLS01]: a local approach (generalization of some existing works: [JV99; BH01]), and a global approach introduced in [CLS01]. We propose, in section 4, a notion of gradual acceptability using these gradual valuations.

2 Dung's acceptability

2.1 Dung's framework

We consider the abstract framework introduced in [Dun95]. An *argumentation system* $<\mathcal{A}, \mathcal{R}>$ is a set $\mathcal{A}$ of arguments and a binary relation $\mathcal{R}$ on $\mathcal{A}$ called an *attack relation*: let A_i and $A_j \in \mathcal{A}$, $A_i \mathcal{R} A_j$ means that A_i attacks A_j (the notion of attack between A_i and A_j implies that there is a "conflict" between A_i and A_j). We will precise neither the form of the arguments, nor the attack relation.

Notations: Let $A \in \mathcal{A}$, the set $\{A_i \in \mathcal{A} | A_i \mathcal{R} A\}$ is denoted by $\mathcal{R}^-(A)$ and the set $\{A_i \in \mathcal{A} | A \mathcal{R} A_i\}$ is denoted by $\mathcal{R}^+(A)$. $<\mathcal{A}, \mathcal{R}>$ defines a directed graph $\mathcal{G}$ (called *attack graph*).

Def. 1 *Let $\mathcal{G}$ be the attack graph associated to the argumentation system $<\mathcal{A}, \mathcal{R}>$:*

- *$A \in \mathcal{A}$ such that $\mathcal{R}^-(A) = \varnothing$ is a* leaf *of $\mathcal{G}$.*
- *A* path *from A to B is a sequence of arguments $\mathcal{C} = A_1 - \ldots - A_n$ such that $A = A_1$, $A_1 \mathcal{R} A_2$, …, $A_{n-1} \mathcal{R} A_n$, $A_n = B$. The* length *of this path is the number of edges that are used in the path and is denoted by $l_\mathcal{C}$. The set of the paths from A to B is denoted by $\mathcal{C}(A, B)$.*
- *Let two paths $\mathcal{C}_A \in \mathcal{C}(A_1, A_n)$ and $\mathcal{C}_B \in \mathcal{C}(B_1, B_m)$. These paths are* dependent *iff $\exists A_i \in \mathcal{C}_A$, $\exists B_j \in \mathcal{C}_B$ such that $A_i = B_j$.* Independent *otherwise.*
  These paths are root-dependent *in A_n iff $A_n = B_m$ and $\forall A_i \neq A_n \in \mathcal{C}_A$, $\nexists B_j \in \mathcal{C}_B$ such that $A_i = B_j$.*
- *A* cycle *is a path $\mathcal{C} = A_1 - \ldots - A_n - A_1$ such that $\forall i, j \in [1, n], i \neq j$, $\nexists A_i, A_j \in \mathcal{C}$ such that $A_i = A_j$.*
  A cycle $\mathcal{C}$ is isolated *iff $\forall A_i \in \mathcal{C}$ if $\exists B$ such that $B \mathcal{R} A_i$ then $B \in \mathcal{C}$.*
  Two cycles $\mathcal{C}_A = A_1 - \ldots - A_n - A_1$ and $\mathcal{C}_B = B_1 - \ldots - B_m - B_1$ are interconnected *iff they are dependent.*

We now introduce the notions of direct and indirect attacks and defences. The notions introduced here are inspired by related definitions first introduced in [Dun95] but are not strictly equivalent: in [Dun95]'s work, direct attacks (resp. defences) are also indirect attacks (resp. defences) which is not true in our definitions:

Def. 2 *Let $A \in \mathcal{A}$. The* direct defeaters *of A are the elements of $\mathcal{R}^-(A)$.*
The direct defenders *of A are the direct defeaters of the elements of $\mathcal{R}^-(A)$.*
The indirect defeaters *of A are the elements A_i such that: $\exists \mathcal{C} \in \mathcal{C}(A_i, A)$ such that $l_\mathcal{C} = 2k+1$, with $k \geq 1$.*
The indirect defenders *of A are the elements A_i such that: $\exists \mathcal{C} \in \mathcal{C}(A_i, A)$ such that $l_\mathcal{C} = 2k$, with $k \geq 2$.*

Def. 3 *Let $A \in \mathcal{A}$, an* attack branch *(resp.* defence branch*) for A is an odd (resp. even) length path from a leaf to A. When such a branch exists, A is said to be the root of the branch.*

2.2 Collective acceptability

The collective acceptability of an argument depends on its membership of some sets (*acceptable sets* or *extensions*) characterized by particular properties such as:

Def. 4 *Let $<\mathcal{A}, \mathcal{R}>$ be an argumentation system. A set $E \subseteq \mathcal{A}$ is* conflict-free *iff $\nexists A, B \in E$ such that $A\mathcal{R}B$.*
Let $E \subseteq \mathcal{A}$, $A \in \mathcal{A}$. E defends (collectively) *A iff $\forall B \in \mathcal{A}$, if $B\mathcal{R}A$, $\exists C \in E$ such that $C\mathcal{R}B$.*

[Dun95] defines several semantics for the collective acceptability ; among them, we have the *admissible semantics*, the *preferred semantics* and the *stable semantics* (with, for respective extensions, the admissible sets, the preferred extensions and the stable extensions):

Def. 5 *Let $<\mathcal{A}, \mathcal{R}>$ be an argumentation system. Let $E \subseteq \mathcal{A}$. E is* admissible *iff E is conflict-free and E defends all its elements.*
E is a preferred extension *iff E is maximal for the set-inclusion among the admissible sets.*
E is a stable extension *iff E is conflict-free and E attacks each argument which doesn't belong to E ($\forall A \in \mathcal{A} \setminus E$, $\exists B \in E$ such that $B\mathcal{R}A$).*

See in [Dun95] the properties of these semantics.

2.3 Different levels of acceptability

Primary levels Under a given semantics, an argument can be *uni-accepted* (it belongs to all the extensions), or it can be *exi-accepted* (it belongs to at least one extension), or it can be *not-accepted* (it doesn't belong to any extension).
These levels don't allow to distinguish two arguments A and B such that $A\mathcal{R}B$ and $B\mathcal{R}A$.

First refinement We propose a new definition which takes into account the situation of the argument w.r.t. its defeaters. So, we can refine the class of the exi-accepted arguments under a given semantics S.

Def. 6 *A is* cleanly-accepted *for S iff A belongs to at least one extension of S and $\forall B \in \mathcal{A}$ such that $B\mathcal{R}A$, B doesn't belong to any extension of S.*

So, an argument will be better, in the point of view of the acceptability, if its defeaters are not-accepted.

Prop. 1 ([CLS02b]) *Let S be a semantics in the sense of [Dun95]. Let $A \in \mathcal{A}$. If A is uni-accepted then A is cleanly-accepted. The contrary is false.*

This notion allows the refinement of the primary levels. For a semantics S and an argument A, A can be *uni-accepted*, or *cleanly-accepted*, or *only-exi-accepted* (if A is not cleanly-accepted but exi-accepted), or *not-accepted*.

Some particular cases In all the cases where there is only one extension, the first three levels merge. We also know some cases in which the uni-accepted level merges with the cleanly-accepted level. See the proofs in [Dou02; CLS02b].

3 Interaction-based gradual valuations

We have proposed in [CLS01] two kinds of valuation in order to take into account the quality of the defeaters and of the defenders of an argument.

3.1 The "local" approach (generic valuation)

This approach proposes to compute the value of an argument using only the values of its direct defeaters. It is a generic approach generalizing some existing works ([BH01; JV99]). We formalize this valuation using the following assumption: Let $(W, \geq)$ be a totally ordered set, with a minimum element (V_{Min}) and a subset V of W that contains V_{Min} and with a maximum element V_{Max}.

Def. 7 *Let $<\mathcal{A}, \mathcal{R}>$ be an argumentation system. A valuation is a function $v : \mathcal{A} \rightarrow V$ such that:*

- *$\forall A \in \mathcal{A}$, $v(A) \geq V_{Min}$*
- *$\forall A \in \mathcal{A}$, if $\mathcal{R}^-(A) = \varnothing$, then $v(A) = V_{Max}$*
- *$\forall A \in \mathcal{A}$, if $\mathcal{R}^-(A) = \{A_1, \ldots, A_n\} \neq \varnothing$, then $v(A) = g(h(v(A_1), \ldots, v(A_n)))$*

where $h : V^* \rightarrow W$ such that (V^* denotes the set of all finite sequences of elements of V) $h(x) = x$, $h() = V_{Min}$ and $h(x_1, \ldots, x_n, x_{n+1}) \geq h(x_1, \ldots, x_n)$,
and $g : W \rightarrow V$ such that $g(V_{Min}) = V_{Max}$, $g(V_{Max}) < V_{Max}$ et g is decreasing (if $x \leq y$ then $g(x) \geq g(y)$).

A valuation v given by the def. 7 induces a complete preordering $\succeq$ on the set of arguments $\mathcal{A}$ defined by: $A \succeq B$ iff $v(A) \geq v(B)$ (see the proof in [CLS01]).

3.2 The "global" approach (tuple-based valuation)

Here, we want to define for each argument a "label" reflecting the structure of the attack graph leading to this argument.

Tuple-based value of an argument The label of an argument A is thus a tuple listing the lengths of all the branches of attack and all the branches of defence leading to A. The case of an attack graph with cycles is the subject of specific rules in which the cycle is considered as a meta-argument with two branches, one attack branch of length 1 and one defence branch of length 2 (each argument of the cycle is, at the same time, attacked and defended by the cycle) on which is added the impact of the defeaters which don't belong to the cycle and that of the other interconnected cycles. All these ideas lead in [CLS01] to the following rules:

Rule 1 *Let A be an argument not belonging to a cycle. If A is not attacked then $v(A) = ()$ else (A has the direct defeaters $B_1, \ldots, B_n$ whose respective values are the tuples $(b_1^1, \ldots, b_{m_1}^1), \ldots, (b_1^n, \ldots, b_{m_n}^n)$)*

$$v(A) = (b_1^1 + 1, \ldots, b_{m_1}^1 + 1, \ldots, b_1^n + 1, \ldots, b_{m_n}^n + 1)$$

Rule 2 *Let A be an argument belonging to the cycles $\mathtt{C}_1 \ldots \mathtt{C}_m$. Let $\mathtt{C}'_1 \ldots \mathtt{C}'_n$ other cycles interconnected with one of the $\mathtt{C}_i$ or between themselves*[1]*. Let $X^1 \ldots X^p$ be arguments which don't belong to the cycles but which are direct defeaters of one of the cycles*[2]*. We denote:*

- *l_i: minimal length of a path from an element of $\mathtt{C}'_i$ to A,*
- *l_{X^i} (length*[3] *of a path from X^i to A),*
- *the value of each argument X^i is: $v(X^i) = (x_1^i, \ldots, x_{k^i}^i)$.*

So, we have:

$$v(A) = (\overbrace{1, 2, \ldots, 1, 2,}^{m \text{ times}} 1 + l_1, 2 + l_1, \ldots, 1 + l_n, 2 + l_n, \left.\begin{matrix} x_1^1 + l_{X^1}, \ldots, x_{k^1}^1 + l_{X^1}, \ldots, \\ x_1^p + l_{X^p}, \ldots, x_{k^p}^p + l_{X_p}) \end{matrix}\right\} \star$$

($\star$ as many time as there are paths from X^i to A)

Def. 8 *Let v a function from $\mathcal{A}$ to the set of all the finite tuples of integers, v is a* tuple-based valuation *iff v satisfies the rules 1 and 2.*

Comparison of arguments using tuples Since, in a tuple, the even values (corresponding to the branches of defence) and the odd values (corresponding to the branches of attack) do not play the same role, we cannot perform a simple lexicographical comparison. So, we compare the tuples in two quite distinct steps by adopting a *careful attitude*:

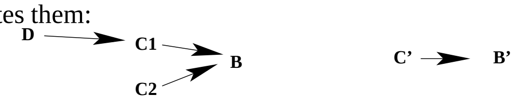
- A "first step" makes it possible to determine and *to compare the number of branches of attack and the number of branches of defence* of each one. That gives us two criteria (for the defence and for the attack). If these two criteria are in agreement, i.e. one of the tuples has more branches of defence and less branches of attack than the other, then we conclude. In the same way, if the two criteria are in disagreement (one of the tuples has more branches of defence and more branches of attack than the other), the tuples are regarded as incomparable.
- If not, both tuples having the same number of attacks and the same number of defences, a "second step" *compares the respective quality of the attacks and defences*. The comparison relates on the one hand on the "even parts" of the tuples, on the other hand on the "odd parts" of the tuples. Here also, in case of dissension, the conclusion is the incomparableness of the tuples ! This comparison uses a lexicographical principle.

This method leads to an algorithm described in [CLS01; CLS02b].

Some properties (see proofs in [CLS01; CLS02b]) A tuple-based valuation associated to the algorithm described in [CLS01; CLS02a] induces a partial preordering $\succeq$ on the set of the arguments whose maximum elements will be the leaves of value $()$. We have also a property showing a kind of "graphical independence": A and A' have the same value $(2, 2)$ although they are root of different subgraphs:

3.3 Essential differences between "local" and "global" valuations

Let us recall on an example the essential point which differentiates them:



On this example, with the local approach, one notes that B has two direct defeaters (C_2 of maximum value and C_1) whereas B' has only one (C' of maximum value). Thus B' is better than B. With the global approach, two branches (one of attack and one of defence) lead to B whereas only one branch of attack which leads to B'. Thus B is better than B' (since it has at least a defence whereas B' has no). In this case, the defeater C_1 lost its negatif status of defeater, since it is in fact "carrying a defence" for B !

4 Gradual acceptability

The exploitation of the gradual interaction-based valuation allows us to define new levels of collective acceptability.

4.1 Comparison based on gradual valuation

Let v be a gradual valuation, we can define a *v-preference* using the preordering $\succeq$ issued of v:

Def. 9 *Let $A, B \in \mathcal{A}$, A is* v-preferred to *B iff $B \not\succeq A$.*

If the v valuation induces a complete preordering on the set of the arguments, we have: A is *v-preferred to* B iff $A \succeq B$.
The v-preference can be used within a level of acceptability (for example, on the exi-accepted arguments) to obtain better accepted arguments.
If the v-preference is applied without respecting the levels of acceptability defined in section 2.3, counter-intuitive situations may arrive.

---

[1] A doesn't belong to any of the $\mathtt{C}'_i$.

[2] They are direct defeaters of some elements of the cycles: $\forall X^i$, $\exists \mathtt{C} \in \{\mathtt{C}_1, \ldots, \mathtt{C}_m, \mathtt{C}'_1, \ldots, \mathtt{C}'_n\}$ such that $\exists Y \in \mathtt{C}$ et $X^i \mathcal{R} Y$.

[3] If there are several paths from X^i to A, we take into account the length of each path.

4.2 Situation of an argument w.r.t. its defeaters

One can also directly use the v-preference to compare, from the point of view of acceptability, an argument and its defeaters[4].

Def. 10 *Let* $A \in \mathcal{A}$, A *is* well-defended for v *iff* $\forall B \in \mathcal{A}$ *such that* $B\mathcal{R}A$, A *is* v*-preferred to* B.

Thus, we capture the idea that an argument will be better accepted if it is at least as good as its direct defeaters (or incomparable with them in the case of a partial ordering !). The set of the well-defended arguments will depend on the selected valuation.

4.3 Compatibility between collective acceptability and gradual valuation

We thus have a new concept of acceptability (def. 10) and it is necessary to study its compatibility with the partition in levels proposed in section 2.3.
Examples easily show that these two approaches are not compatible in the case of the gradual valuations (global or local).

Particular cases leading to compatibility In the context of an argumentation system with a finite relation $\mathcal{R}$ and not admitting any cycle[5], the stable semantics and the preferred one propose only one extension and the levels of acceptability uni-accepted, exi-accepted, cleanly-accepted merge. So, in this case, an exi-accepted argument is also a well-defended argument and vice-versa (proof in [CLS02b]).

5 Conclusion

In this paper, we studied a very natural notion in argumentation: the graduality of acceptability of an argument. For that, we started with the concept of collective acceptability of [Dun95] which already makes it possible to have 3 levels of acceptability (uni-accepted, exi-accepted, not-accepted). We then took into account, at the same time, the status of the argument w.r.t. its defeaters but also a gradual valuation of the arguments in the concept of collective acceptability. That led us to propose various concepts applicable to an argument:

- the *cleanly-accepted* arguments: those whose defeaters are not accepted,
- the v*-preference* between arguments resulting directly from a valuation v,
- the *well-defended* arguments: those which are v-preferred to their defeaters (for a gradual valuation v).

The first concept makes it possible to refine the level of exi-accepted in two sublevels (cleanly-accepted and only-exi-accepted). The second concept allows graduality in all the levels of acceptability but only by using it inside each level. The third concept makes it possible to define two new levels of acceptability (well-defended and not-well-defended). Unfortunately, we noted on various examples that, in the general case, these levels were not compatible with those proposed by [Dun95] and refined using the first concept (clean acceptability) ; and that, although the concepts of well-defended argument and cleanly-accepted argument are both built starting from the interactions between arguments. Some particular cases on which compatibility exists were identified.

[4] This idea is also used in the notion of "defeat" proposed by [BC02]. So, there is a link between a "well-defended argument" and an argument which is not "defeated" in the sense of [BC02] by its direct defeaters. Note that, in [BC02], the valuation is an extra knowledge added in the argumentation framework. Contrastedly, here, the v-preference is extracted from the attack graph.

[5] So, $(\mathcal{A}, \mathcal{R})$ is well-founded.

References

[AC98] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In G. F. Cooper and S. Moral, editors, *Proc. of the 14th Uncertainty in Artificial Intelligence*, pages 1–7, Madison, Wisconsin, 1998. Morgan-Kaufmann.

[BC02] Trevor J.M. Bench-Capon. Value based argumentation frameworks. In Benferhat and Giunchiglia, editors, *Proc. of the 9th International Workshop on Nonmonotonic Reasoning (session on Argument, Dialogue and Decision)*, pages 444–453, Toulouse, France, 2002.

[BH01] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128 (1-2):203–235, 2001.

[CLS01] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Évaluation graduelle d'arguments. Rapport de recherche 2001-18-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, November 2001.

[CLS02a] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Gradual handling of contradiction in argumentation frameworks. In *Proc. of the 9th IPMU (Information Processing and Management of Uncertainty in Knowledge-Based Systems)*, pages 83–90, Annecy, France, 2002.

[CLS02b] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Vers une acceptabilité graduelle des arguments dans les sytèmes d'argumentation. Rapport de recherche 2002-42-R, Institut de Recherche en Informatique de Toulouse (I.R.I.T.), France, December 2002.

[Dou02] Sylvie Doutre. *Autour de la sémantique préférée des systèmes d'argumentation*. Thèse, Université Paul Sabatier, IRIT, 2002.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[JV99] Hadassa Jakobovits and Dick Vermeir. Robust semantics for argumentation frameworks. *Journal of logic and computation*, 9(2):215–261, 1999.

[KAEF95] Paul Krause, Simon Ambler, Morten Elvang, and John Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.

[Par97] Simon Parsons. Normative argumentation and qualitative probability. *LNAI*, 1244:466–480, 1997.

[PS97] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.