Argumentation Games for Admissibility and Cogency Criteria

Gustavo BODANZA, [a,b] Fernando TOHMÉ [a,b] and Guillermo R. SIMARI [a,1]
[a] *Universidad Nacional del Sur, Argentina*
[b] *CONICET, Argentina*

Abstract. In this work, we develop a game-theoretic framework in which *pro* and *con* arguments are put forward. This framework intends to capture winning strategies for different defense criteria. In turn, each possible extension semantics satisfies a particular defense criterion. To ensure that only semantics satisfying given criteria are obtained, protocols for playing have to be defined, being the ensuing winning strategies full characterizations of the corresponding criteria. Admissibility and strong admissibility are two of those criteria proposed in the literature for the evaluation of extension semantics; here, we also introduce two weaker criteria, *pairwise* and *weak cogency*, for the evaluation of non-admissible semantics, and we define specific game protocols capturing them.

Keywords. Argumentation frameworks, Argumentation games, Extension semantics, Defense criteria

Introduction

Argumentation is a process in which arguments for and against some claim are put forward by contending parties. While it might adopt different patterns, a very general characterization of argumentation starts from a general set of arguments and a given "attack" relation among them. In [8], Dung formalizes an *argumentation framework* as a tuple $AF = \langle \mathcal{A}, \succ \rangle$, which consists of a set of arguments $\mathcal{A}$ and the *attack relation* $\succ \subseteq \mathcal{A} \times \mathcal{A}$. Given $A, B \in \mathcal{A}$, $A \succ B$ means that A attacks B. For simplicity, we will only consider *finite* argumentation frameworks, *i.e.*, frameworks based in a finite set $\mathcal{A}$.[2]

The main problem for argumentation frameworks is the determination of which arguments can be accepted in the framework. This notion refers to the possibility of shielding arguments from attacks, and the ways in which this can be achieved lead to alternative notions of acceptance. The usual way is by defining an *extension semantics* $\mathcal{S}$ for argumentation frameworks, yielding, for every argumentation framework AF, a family $\mathcal{E}_{\mathcal{S}}(AF) \subseteq 2^{\mathcal{A}}$ of "extensions". It is said that an argument is *credulously* accepted in a semantics $\mathcal{S}$ iff it belongs to some extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$, and it is said to be *skeptically* accepted iff it belongs to every extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$.

[1] Corresponding Author: Departamento de Cs. e Ing. de la Computación, Universidad Nacional del Sur, 8000 Bahía Blanca; E-mail: grsimari@gmail.com.

[2] From now on 'AF' will denote an arbitrary argumentation framework, unless the contrary is stated.

Argumentation semantics have been evaluated in different ways, usually via canonical examples but also through the confrontation with human behavior ([14]). On the other hand, Baroni and Giacomin ([3]) have proposed several criteria or "principles" for the evaluation of extension semantics. We are interested in this last approach, particularly in the defense-related criteria concerning admissibility (a widely accepted criterion introduced in Dung's seminal work). These authors distinguish two versions of admissibility, a weak one (every argument that belongs to an extension is defended by the extension) and a strong one (every argument that belongs to an extension is defended by another argument from the extension). But we are also concerned with weaker forms of defense; admissibility is too skeptic towards arguments attacked by self-attacking arguments; furthermore, it treats differently cycles of attack of even and odd length. Arguably $CF2$ [2] and *stage* [17] semantics are the best known amendments for those rather undesirable features, but others were also proposed in [4] and [5]. In this paper we introduce, as general criteria of defense, two versions of what we call "cogency": *pairwise cogency* and *weak cogency*. These criteria, satisfied only by extensions that overcome the problems introduced by cycles of attack, will be formally introduced in Section 1.

Acceptance can also be characterized via proof-procedures. The most common among them are based on *labelings* [11,6] and *dialogue games* [1,7,9,10,12,15,16,18,19]. One of the goals of this paper is to provide a proof-theoretical counterpart to pairwise and weak cogency in the form of dialogue games. The formal model will be defined in Section 2, while in Section 3 specific protocols are defined, capturing in a sound and complete way Baroni and Giacomin's criteria as well as pairwise and weak cogency.

1. Defense criteria

In [3], Baroni and Giacomin have proposed several significant criteria to evaluate extension semantics. We will consider here those related to admissibility:

Definition 1. For any argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$ and $S \subseteq \mathcal{A}$, let $F(S) = \{A : \forall B(B \succ A \Rightarrow \exists C \in S \ C \succ B)\}$ ([8]'s characteristic function). A subset $T \subseteq \mathcal{A}$ is *defensible* iff T satisfies the following condition:

$$T \subseteq F(T) \tag{1}$$

A semantics $\mathcal{S}$ satisfies the *admissibility* criterion iff for any argumentation framework AF and for every $E \in \mathcal{E}_{\mathcal{S}}(AF)$, E is defensible.

Another important criterion considered in [3] demands that all the extensions sanctioned by a semantics must be conflict-free. Although not strictly related with defense, all the extension semantics in the literature satisfy this principle:

Definition 2. A set of arguments T is *conflict-free* iff $\forall A, B \in T \ (A \not\succ B)$. A semantics $\mathcal{S}$ satisfies the conflict-free criterion iff for any argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$ and for every $E \in \mathcal{E}_{\mathcal{S}}(AF)$, E is conflict-free.

Dung ([8]) uses the term 'admissible' to call sets of arguments which are both defensible and conflict-free. Hence notice that any semantics which extensions are all ad-

missible sets (in Dung's sense) satisfies the admissibility criterion, but the converse is not necessarily true.

A strong criterion of defense is obtained also in terms of classes of arguments:

Definition 3. Given a class of arguments $S \subseteq \mathcal{A}$, the set of arguments *strongly defended* by S is the set $sd(S) \subseteq \mathcal{A}$ such that $A \in sd(S)$ iff for all $B \in \mathcal{A}$, if $B \succ A$ then there exists some $C \in S - \{A\}$ such that $C \succ B$ and $C \in sd(S - \{A\})$.

Definition 4. For any argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$, a subset $T \subseteq \mathcal{A}$ is *strongly admissible* iff T satisfies the following condition:

$$A \in T \;\Rightarrow\; A \in sd(T) \tag{2}$$

A semantics $\mathcal{S}$ satisfies the *strong admissibility* criterion iff for any argumentation framework AF and for every $E \in \mathcal{E}_{\mathcal{S}}(AF)$, E is strongly admissible.

Among Dung's semantics, all of them satisfy the admissibility criterion while only grounded semantics satisfies the strong admissibility criterion.

On the other hand, there also exist extension semantics that do not satisfy admissibility. Examples are *stage* [17], $CF2$ [2], *sustainable*, *tolerant* [5], and *lax* [4] semantics. We propose here two criteria to evaluate them, *pairwise cogency* and *weak cogency*. Informally, we call "cogent" an argument that is accepted unless some of its attackers have a better defense than itself. For example, an argument supported by a well established scientific theory, *e.g.*, the argument from special relativity theory that no particle can exceed the speed of light is not defeated by an argument drawn from experiments showing an "anomaly", say neutrinos traveling faster than light, unless it is supported by a rival, stronger scientific theory. Cogency is then a notion in which admissibility exerts only a relative, contextual, authority:

Definition 5. Given an argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$, and two subsets S, S' $\in 2^{\mathcal{A}}$, we say that S is *at least as cogent as* S', in symbols, $S \geq_{cog} S'$, iff S is admissible in the restricted argumentation framework $\langle \mathcal{A}, \succ_{|_{S \cup S'}} \rangle$. We say that S is *strictly more cogent than* S', in symbols, $S >_{cog} S'$, iff $S \geq_{cog} S'$ and not $S' \geq_{cog} S$.

We obtain a new criterion of defense, weaker than admissibility:

Definition 6. For any argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$, a subset $T \subseteq \mathcal{A}$ is *pairwise cogent* iff T is maximal w.r.t. $\geq_{cog}$, i.e.:

$$\forall S \subseteq \mathcal{A}\ S \not> _{cog}\ T \tag{3}$$

A semantics $\mathcal{S}$ satisfies the *pairwise cogency* criterion iff for any argumentation framework AF and for every $E \in \mathcal{E}_{\mathcal{S}}(AF)$, E is pairwise cogent.

Let $Cog(AF) = \{E \subseteq \mathcal{A} : E \text{ is pairwise cogent}\}$. It is easy to see that if E is admissible then $E \in Cog(AF)$ (moreover, $E \geq_{cog} S$ for every $S \subseteq \mathcal{A}$). *Sustainable* semantics, in particular, takes as extensions all the maximal (w.r.t. $\subseteq$) subsets of $Cog(AF)$.

The salient behavior of sustainable semantics is the avoidance of undesired interferences of self-attacking arguments. For example, the argumentation framework

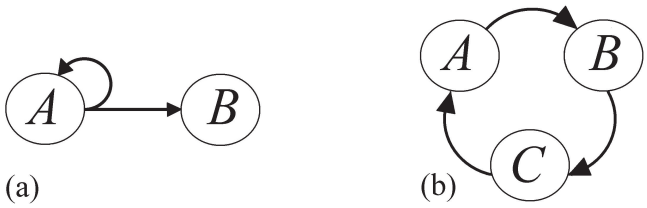

Figure 1.

$AF = \langle\{A, B\}, \{(A, A), (A, B)\}\rangle$ (Figure 1(a)) has only one sustainable extension, $\{B\}$. Note that the only subsets satisfying (3) are $\emptyset$ and $\{B\}$, but the latter is the only maximal one. On the other hand, note that we have both $\{B\} \not>_{cog} \{A\}$ and $\{A\} \not>_{cog} \{B\}$. This explains why $\{B\} \in Cog(AF)$, even if it is not admissible.

Even weaker (but sensible) defense criteria can be conceived:

Definition 7. For any argumentation framework $AF = \langle\mathcal{A}, \succ\rangle$, a subset $T \subseteq \mathcal{A}$ is *weakly cogent* iff T satisfies the following condition:

$$\forall S \subseteq \mathcal{A}\,(S >_{cog} T \Rightarrow S \notin Cog(AF)) \tag{4}$$

A semantics $\mathcal{S}$ satisfies the *weak cogency* criterion iff for any argumentation framework AF and for every $E \in \mathcal{E}_{\mathcal{S}}(AF)$, E is weakly cogent.

The intuition behind weak cogency is that a set of arguments is acceptable unless it can be defeated by a pairwise cogent one. Notice that every weakly cogent set T is such that it is either pairwise cogent or for every $S >_{cog} T$ there exists T' such that $T' >_{cog} S$.

Every semantics that satisfies pairwise cogency also satisfies weak cogency. *Lax* semantics is characterized as the class of maximal (w.r.t. $\subseteq$) subsets of arguments satisfying (4). It determines the acceptance of a set of arguments if every other set which is strictly more cogent is not pairwise cogent. For example, the argumentation framework $\langle\{A, B, C\}, \{(A, B), (B, C), (C, A)\}\rangle$ (Figure 1(b)) has three lax extensions, $\{A\}$, $\{B\}$ and $\{C\}$. Note that $\{A\} >_{cog} \{B\} >_{cog} \{C\} >_{cog} \{A\}$ and no other subset of $\mathcal{A}$ is strictly more cogent than each extension. Moreover, none of them is pairwise cogent, hence all of them are weakly cogent. The criterion makes sense especially in situations of practical decisions:

Example 1. Assume you have to choose a school for your children and your candidates are s_1, s_2 and s_3. You evaluate them according to three criteria: nearness, tuition fee and social environment. A candidate is preferred to another if it is better with respect to most of the criteria. Assume now that after a one-to-one comparison you build these three arguments:
A: "s_2 is better than s_1 with respect to nearness, but s_1 is better than s_2 with respect to tuition fee and environment; then choose s_1";
B: "s_3 is better than s_2 with respect to environment, but s_2 is better than s_3 with respect to nearness and tuition fee; then choose s_2";

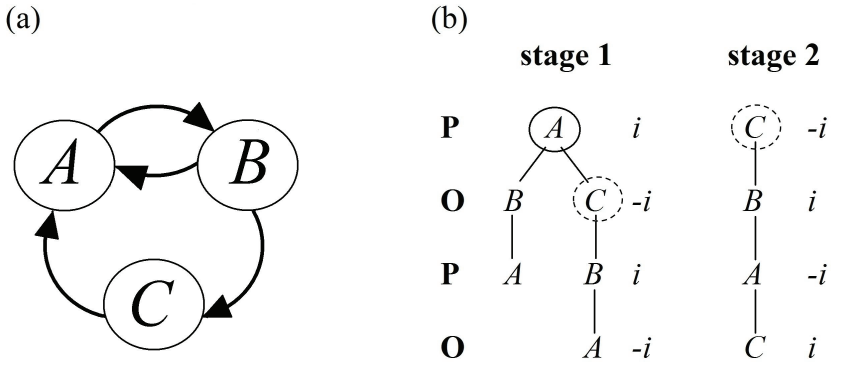

Figure 2.

C: "s_1 is better than s_3 with respect to tuition fee, but s_3 is better than s_1 with respect to environment and nearness; then choose s_3".
The situation can be modeled through the argumentation framework depicted in (Figure 1(b)).

$CF2$ semantics also obtains the extensions $\{A\}$, $\{B\}$ and $\{C\}$. Nevertheless, it fails to satisfy both pairwise and weak cogency. The argumentation framework $\langle\{A, B, C\}, \{(A, B), (B, A), (B, C), (C, A)\}\rangle$ (Figure 2(a)) offers a counterexample: $CF2$ semantics sanctions the three extensions $\{A\}$, $\{B\}$ and $\{C\}$ again, while $\{C\}$ is not weakly cogent (note that $\{B\} >_{cog} \{C\}$ and $\{B\} \in Cog(AF)$).

2. Argumentation games on argumentation frameworks

The abstract character of argumentation frameworks calls for highly general notions of acceptance. Here we enumerate the most common features of dialogical models of proof procedures:

– *Two party*: one is the "proponent" (**P**), who defends an argument, and the the other the "opponent" (**O**), who aims to defeat it.

– *Zero-sum game*: only one of the players wins.

– *Finiteness*: the number of arguments put forward by both **P** and **O** is finite.

These features seem to provide a very basic underlying framework for any dialogue game ([12], etc.). Other, more comprehensive notions of acceptance can be defined on the basis of this model by adding further *protocol rules*, without changing or removing any rule of the original game. Next we redefine argumentation games in precise game-theoretical terms as two party zero-sum games [13], following in general the formalism in [16]. In subsequent sections we will deal with our main goal, which is to find specific protocol rules making argumentation games finite and to show their relationships with the defense criteria mentioned in Section 1.

Definition 8. An *argumentation game* on an argumentation framework $AF = \langle \mathcal{A}, \succ \rangle$ is a zero-sum extensive game in which:

1. There are two players, i and $-i$, who play the roles of **P** and **O**, respectively. The following rules state the legal moves of the players according to the role they assume in the game (for the sake of simplicity, we will call them 'player' **P** and 'player' **O**, regardless which player, i or $-i$, assumes each role).[3]
2. A *history* in the game is any sequence $A_0, A_1, A_2, \ldots, A_{2k}, A_{2k+1}, \ldots$ of choices of arguments in $\in \mathcal{A}$ made by the players in the game. A_{2k} corresponds to **P** and A_{2k+1} to **O**, for $k = 0, 1, \ldots$.
3. At any history, A_0 is the argument that player **P** intends to defend.
4. In a history the choices by a player i at a level $k > 0$ are $C_i(k) = \{A \in \mathcal{A} : \exists B \in C_{-i}(k-1), A \succ B\}$.
5. A history of finite length K, $A_0, \ldots, A_K$, is *terminal* if A_K corresponds to player $-i$ and $C_i(K+1) = \emptyset$.
6. Payoffs are determined at terminal histories: at $A_0, \ldots, A_K$, **P**'s payoff is 1 if K is even (*i.e.*, **O** can not reply the last **P**'s argument), and -1 otherwise. In turn, **O**'s payoff at $A_0, \ldots, A_K$ is 1 if K is odd and -1 otherwise.

A game in which **P** intends to defend an argument A can be seen as a *rooted tree*, in which A is the root. Each non-terminal node at level l consists of a history $A_0, \ldots, A_l$ and its children are all the histories $A_0, \ldots, A_l, A_{l+1}$. The terminal nodes are, of course, the terminal histories.[4]

Definition 9. A *strategy* for a player i is a function that assigns an element $A_{l+1} \in C_i(l)$ at each non-terminal history $A_0, \ldots, A_l$ where A_l corresponds to player $-i$. A strategy W of the player i in game in which **P** intends to defend argument A will be denoted by '$W(A)_i$'.

Note that any pair of strategies, one for each player, say $W_i(A)$ and W_{-i} determine a unique terminal history.

Definition 10. A strategy $W(A)_i$ of player i in a game in which **P** defends A is said a *winning strategy* if for every strategy chosen by $-i$, $W(A)_{-i}$, the ensuing terminal history yields a payoff 1 for player i.

If **P** has a winning strategy, it means that her initial argument can be defended against any possible attack. On the contrary, if **O** has a winning strategy, it means that **P** fails at defending her initial argument.

Notice that winning strategies for either **P** or **O** cannot be ensured to exist if the game tree is infinite. Even being finite, an argumentation framework which is not free of cycles in the attack relation can yield an infinite tree.

Let $W(A)_{\mathbf{P}}$ the set of arguments played by **P** in a winning strategy at a game in which she defends argument A. It is easy to see that **P** has a winning strategy for every argument belonging to $W(A)_{\mathbf{P}}$:

[3] We follow here the usual convention of calling a generic player i and the other $-i$.

[4] The usual custom is to identify each node with the *last* argument in the corresponding history. So for instance the node for $A_0, \ldots, A_l$ is denoted just A_l.

Lemma 1 *Let A be an argument that* **P** *can defend with a winning strategy. Then* **P** *has a winning strategy for every argument $B \in W(A)_{\mathbf{P}}$.*

Proof. Trivial. Suppose that **P** does not have a winning strategy for $B \in W(A)_{\mathbf{P}}$. It means that once **P** plays B, she can no longer force the game towards a terminal history in which she gets 1. But then, when the game for argument A leads **P** to utter B, the game can follow a history in which **P** does not win, contradicting the assumption that $W(A)_{\mathbf{P}}$ is a winning strategy. □

Lemma 2 *The set $W(A)_{\mathbf{P}}$ of all the arguments played by* **P** *in her winning strategy $W(A)$ constitutes an admissible and strongly admissible set of arguments.*

Proof. First we prove that $W(A)_{\mathbf{P}}$ satisfies condition (1) (the conflict-free condition is implicit). Assume **P** has a winning strategy $W(A)$ to defend A and assume by contradiction that $W(A)_{\mathbf{P}}$ is not admissible. This implies two alternative cases: a) $W(A)_{\mathbf{P}}$ is not conflict-free; b) there exists some argument $B \in W(A)_{\mathbf{P}}$ which is not acceptable w.r.t. $W(A)_{\mathbf{P}}$. But then:
a) The hypothesis that $B, C \in W_{\mathbf{P}}$ are such that $B \succ C$ leads to a contradiction, since it implies that if **P** plays C then **O** can win the game by playing B and following the same sequence of moves that **P** would play in $W(A)$ if she were to defend B.
b) It follows that there exists some $C \in \mathcal{A}$ such that $C \succ B$ and for every $D \in W(A)_{\mathbf{P}}$ it is not the case that $D \succ C$. Then **O** can win the game by playing C when **P** plays B. This, in turn, implies that B does not belong to a winning strategy of **P**, contradicting the hypothesis.
To prove that $W(A)_{\mathbf{P}}$ satisfies condition (2) just note that no history built on this strategy can have cycles, otherwise it would be infinite and the strategy would not be winning. □

Proposition 1 *Let $S \subseteq \mathcal{A}$ be the set of all the arguments that can be defended by* **P** *with a winning strategy. The set $\bigcup_{A \in S} W(A)_{\mathbf{P}}$ is an admissible and strongly admissible set of arguments.*

Proof. Given lemma 2 we need only prove that $\bigcup_{A \in S} W(A)_{\mathbf{P}}$ is conflict-free. Let $A, B \in S$, and suppose by contradiction that $A' \in W(A)_{\mathbf{P}}$ and $B' \in W(B)_{\mathbf{P}}$ are such that $A' \succ B'$. By lemma 1 we have that **P** has a winning strategy $W(A')$ for A'. But then $W(A')$ is also a winning strategy for **O** in the game against B'. Hence, $W(B)$ is not a winning strategy for **P**. Contradiction. □

On the other hand, an argument that belongs to a defensible and strongly defensible but not conflict-free set is not ensured to be defended by **P** with a winning strategy. To see this, consider the argumentation framework $\langle \{A, B, C\}, \{(A, B), (B, C), (C, A)\} \rangle$ in which some (strange) semantics could sanction the only extension $\{A, B, C\}$, satisfying both admissibility and strong admissibility; nevertheless, neither argument in that extension can be defended by **P** (the cycle of attacks that includes them makes the game infinite). Of course, this semantics does not satisfy the conflict-free criterion, unlike all the known semantics in the literature.

3. Defining different protocols for finite argumentation games

Argumentation games, as defined in definition 2, are not necessarily finite. Given the rules of the game, it is clear that the source of non-finiteness is the possible existence of cycles in the attack relation. So there are basically two ways to avoid infiniteness: a) restricting argumentation games to argumentation frameworks in which the attack relation is acyclic; b) adding rules forbidding either one or both of the players to repeat arguments in some specific way. As the first alternative seems too drastic, we (as usual in the literature) will follow the latter one.

3.1. Protocols capturing admissibility related criteria

Dung's grounded semantics is the best known strongly admissible semantics. It sanctions only one extension, the least fixed point of the characteristic function $F(S)$ (def. 1). No protocol needs to be added to the rules 1-6 to capture grounded semantics. This is shown, *mutatis mutandis*, in [12] and [16]. It is also known that every semantics satisfying strong admissibility is "covered" by grounded semantics (cf. [3])). So, the game protocol needs only be extended to ensure the finiteness of the game.

PROTOCOL 1 *Rules 1-6 plus*

7. ***P** is not allowed to play an argument that was previously played by either player.*

The following two results are immediate considering what we said above about grounded semantics (proofs omitted):

Proposition 2 *Let $A \in E$ for any extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$ of a semantics $\mathcal{S}$ satisfying the conflict-free, admissibility and strong admissibility criteria. Then **P** has a winning strategy for A under protocol 1.*

Proposition 3 *If **P** has a winning strategy for A under protocol 1 then $A \in E$ for some set $E \subseteq \mathcal{A}$ satisfying strong admissibility (i.e. condition* (2)*).*

The next protocol captures admissibility and its ensuing properties were, in essence, derived by Vreeswijk and Prakken [19].

PROTOCOL 2 *Rules 1-6 plus*

7. *Neither player is allowed to advance an argument that was previously played by **O** (if **P** repeats **O** then she is introducing a conflict within her own strategy; if **O** repeats **O** then she is repeating an unsuccessful counter-argument).*
8. *Neither player is allowed to move if the last argument in the sequence was previously played by **P** (*i.e.*, if the next move corresponds to **O** and **P** has repeated herself then **O** looses –her strategy failed; if the next move corresponds to **P** and **O** has repeated an argument that was previously played by **P** then **P** looses –**O** has made an* eo ipso *move).*

The corresponding results arise from what is known about games for preferred semantics ([19]) (proofs omitted):

Proposition 4 *Let $A \in E$ for any extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$ of a semantics $\mathcal{S}$ satisfying the admissibility criterion. Then **P** has a winning strategy for A under protocol 2.*

Proposition 5 *If **P** has a winning strategy for A according to protocol 2 then $A \in E$ for some set $E \subseteq \mathcal{A}$ satisfying admissibility (i.e. condition* (1)*).*

3.2. Protocols capturing cogency-related criteria

The only case in which a pairwise cogent set S of arguments is not admissible is when S is attacked by some self-attacking argument: Assume S is pairwise cogent but not admissible. Then there exist $A \in S$ and $B \in \mathcal{A}$ such that $B \succ A$ and $S \not\succ B$. But if $\{B\}$ is conflict-free then $\{B\} >_{cog} S$. Nevertheless, S is pairwise cogent hence $\{B\} \not>_{cog} S$. Therefore, $\{B\}$ is not conflict-free, implying that B is self-attacking. This leads us to the following protocol:

PROTOCOL 3 *Protocol 2 plus*

9. *Neither player is allowed to play a self-attacking argument.*

Proposition 6 *Let $A \in E$ for any extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$ of a semantics $\mathcal{S}$ satisfying the pairwise cogent criterion. Then **P** has a winning strategy for A under protocol 3.*

Proof. Assume that **P** has no winning strategy for A. Then, there exists some history leading **O** to win, which means that in this history either a) **O** played an *eo ipso*, or b) all the arguments that attack the last move of **O** were already deployed by **O**, or c) all the arguments that attack the last move of **O** are self-attacking. Cases a) and b) imply that there exists an odd-length cycle of attacks involving an argument played by **P**. Hence, for every set S that **P** tries to build to defend A, there exists some set S' such that $S' >_{cog} S$. Hence A cannot belong to a pairwise cogent set. For case c), assume B is the last argument played by **O** and for every argument C, if $C \succ B$ then $C \succ C$. Then $\{B\}$ is pairwise cogent and for every set S that **P** tries to build for A, $\{B\} \succ S$. □

Proposition 7 *If **P** has a winning strategy for A according to protocol 3, then $A \in E$ for some set $E \subseteq \mathcal{A}$ satisfying pairwise cogency (i.e. condition* (3)*).*

Proof. Assume **P** has a winning strategy $W(A)_{\mathbf{P}}$ for A following protocol 3 and consider any history H of W in which **O** plays optimally. Let $H_{\mathbf{P}}$ be the set of arguments used by **P** in H and assume by contradiction that $H_{\mathbf{P}} \notin Cog(AF)$. That implies that there exists some set $S >_{cog} H_{\mathbf{P}}$. But since **O** has not played an *eo ipso* move in H, $H_{\mathbf{P}}$ is conflict-free, hence there exist some arguments $B \in S$ and $C \in H_{\mathbf{P}}$ such that $B \succ C$ and $H_{\mathbf{P}} \not\succ B$. Then **O** could win the game by playing B against C. Contradiction. □

To show that an argument A belongs to some weakly cogent extension, in a first stage a player i (playing the role of **P**) must defend A by constructing a pairwise cogent set. But if she fails, a second stage begins in which i will try to shift the burden of proof forcing $-i$ to sustain her opposition on a better "theory", that is, showing that her attacking arguments come from a pairwise cogent set. Then i will assume now the role of **O**). Finally, the game is won by the player who succeeds at this stage. The protocol is presented in algorithmic form for the sake of clarity of exposition.

PROTOCOL 4 *The game follows this algorithm:*

begin `% The game starts distributing the roles of proponent and`
`% opponent to the players`
P := i
O := $-i$
Play *Protocol 3* `% Stage 1:` i `must defend` A
if **P** *succeeds at defending* A then
Payoff(**P**) := 1
Payoff(**O**) := −1
else `% ` i `shifts the burden of proof on` $-i$
P *chooses an argument* B *among those non* eo ipso *moves played by* **O** *at Stage 1.*
P := $-i$
O := i
Play *Protocol 3* `% Stage 2:` $-i$ `must defend` B
if **P** *succeeds at defending* B then
Payoff(**P**) := 1
Payoff(**O**) := −1
else
Payoff(**P**) := −1
Payoff(**O**) := 1
end.

For example, let us suppose that player i (**P**) is aimed to defend A in the argumentation framework depicted in Figure 2(a). Figure 2(b) shows the game tree following Protocol 4. At stage 1, the right branch of the tree is obtained up from the only strategy for **O** that can win the stage game. At stage 2, i shifts the burden of proof challenging $-i$ to defend her argument C (the only non *eo ipso* move of $-i$). Then $-i$ assumes the role of **P** but fails to defend C. Therefore, i wins the overall game. The reason why $-i$ should not defend her *eo ipso* move at stage 2 is that it does not make sense to oblige an opponent to defend the same arguments defended by the proponent. Moreover, note that the choice by i among the arguments played by $-i$ at stage 1 is not necessarily deterministic: any legal choice would yield the same result. Nevertheless, that indeterminacy can be avoided for computational purposes by forcing the choice of, for instance, the first argument played by $-i$.

Proposition 8 *Let $A \in E$ for any extension $E \in \mathcal{E}_{\mathcal{S}}(AF)$ of a semantics $\mathcal{S}$ satisfying the weakly cogent criterion. Then* ***P*** *has a winning strategy for A under protocol 4.*

Proof. ~~By way of contradiction, assume that **P** has no winning strategy for A.~~ If A is self-attacking then is clearly excluded from any weakly cogent set. Otherwise there exists some history in which **O** wins the second stage game. This means that **P** failed to build a pairwise cogent set for A at the first stage game, which implies that for any set S such that $A \in S$, **O** can play an argument B such that for some set S', $B \in S'$ and $S' >_{cog} S$. At stage 2, let B be the argument that **O** defends. Since **P** had no winning strategy, no matter which set of arguments T is used by **P** in her strategy in the second stage game, **O** can play a strategy with arguments belonging to some set of arguments

By contraposition

T' such that $B \in T'$ and $T' \geq_{cog} T$. Therefore, A cannot belong to any weakly cogent set of arguments. ~~Contradiction.~~ □

Proposition 9 *If* ***P*** *has a winning strategy for A following protocol 4 then $A \in E$ for some set $E \subseteq \mathcal{A}$ satisfying weak cogency (i.e. condition* (4)*).*

Proof. Assume **P** has a winning strategy for A in a game played according to protocol 4. Let us analyze two cases: a) **P** wins at stage 1; b) **P** wins only at stage 2.
a) In this case, the winning strategy $W(A)_{\mathbf{P}}$ is a pairwise cogent set and, hence, a weakly cogent set too.
b) Assume $i = \mathbf{P}$ has no winning strategy for A at stage 1 but has one at stage 2. First note that if **P** cannot win the game at stage 1 she will be unable to build a set for A in $Cog(AF)$, that is, for every set S such that $A \in S$ there has to exist some set S' such that $S' >_{cog} S$. Now let B be an optimal choice of **P** among the non *eo ipso* arguments played by **O** at stage 1. This choice forces $-i$ to defend B at stage 2. Since at the second stage game $i = \mathbf{O}$ wins the game, it must be that $-i = \mathbf{P}$ will not be able to build a set in $Cog(AF)$ for B. Hence, for any set S' such that $B \in S'$ and $S' >_{cog} S$ for any set S such that $A \in S$, there exists some subset S'' (containing some of the arguments played by **P** at stage 2) such that $S'' >_{cog} S'$. This implies that S satisfies (4). □

4. Conclusion

Dialogue games are frequently studied as proof-theoretical counterparts of extension semantics. Our work goes a step further, relating dialogue games and general criteria of argument defense. We have redefined argumentation games in precise game-theoretical terms, particularly in the already commonly assumed framework: two-party, zero-sum games. We have shown that particular protocols of play can be defined in such way that the arguments in the ensuing winning strategies constitute the extensions that satisfy general criteria of defense. Moreover, all the protocols presented here ensure the finiteness of the games.

In particular, we have shown this characterization for two defense criteria weaker than admissibility: pairwise and weak cogency. These criteria exploit the intuition that acceptability depends on how two opposing argument strategies are able to shield their arguments. Unlike admissibility, which requires to defend an argument against *any* attack, cogency requires to defend an argument only against *coordinated* attacks. This yields an immediate dialectical interpretation: the opponent is urged to build an argument strategy at least as convincing as the proponent's one. Pairwise cogency offers a rational criterion for solving undesired interferences of self-attacking arguments. Weak cogency also enables *rationalizable* choices, in the game-theoretical meaning of the term, of arguments involved in odd-length cycles of attack. Among non admissible extension semantics, sustainable semantics [5] satisfies pairwise cogency while lax semantics [4] satisfies weak cogency. $CF2$ [2] and stage [17] semantics, on the other hand, do not satisfy weak cogency, though their behavior is often similar to that of lax semantics with respect to self-attacking arguments and odd-length cycles of attack. With respect to $CF2$ semantics we already mentioned the counterexample offered by Figure 2 (a). For stage semantics, check the argumentation framework $\langle\{a, b, c\}, \{(a, a), (a, b), (b, c), (c, a)\}\rangle$

where the stage extension $\{c\}$ is not weakly cogent because $\{b\} >_{cog} \{c\}$ and $\{b\}$ is pairwise cogent.

Among future research lines we stress the development of detailed algorithms for computing the arguments that can be defended by the proponent's winning strategies under each protocol.

Aknowledgements

We thank three anonymous reviewers for helping criticisms that improved the paper.

References

[1] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, *Annals of Mathematics and Artificial Intelligence* **34** (1–3) (2002), 197–215.

[2] P. Baroni, M. Giacomin, G. Guida, SCC-recursiveness: a general schema for argumentation semantics, *Artificial Intelligence* **168** (1–2) (2005), 162–210.

[3] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, *Artificial Intelligence* **171** (10–15) (2007), 675–700.

[4] G. Bodanza, The "lottery paradox" paradox and other self-attacking arguments from the point of view of defeasible argumentation frameworks, in W. Carnielli, M. Coniglio, I. D'Ottaviano (eds), *The Many Sides of Logic*. Studies in Logic 21, College Publications, 2009, 481–496.

[5] G. Bodanza, F. Tohmé, Two approaches to the problems of self-attacking arguments and general odd-length cycles of attack, *Journal of Applied Logic* **7** (4) (2009), 403–420.

[6] M. Caminada, On the issue of reinstatement in argumentation, in *Procs. of Logics in Artificial Intelligence, 10th European Conference, JELIA*, LNCS 4160, Springer (2006), 111–123.

[7] C. Cayrol, S. Doutre, J. Mengin, On decision problems related to the preferred semantics for argumentation frameworks, *Journal of Logic and Computation* **13** (3) (2003), 377–403.

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

[9] P.M. Dung, P. Mancarella, F. Toni, Computing ideal sceptical argumentation, *Artificial Intelligence* **171** (10–15) (2007), 642–674.

[10] P.E. Dunne, T.J.M. Bench-Capon, Two party immediate response disputes: Properties and eficiency, *Artificial Intelligence* **149** (2) (2003), 221–250.

[11] H. Jakobovits, D. Vermeir, Dialectic semantics for argumentation frameworks, in *Procs. of the 7th ICAIL*, Oslo, ACM, 1999, 53–62.

[12] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in I. Rahwan, G. R. Simari (eds), *Argumentation in AI*, Springer–Verlag, 2009, 105–132.

[13] M. Osborne, A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge (MA), 1994.

[14] I. Rahwan, M. I. Madakkatel, J.-F. Bonnefon, R.N. Awan, S. Abdallah, Behavioral experiments for assessing the abstract argumentation semantics of reinstatement, *Cognitive Science* **34** (2010), 1483–1502.

[15] P.M. Thang, P.M. Dung, N.D. Hung, Towards a common framework for dialectical proof procedures in abstract argumentation, *Journal of Logic and Computation* **19** (6) (2009), 1071–1109.

[16] I. Viglizzo, F. Tohmé, G. R. Simari, The foundations of DeLP: defeating relations, games and truth values, *Annals of Mathematics and Artificial Intelligence* **57** (2) (2010), 181–204.

[17] B. Verheij, Two approaches to dialectical argumentation: admissible sets and argumentation stages, in J.-J.Ch.Meyer and L.C. van der Gaag (eds), *Procs. of the Eighth Dutch Conference on Artificial Intelligence (NAIC '96)*, Utrecht University, 1996, 357–368.

[18] G. A. W. Vreeswijk, Defeasible dialectics: A controversy-oriented approach towards defeasible argumentation, *Journal of Logic and Computation* **3** (1993), 3–27.

[19] G. A. W. Vreeswijk, H. Prakken, Credulous and sceptical argument games for preferred semantics, in *Procs. 7th European Workshop on Logic for Artificial Intelligence* (2000), 239–253.