What Is a Reasonable Argumentation Semantics?

Sarah Alice Gaggl, Sebastian Rudolph, and Michaël Thomazo*

Technische Universität Dresden, Computational Logic Group, Germany

Abstract. In view of the plethora of different argumentation semantics, we consider the question what the essential properties of a "reasonable" semantics are. We discuss three attempts of such a characterization, based on computational complexity, logical expressivity and invariance under partial duplication, which are satisfied by most, if not all, known semantics. We then challenge each of these proposals by exhibiting plausible semantics which still not satisfy our criteria, demonstrating the difficulty of our endeavor.

Keywords: abstract argumentation, complexity, expressiveness, invariance under modifications.

1 Introduction

Since initiated by Dung's seminal paper [12], the field of abstract argumentation has attracted a lot of interest from researchers all over the world. One striking phenomenon in the community that sets it apart from other fields related to logic and knowledge representation is the past and ongoing proliferation of the proposed different semantics [3,15].

Typically, new argumentation semantics are motivated by providing scenarios where existing semantics do not exhibit the wanted behavior. This phenomenological and case-based approach is certainly worthwhile in a "pioneering phase", where the space of possibilities needs to be explored. However, with the field advancing and becoming more mature, it becomes more and more important to categorize and compare the different proposals for argumentation semantics as well as to identify common principles.

There is a lot of ongoing work on this along different lines. Most notably, argumentation semantics can be distinguished and classified according to the computational complexities of the associated reasoning tasks [14]. Note that evaluation criteria and rationality postulates have been discussed for abstract argumentation and its many variations and extensions [4,9].

With the wide range of existing argumentation semantics and many criteria around that help distinguishing them, it seems interesting to ask for commonalities shared by all semantics that are considered "reasonable". Is it possible to

* The third author is supported by the Alexander von Humboldt Foundation.

T. Eiter et al. (Eds.): Brewka Festschrift, LNAI 9060, pp. 309–324, 2015.

identify a "common core" of criteria that would characterize minimal requirements to an argumentation semantics? This paper discusses three properties shared by most, if not all, current semantics:

- Computational complexity. Reasoning tasks associated with argumentation semantics seem to be situated at a rather low (mostly the first, not more than the second) level of the polynomial hierarchy.
- Expressibility in monadic second-order logic (MSO). This logic has been propagated as an appropriate language for defining argumentation semantics [17]. This proposal insinuates that any "reasonable" semantics should be MSO-expressible.
- Invariance under duplication of parts of the framework. To the best of our knowledge, this criterion has not been proposed in the literature, but we found it intuitive and indeed, widely applicable.

For each of the three criteria, we show that they are satisfied by a majority of argumentation semantics. On the other hand, we critically scrutinize their universal validity and succeed in coming up with semantics which violate them while still being intuitively reasonable.

The paper is organized as follows. In Section 2 we recall the background on abstract argumentation frameworks and computational complexity. In Section 3 we introduce a semantics based on a game-theoretic approach and show that its computational complexity is higher than for the standard semantics. Then, in Section 4 we consider MSO-expressibility and exhibit a seemingly natural semantics which is not expressible in MSO logic. Section 5 is dedicated to the study of the behavior of semantics when an AF contains "structural duplicates". Finally, in Section 6 we conclude the article and point out possible future directions.

2 Preliminaries

In this section we introduce the basics of abstract argumentation, the semantics we need for further investigations and recall necessary notions from complexity theory.

Abstract Argumentation. We start with a definition of abstract argumentation frameworks following [12].

Definition 1. *An* argumentation framework *(AF) is a pair $F = (A, R)$, where A is a finite set of arguments and $R \subseteq A \times A$. The pair $(a, b) \in R$ means that a attacks b. A set $S \subseteq A$* defeats *b (in F) in symbols $S \rightarrowtail b$, if $\exists a \in S$, s.t. $(a, b) \in R$. An $a \in A$ is* defended *by $S \subseteq A$ (in F) iff, $\forall b \in A$, it holds that, if $(b, a) \in R$, then S defeats b (in F). An $a \in A$ is* in conflict with *$b \in A$, if either $(a, b) \in R$ or $(b, a) \in R$.*

The inherent conflicts between the arguments are solved by selecting subsets of arguments, where a semantics σ assigns a collection of sets of arguments to an AF F. The basic requirement for all semantics is that the sets are conflict-free.

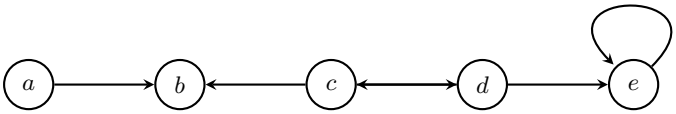

Fig. 1. AF F from Example 1

Definition 2. *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is said to be* conflict-free *(in F), if there are no $a, b \in S$, such that $(a, b) \in R$. We denote the collection of sets which are conflict-free (in F) by $cf(F)$. A set $S \subseteq A$ is maximal conflict-free or* naive*, if $S \in cf(F)$ and for each $T \in cf(F)$, $S \not\subset T$. We denote the collection of all naive sets of F by $naive(F)$. For the empty AF $F_0 = (\emptyset, \emptyset)$, we set $naive(F_0) = \{\emptyset\}$.*

Towards definitions of the semantics we introduce the following formal concepts [12].

Definition 3. *Given an AF $F = (A, R)$ and some $S \subseteq A$, the* characteristic function $\mathcal{F}_F : 2^A \to 2^A$ *of F is defined as $\mathcal{F}_F(S) = \{x \in A \mid x \text{ is defended by } S\}$.*

We consider the following semantics.

Definition 4. *Let $F = (A, R)$ be an AF. A set $S \in cf(F)$ is said to be*

- *a* stable *extension (of F), i.e. $S \in stable(F)$, if $S^+_R = A$ where $S^+_R = S \cup \{a \mid \exists b \in S : (b, a) \in R\}$ is the* range *of S;*
- *an* admissible *extension, i.e. $S \in adm(F)$ if each $a \in S$ is defended by S;*
- *a* complete *extension (of F), i.e. $S \in comp(F)$, if each $S \in adm$ and for each $a \in A$ defended by S (in F), $a \in S$ holds;*
- *a* preferred *extension, i.e. $S \in prf(F)$ if $S \in adm(F)$ and for each $T \in adm(F)$, $S \not\subset T$;*
- *the* grounded *extension (of F), i.e. the unique set $S \in grd(F)$, is the least fixed point of the characteristic function $\mathcal{F}_F$.*

AFs are typically represented as directed graphs where the nodes correspond to the arguments and the edges to the attacks.

Example 1. Let $F = (A, R)$ be an AF with arguments $A = \{a, b, c, d, e\}$ and attacks $R = \{(a, b), (c, b), (c, d), (d, c), (d, e), (e, e)\}$. The corresponding graph is depicted in Figure 1. F has the following sets of extensions for the introduced semantics, $stable(F) = \{\{a, d\}\}$, $prf(F) = \{\{a, c\}, \{a, d\}\}$, $comp(F) = \{\{a\}, \{a, c\}, \{a, d\}\}$, $grd(F) = \{\{a\}\}$ and $adm(F) = \{\{\}, \{a\}, \{c\}, \{d\}, \{a, c\}, \{a, d\}\}$.

Computational Complexity. We assume the reader to be familiar with standard complexity classes, i.e. P, NP, coNP and PSPACE (polynomial space). Nevertheless, we briefly recapitulate the concept of oracle machines and some related

Table 1. Complexity of decision problems ($\mathcal{C}$-c denotes completeness for class $\mathcal{C}$)

	Ver_σ	$Cred_\sigma$	$Skept_\sigma$	$Exists_\sigma^{\neg\emptyset}$
naive	in P	in P	in P	in P
stable	in P	NP-c	coNP-c	NP-c
adm	in P	NP-c	trivial	NP-c
comp	in P	NP-c	P-c	NP-c
prf	coNP-c	NP-c	Π_2^{P}-c	NP-c

complexity classes. Let $\mathcal{C}$ notate some complexity class. By a $\mathcal{C}$-oracle machine we mean a (polynomial time) Turing machine which can access an oracle that decides a given (sub)-problem in $\mathcal{C}$ within one step. We denote the class of decision problems, that can be solved by such machines, as $\mathrm{P}^{\mathcal{C}}$ if the underlying Turing machine is deterministic and $\mathrm{NP}^{\mathcal{C}}$ if the underlying Turing machine is non-deterministic. The class $\Sigma_2^{\mathrm{P}} = \mathrm{NP}^{\mathrm{NP}}$, denotes the problems which can be decided by a non-deterministic polynomial time algorithm that has access to an NP-oracle. The class $\Pi_2^{\mathrm{P}} = \mathrm{coNP}^{\mathrm{NP}}$ is defined as the complementary class of Σ_2^{P}, i.e. $\Pi_2^{\mathrm{P}} = \mathrm{co}\Sigma_2^{\mathrm{P}}$. The relations between the complexity classes used in this work are $\mathrm{P} \subseteq \mathrm{NP}$ (coNP) $\subseteq \Sigma_2^{\mathrm{P}}$ (Π_2^{P}) $\subseteq$ PSPACE.

We are interested in the following decision problems (for a semantics σ).

- $Cred_\sigma$: Given AF $F = (A, R)$ and $a \in A$. Is a contained in *some* $S \in \sigma(F)$?
- $Skept_\sigma$: Given AF $F = (A, R)$ and $a \in A$. Is a contained in *each* $S \in \sigma(F)$?
- Ver_σ: Given AF $F = (A, R)$ and $S \subseteq A$. Is $S \in \sigma(F)$?
- $Exists_\sigma^{\neg\emptyset}$: Given AF $F = (A, R)$. Does there exist a set $S \subseteq A, S \neq \emptyset$ such that $S \in \sigma(F)$?

The complexity landscape for the semantics considered in this article is given in Table 1 (see [10,11,13]).

3 About Computational Requirements

The decision problems associated with most of the classical Dung semantics have relatively low computational complexity. Indeed, most of the problems belong to the first level of the polynomial hierarchy, with the exception of the skeptical acceptance for preferred semantics, which is complete for the second level of the polynomial hierarchy (see Table 1). This is an appreciable feature, since it allows one to use efficient solvers developed in other communities, such as satisfiability solvers or answer-set programming solvers, either directly or used as oracles in a more complex algorithm [18,16].

In our task of compiling a set of properties that reasonable semantics should fulfill, should we include an upper-bound on the complexity of classical reasoning problems? To tackle this question, we adopt a game-oriented approach, as in [19].

As already argued in the literature, games are a natural approach to argumentation, since they fit with the intuition of an iterative process. Their main use with respect to abstract argumentation frameworks has been to provide alternative characterizations of credulously and skeptically accepted arguments according to a variety of known semantics. We adopt a dual approach here, using games to define novel semantics. We then explore the complexity of reasoning under such semantics.

This section is organized as follows:

- first, we argue that shifting the focus from the extension to the way it has been built may allow to distinguish between extensions that would otherwise be similar; in particular, we present a way to structurally (and partially) rank preferred extensions;
- second, we explore semantics that can be expressed thanks to games. In particular, we introduce a semantics whose credulous acceptance problem is PSPACE-complete. While this semantics is maybe not immediate, we believe that it demonstrates a still unexplored space of game-based semantics of high complexity.

Dynamics of Preferred Extensions. In the argumentation semantics that we have presented so far, all arguments are chosen simultaneously. That is, an acceptance condition is checked on a potential extension, but the way this extension has been created is ignored. In [19], this is indicated as not fully intuitive, as argumentation often refers to an iterative process, where arguments are given one at a time. We provide another reason to pay attention to the generation process and not only to its result: it provides further structural insights on how to distinguish extensions.

To illustrate our point, let us consider the argumentation framework F_d drawn in Figure 2. This framework has two preferred extensions, which are $\{1, 2\}$ and $\{3, 4\}$. By looking only at these two sets, there is no reason to distinguish one from the other: both are preferred, both are of the same size, and we assume not to have any preference information on the arguments. However, let us assume that arguments are added one at a time. There are two ways to generate the first extension: either choose 1 then 2, or choose 2 and then 1. Similarly, there are two ways to generate the second extension. By looking at these sequences, the two extensions can then be distinguished: at any step, the set built towards the second extension is admissible. Indeed, $\{3\}$ is admissible, as well as $\{3, 4\}$. This is the case neither for $\{1\}$ nor for $\{2\}$. This means that the extension $\{3, 4\}$ can be generated by constructing only admissible sets, whereas $\{1, 2\}$ cannot. Let us define formally a malus function on preferred extensions.

Definition 5 (Malus of a preferred extension). *Let F be an argumentation framework, and let S be a preferred extension of F. Let $(s_1, \ldots, s_n)$ be a sequence of elements of S where each element of S appears exactly once. The malus of $(s_1, \ldots, s_n)$ is the number of indices i such that $1 \leq i \leq n$ and $\{s_1, \ldots, s_i\}$ is not an admissible set. The malus of S is the minimal malus of any such a sequence $(s_1, \ldots, s_n)$.*

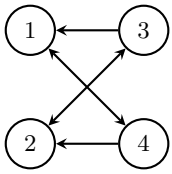

Fig. 2. The argumentation framework F_d

Equipped with this notion of malus, we can now define a notion of preference on preferred extensions.

Definition 6. *Let F be an argumentation framework, S_1 and S_2 be two preferred extensions of F. S_1 is preferred to S_2 if the malus of S_1 is smaller than the malus of S_2.*

The intuition behind this notion of preference is that a (linear) argumentation, where at each step the obtained set of arguments is admissible, is more solid than one for which this is not the case.

From Semantics to Games and Back. A natural way to use information on how an extension has been generated is to look at it as the result of a game. We already mentioned that games have been used to characterize credulous and skeptical acceptance for a variety of semantics. Let us give some more details about this approach, as presented in [19]. Two player games are used, where both players, "PRO" and "CON", play alternately. PRO aims at proving credulous acceptance of an argument, while CON tries to disprove this. The set of moves allowed is defined in order to capture a given semantics. Credulous acceptance is then defined with respect to *winning strategies* of PRO. Hence, games have been used to shed a new light on already existing semantics. We adopt here a dual approach, where we start from a family of games, and explore which semantics may be defined in this way.

In the literature, an extension is defined from the arguments that PRO uses. We adopt a slightly different approach: both players "collaborate" to create an extension, but their contributions are chosen with the goal to maximize their own satisfaction. By collaborating, we mean that both players are adding arguments to what will become an extension. At each step, the arguments they can add depends on the structure of the graph and on the already played arguments. This is specified thanks to the definition of *legal sequences*. However, both players may have different objectives: this is represented by a payoff function, that associates each outcome of the game with a payoff for each player. An extension is then the set of arguments that have been played during an *optimal* play of both players.

Let us now formally define the three ingredients of an *argumentation game* $\mathcal{G}$ which we introduced informally above: legal sequences, payoff function and optimal play.

Definition 7 (Legal sequences). *A set of* legal sequences $\mathcal{S}$ *on an argumentation framework F is a finite set of tuples of arguments of F that is prefix-closed, that is, if $(a_1, \ldots, a_n) \in \mathcal{S}$, then $(a_1, \ldots, a_{n-1})$ also belongs to $\mathcal{S}$. The* outcomes $\mathcal{O}$ *of the game are sequences that are not a strict prefix of any other sequence in $\mathcal{S}$.*

A move of the game is the transition from a legal sequence to another legal sequence by adding an argument at the end. Moves resulting in a sequence of odd size are played by Player 1, while other moves are played by player 2.

Note that we could also define infinite games, but we stick to the finite case for simplicity. We now introduce payoff functions, that describe the "satisfaction" of each player after playing a given sequence.

Definition 8 (Payoff function). *Let $\mathcal{S}$ be a set of sequences, $\mathcal{O}$ be the set of outcomes. A payoff function is a function from O to $\mathbb{N} \times \mathbb{N}$. The first component is the payoff for Player 1, while the second is for Player 2.*

Players aim at maximizing their payoff, and use *strategies* in order to do so. Strategies define what to play in each given situation.

Definition 9 (Strategy). *A strategy for Player 1 (resp. Player 2) is a function that associates to each legal sequence of even length (resp. odd length) another legal sequence that can be reached by a move of Player 1 (resp. Player 2).*

Strategies of particular interest are the so-called optimal strategies.

Definition 10 (Optimal strategy). *A strategy is optimal if it maximizes the minimal payoff a player may get by playing it, whatever the opponent's strategy is.*

We now have all the tool to define the semantics associated with a game.

Definition 11 (Game semantics). *Let $\mathcal{G}$ be an argumentation game. Let $F = (A, R)$ be an argumentation framework. The extensions of F according to $\mathcal{G}$ are the set of arguments E that can be played when both players are following an optimal strategy.*

Thus, the choices of a set of legal sequences and a payoff function define a semantics for argumentation frameworks. Let us notice that some choices may violate even the most widely accepted properties of a semantic, such as language independence [4]. It is however possible to regain such a property by adequately restricting the set of legal sequences and the set of payoff functions one may use.

We now instantiate the previous definitions to define the *last-word game* and its associated semantics. The aim of each player is the following: either he/she wants to ensure that he/she will choose the last argument, or, if that cannot be ensured, he/she wants the extension to be as large as possible. At each time, they can choose any argument that maintain conflict-freeness, and that is attacked by at least one argument that was attacked by the previously chosen argument.

Definition 12 (Legal sequence for the last-word game). *Let $F = (A, R)$ be an argumentation framework. A legal sequence for the last word game is defined inductively as follows:*

- *the empty sequence is a legal sequence;*
- *(a_1) is a legal sequence for any $a_1 \in A$ such that $(a_1, a_1) \notin R$;*
- *if $(a_1, \ldots, a_n)$ is a legal sequence, and there exists $b, a_{n+1} \in A$ such that $(a_n, b) \in R$ and $(b, a_{n+1}) \in R$, and $(a_i, a_{n+1}) \notin R$ for any i with $1 \leq i \leq n+1$, then $(a_1, \ldots, a_n, a_{n+1})$ is a legal sequence.*

Definition 13 (Payoff for the last-word game). *Let $(a_1, \ldots, a_n)$ be an outcome for the last word game. The last-word payoff f_{lw} is defined as follows:*

- *if n is odd, then $f_{lw} = (|A| + n, n)$;*
- *if n is even, $f_{lw} = (n, |A| + n)$.*

Theorem 1. *Credulous acceptance for the last-word semantics is* PSPACE*-complete.*

Proof. (Sketch) Membership is direct. For hardness, we reduce the problem of the existence of a winning strategy for the first player in GENERALIZEDGEOGRAPHY to credulous acceptance under the last-word semantics. Let us first recall that GENERALIZEDGEOGRAPHY is a two player game played on a directed graph. At each step, a player chooses a non-visited vertex that is a successor of the last played vertex. The last player who can play wins. An example of instance is given in Example 2. We first create an instance G^* of GENERALIZEDGEOGRAPHY such that the first player has a winning strategy starting from one of two special moves depending on the existence of a winning strategy in the original instance G. We thus create an argumentation framework by replacing each edge in G^* by two attacks.

□

Example 2. Figure 3 presents an example of instance for GENERALIZEDGEOGRAPHY. Player 1 could play 1. Player 2 has two choices: either 2 or 3. If Player 2 plays 2, Player 1 plays 3 and wins. If Player 2 plays 3, Player 2 wins. A winning strategy for Player 1 is to play 3 from the beginning.

Figure 4 is the argumentation framework obtained from the instance of Figure 3 by the reduction used in the proof of Theorem 1. a_1, a_2 and a_3 correspond to vertices of the original instance. a_{v_1} and a_{v_2} correspond to vertices added to ensure that Player 1 has a winning strategy, starting either with a_{v_1} or a_{v_2}. Other vertices corresponds to edges in the original instance.

While possibly not overly intuitive, we believe that this semantics helps making a case for interesting semantics that incorporate information on how an extension may have been created, and such semantics are likely to have a higher computational complexity than the classical ones.

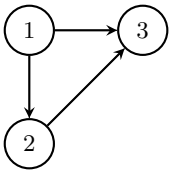

Fig. 3. An instance of GeneralizedGeography

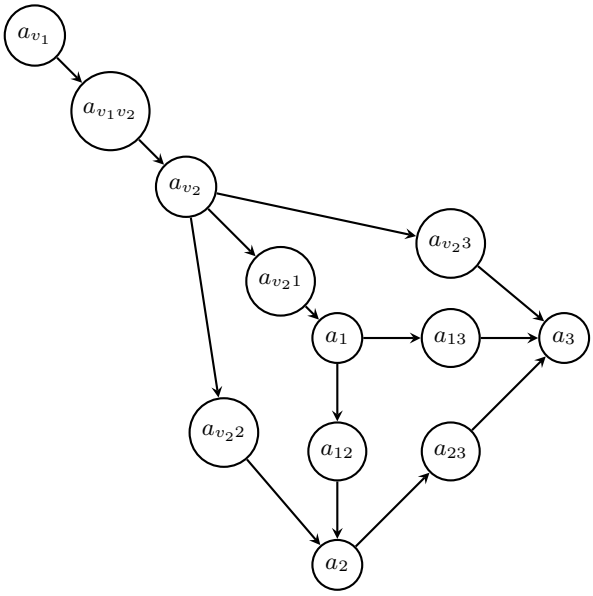

Fig. 4. The argumentation framework obtained from the instance of Example 2

4 Expressiveness of Argumentation Semantics

Another angle from which one can investigate argumentation semantics is the logical expressivity needed to define them. Note that each argumentation framework $F = (A, R)$ can be seen as a relational structure with one binary relation and therefore as a logical interpretation for a signature containing one binary predicate symbol. This perspective allows for characterizing argumentation semantics in terms of the logical expressiveness needed for defining them.

It has been argued before [17] that a significant number of semantics can be expressed by monadic second-order (MSO) logic formulae. MSO logic is an extension of first-order predicate logic by *set variables* (usually denoted by upper case letters like X) which are used to represent sets of domain elements. They can be quantified over and used in *membership atoms* of the form $x \in X$ which are interpreted in the intuitive way.

Now, an MSO formulae $\varphi[X]$ with one free set variable X can be seen as a formal definition of some semantics σ as follows: for any AF $F = (A, R)$ and $S \subseteq A$ the following holds: $S \in \sigma(F)$ iff F satisfies $\varphi[X]$ under the variable assignment $X \mapsto S$. Finally, a semantics σ is called MSO-expressible if such a defining MSO formula exists for it.

As stated above, virtually all mainstream argumentation semantics are MSO-expressible. As an example, the admissible semantics can be expressed by the following formula:

$$\forall x, y \ \big(R(x,y) \wedge (y \in X) \ \rightarrow \ \neg(x \in X) \wedge \exists z.(R(z,x) \wedge z \in X)\big)$$

Note that MSO-expressibility guarantees certain properties: the computational complexity of all the reasoning tasks described in Section 2 will be on some fixed level of the polynomial hierarchy. By contraposition, we can infer that any semantics with a complexity of PSPACE or above cannot be expressed in MSO logic. While we saw an example of this in Section 3, we focus here on the question if there is a reasonable semantics with comparably low reasoning complexity which is nevertheless not MSO-expressible. Indeed, the following semantics satisfies these properties.

Definition 14. *A set $S \in cf(F)$ is said to be a* multi-admissible *extension if $|\{b \in S \mid (a,b) \in R\}| \leq |\{c \in S \mid (c,a) \in R\}|$ holds for every $a \in A \setminus S$.*

In words, a multi-admissible extension S must attack each argument a outside S at least as often as a attacks S. We deem this a rather reasonable semantics, as it is very close to the admissible semantics (in fact, every multi-admissible extension is also admissible), but additionally takes the multiplicity of the attacks carried out by an external argument into account by requiring them to be compensated by an according number of counter-attacks.

It is straightforward to check that verifying if some set S is a multi-admissible extension can be done in polynomial time, which immediately ensures that the complexity of all other reasoning tasks is not worse than on the first level of the polynomial hierarchy.

We will next show that despite this comparably low complexities, this semantics cannot be expressed in MSO logic.

Theorem 2. *There is no MSO formula that expresses the multi-admissible semantics.*

For the proof of this theorem, we use a well known result of Büchi linking regular word languages and MSO logic.

Definition 15. *Let Σ be a finite alphabet. The* word interpretation $\mathcal{I}_w$ *of some word $w = \alpha_1 \dots \alpha_n \in \Sigma^*$ is the relational structure with base set $\{1, \dots, n\}$ the binary relation $<$ defined in the usual way, and unary relations P_α for all $\alpha \in \Sigma$ with $i \in P_\alpha^{\mathcal{I}_w}$ iff $\alpha = \alpha_i$ for every $i \in \{1, \dots, n\}$.*

Theorem 3 (Büchi [8]). *A word language $L \subseteq \Sigma^*$ is regular if and only if there exists an MSO sentence φ satisfying $L = \{w \mid \mathcal{I}_w \models \varphi\}$.*

This result is now leveraged for an indirect proof: we argue that a hypothetical MSO formula expressing multi-admissible semantics could be used to come up with an MSO formula characterizing a non-regular language.

Proof. (Sketch) Assume there is an MSO formula $\varphi[X]$ characterizing multi-admissible extensions. Let $\varphi'[X]$ be the MSO formula obtained from $\varphi[X]$ by replacing every atom $R(x, y)$ by the subformula $(P_b(x) \wedge P_c(y)) \vee (P_c(x) \wedge P_a(y))$. Thereby, we assume an abstract framework where every node is labeled by a, b or c; then we let each b-labeled node attack every c-labeled node and have each c-labeled node attack every a-labeled node.

Finally, for checking if the set of all nodes labeled with a or b can be an extension, let $\psi[X] = \forall x.(x \in X \leftrightarrow P_a(x) \vee P_b(x))$. By construction this is the case, if more nodes are labeled with b than with a.

Then, every word interpretation $\mathcal{I}_w$ corresponding to some word w over $\{a, b, c\}$ satisfies that it is a model of $\exists X.(\varphi'[X] \wedge \psi[X])$ if and only if w contains more bs than as. However, the language of all words with these properties is not regular as can be easily shown using the well-known pumping lemma for regular languages. □

From a more general perspective, MSO logic is known to be incapable of comparing cardinalities of sets of unbounded size. Thus, it will be difficult to cast any semantics relying on such a comparison into MSO logic.

5 Invariant Behavior under Modification of the Framework

The evaluation of AFs is solely based on syntactic properties. For example checking whether a set of arguments is accepted under stable semantics requires that there are no two arguments in the set which attack each other and all arguments not contained in the set are attacked by the set.

Due to the non-monotonic behavior of AFs, modifications, i.e. adding or deleting arguments or attacks, may change the outcome of a semantics in a way that arguments which have been accepted before are not acceptable afterwards. In the literature, most of the work was focused on studying equivalences, where one identifies a kernel of an AF, and if two different AFs posses the same kernel they are strongly equivalent to each other [20].

In this section we study what happens to the extensions if some part of the framework is duplicated. A duplicate will be a set of arguments which has internally and externally the same relations as its original. However, there is no connection i.e. no attacks between the original and the duplicate.

Definition 16. *Let $F = (A, R)$ be an AF. A set $D \subset A$ is a* duplicate *in F, with its original set $\hat{D} = \{\hat{a} \in A \mid a \in D\}$ satisfying $\hat{D} \cap D = \emptyset$ and $|\hat{D}| = |D|$ if there are only the following four types of attacks between arguments $d', e' \in D$, their originals $\hat{d}, \hat{e} \in \hat{D}$ and $x, y \in A \setminus (D \cup \hat{D})$.*

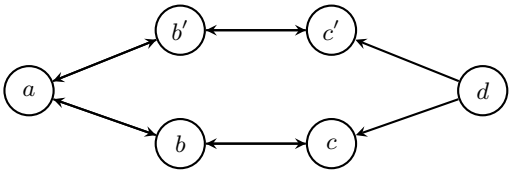

Fig. 5. AF F from Example 3

R1) $(d', e') \in R$ *iff* $(\hat{d}, \hat{e}) \in R$*;*
R2) $(d', x) \in R$ *iff* $(\hat{d}, x) \in R$*;*
R3) $(x, d') \in R$ *iff* $(x, \hat{d}) \in R$*;*
R4) $(x, y) \in R$*.*

In the following we will denote the duplicate of an argument a with a' and the original of a duplicate b with $\hat{b}$. For an AF $F = (A, R)$ and a set $D \subseteq A$ we can add a set of duplicate A' such that $A' = \{a' \mid a \in D\}$, then the obtained AF with duplicates will be denoted by $F' = (A \cup A', R \cup R')$, where R' is as in $R1 - R3$ in Definition 16.

Example 3. Consider AF from Figure 5. There, the set $A' = \{b', c'\}$ is a duplicate in F with its original set $\hat{A} = \{b, c\}$.

We say a semantics σ is *weakly duplicate invariant* if for any AF F and its related AF F' with duplicates, each σ-extension S of F is related to a σ-extension S' of F' in such a way that all arguments from S are accepted as well as those duplicates from A' if their original argument was contained in S.

Definition 17. *A semantics* σ *is* weakly duplicate invariant *if for any AF* $F = (A, R)$ *and* $F' = (A \cup A', R \cup R')$ *such that the set* A' *is a duplicate in* F'*,* R' *are the attacks related to duplicates and for any* $S \subseteq A$ *the following holds*

$$S \in \sigma(F) \Rightarrow S' \in \sigma(F'),$$

where $S' = S \cup \{a' \in A' \mid a \in S\}$*.*

Lemma 1. *Conflict-free sets are weakly duplicate invariant.*

Proof. Let $F = (A, R)$ and $F' = (A \cup A', R \cup R')$ be AFs such that the set A' is a duplicate in F', R' are the attacks related to duplicates in F'. Towards a contradiction assume there is a set $S \in cf(F)$ but $S' \notin cf(F')$, where $S' = S \cup \{a' \in A' \mid a \in S\}$, thus $S \subseteq S'$ and we can have the four following cases.

R1: $a, b \in A'$ then it follows that $(\hat{a}, \hat{b}) \in R$ with $\hat{a}, \hat{b} \in S$, a contradiction;
R2: $a \in A'$, it follows that $(\hat{a}, b) \in R$ with $\hat{a} \in S$, a contradiction;
R3: $b \in A'$, it follows that $(a, \hat{b}) \in R$ with $\hat{b} \in S$, a contradiction;
R4: (a, b) is an ordinary attack, thus $a, b \in S$, a contradiction.

□

Theorem 4. *Stable, admissible, preferred, complete and naive semantics are weakly duplicate invariant.*

Proof. Let $F = (A, R)$ and $F' = (A \cup A', R \cup R')$ be AFs such that the set A' is a duplicate in F', R' are the attacks related to duplicates in F'.

For stable semantics: Towards a contradiction assume there is a set $S \in stable(F)$ but $S' \notin stable(F')$, where $S' = S \cup \{a' \in A' \mid a \in S\}$. As $S \in stable(F)$ clearly $S \in cf(F)$, and from Lemma 1 we thus know that $S' \in cf(F')$. Hence, there is an argument $a \notin S'^{+}_{R}$ and due to the definition of the range, $a \notin S'$ and clearly $a \notin S$. Hence, for each $b \in S'$ we have $(b, a) \notin R \cup R'$. We need to consider the following four cases.

R1: $a, b \in A'$ then it follows that $(\hat{b}, \hat{a}) \notin R$, thus $\hat{a} \notin S^{+}_{R}$, a contradiction;
R2: $b \in A'$, it follows that $(\hat{b}, a) \notin R$ and $a \notin S^{+}_{R}$, a contradiction;
R3: $a \in A'$, it follows that $(b, \hat{a}) \notin R$ and $\hat{a} \notin S^{+}_{R}$, a contradiction;
R4: $a, b \in A \setminus A' \cup \hat{A}$ thus $a \notin S^{+}_{R}$, a contradiction.

For admissible semantics: Towards a contradiction assume there is a set $S \in adm(F)$ but $S' \notin adm(F')$, where $S' = S \cup \{a' \in A' \mid a \in S\}$. As $S \in adm(F)$ clearly $S \in cf(F)$, and from Lemma 1 we thus know that $S' \in cf(F')$. Hence, there is an argument $a \in S'$ which is not defended by S'. This means, there is an argument $b \in A \cup A'$ s.t. $(b, a) \in R \cup R'$ but for each $c \in S'$ we have $(c, b) \notin R \cup R'$. We can have the following eight cases.

$C1$: $(c', b') \notin R'$ and $(b', a') \in R'$, then $(\hat{c}, \hat{b}) \notin R$ and $(\hat{b}, \hat{a}) \in R$, thus $\hat{a} \in S$ is not defended by S, a contradiction;
$C2$: $(c', b') \notin R'$ and $(b', a) \in R'$, then $(\hat{c}, \hat{b}) \notin R$ and $(\hat{b}, a) \in R$, with $a \in A \setminus (A' \cup \hat{A})$, thus, $a \in S$ is not defended by S, a contradiction;
$C3$: $(c', b) \notin R'$ and $(b, a') \in R'$, then $(\hat{c}, b) \notin R$ and $(b, \hat{a}) \in R$, with $b \in A \setminus (A' \cup \hat{A})$, thus, $\hat{a} \in S$ is not defended by S, a contradiction;
$C4$: $(c', b) \notin R'$ and $(b, a) \in R$, then $(\hat{c}, b) \notin R$, with $a, b \in A \setminus (A' \cup \hat{A})$, thus, $a \in S$ is not defended by S, a contradiction;
$C5$: $(c, b') \notin R'$ and $(b', a') \in R'$, then $(c, \hat{b}) \notin R$ and $(\hat{b}, \hat{a}) \in R$, with $c \in A \setminus (A' \cup \hat{A})$, thus, $\hat{a} \in S$ is not defended by S, a contradiction;
$C6$: $(c, b') \notin R'$ and $(b', a) \in R'$, then $(c, \hat{b}) \notin R$ and $(\hat{b}, a) \in R$, with $a, c \in A \setminus (A' \cup \hat{A})$, thus, $a \in S$ is not defended by S, a contradiction;
$C7$: $(c, b) \notin R$ and $(b, a') \in R'$, then $(\hat{b}, a) \in R$, with $b, c \in A \setminus A' \cup \hat{A}$, thus, $a \in S$ is not defended by S, a contradiction;
$C8$: $(c, b) \notin R$ and $(b, a) \in R$, with $a, b, c \in A \setminus A' \cup \hat{A}$, thus, $a \in S$ is not defended by S, a contradiction;

For preferred semantics: We show that for each $S \in prf(F)$ it holds that $S' \in prf(F')$ for $S' = S \cup \{a' \in A' \mid a \in S\}$. We know that $S \in adm(F)$ and from above that also $S' \in adm(F')$. Moreover, for each $T \in adm(F)$ we have $T \subseteq S$. As in the primed version of the extension we only add arguments if their originals are contained in the non-primed version, one can easily see that for each $T' \in adm(F)$, $T' \subseteq S'$ as well, where $T' = T \cup \{a' \in A' \mid a \in T\}$. It follows that $S \in prf(F')$.

The proofs of other semantics rely on similar arguments. □

We find that all standard semantics are weakly duplicate invariant. So one can see this as a preference on argumentation semantics. Interestingly, this means that all these semantics have a monotonic behavior when duplicates are added to the framework.

Still, weak duplicate invariance is not a criterion to be taken for granted for all reasonable argumentation semantics: indeed, it is not too hard to see that the multi-admissible semantics introduced in the previous section violates this criterion.

Theorem 5. *The multi-admissible semantics is not weakly duplicate invariant.*

Proof. Consider the AF $F = (A, R)$ with $A = \{a, b, c\}$ and $R = \{(c, a), (b, c)\}$. Obviously $S = \{a, b\}$ is a multi-admissible extension of (A, R). Now consider the AF $F' = (\{a, a', b, c\}, \{(c, a), (c, a'), (b, c)\})$. Clearly, $S' = \{a, a', b\}$ is not a multi-admissible extension of F', thus we have shown our claim. □

What happens if one considers the other direction of duplicate invariance? In the following we define that a semantics σ is *strongly duplicate invariant* if for each σ-extension S' of the AF F' with the duplicate A', the extension S obtained by deleting the duplicate arguments from S' is a σ-extension of the respective AF F without duplicates.

Definition 18. *A weakly duplicate invariant semantics σ is* strongly duplicate invariant *if for any AFs $F = (A, R)$ and $F' = (A \cup A', R \cup R')$ such that the set A' is a duplicate in F', R' are the attacks related to duplicates and $S' \subseteq A \cup A'$, the following holds*

$$S' \in \sigma(F') \Rightarrow S \in \sigma(F)$$

where $S = S' \cap A$.

Theorem 6. *Stable, preferred, complete, admissible and naive semantics are not strongly duplicate invariant.*

Proof. Consider the AF $F' = (A \cup A', R \cup R')$ with arguments $A = \{a, b, c, d\}$ and $A' = \{b', c'\}$, and attacks $R = \{(a, b), (b, a), (b, c), (c, b), (c, d)\}$ and $R' = \{(a, b'), (b', a), (b', c'), (c', b'), (c', d)\}$ as depicted in Figure 6. Let $F = (A, R)$ be an AF obtained from F' without the duplicate A'. Consider the set $S' = \{b, c'\}$ which is a stable (resp. preferred, complete, admissible, naive) extension of F' but the set $S = \{b\}$ obtained from S' by deleting the duplicate argument c' is not a stable (resp. preferred, complete, admissible, naive) extension of F. □

Theorem 7. *The grounded semantics is strongly duplicate invariant.*

Proof. (Sketch) Let F be an argumentation framework, and let F' be obtained by F by duplicating some of its arguments. Let us denote $E_0 = E'_0 = \emptyset$. Let us also define $E_{i+1} = \mathcal{F}_F(E_i)$ as well as $E'_{i+1} = \mathcal{F}_{F'}(E'_i)$, where we recall that $\mathcal{F}_F$ denotes the characteristic function. We prove by induction on i that $E'_i = E_i \cup \{y' \mid \exists y \in E_i : \ y' \text{ is a duplicate of } y\}$. This proves in particular the result for the grounded extensions of F and F'. Weak and strong duplicate invariance are clear from this equality. □

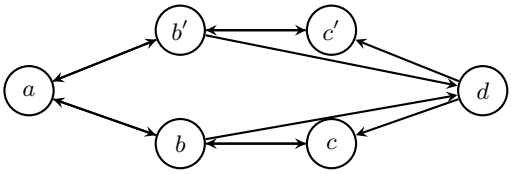

Fig. 6. AF F'

6 Conclusion

On our quest for a better understanding of the gist of argumentation semantics, we have been investigating three characteristics we found to be shared by the mainstream argumentation semantics. These characteristics were based on three general classification schemes typically encountered in theoretical computer science: computational complexity, expressibility in some logical language and invariance under certain transformations.

While this endeavor certainly enhanced our understanding of the matter, we found that for each of the criteria, counterexamples can be constructed which, arguably, still have the "look and feel" of a typical argumentation semantics. In our view, this demonstrates the nontrivial philosophical dimension of an area that tries to capture the essence of "argumentation" on an abstract and formal level.

While our studies focused on the traditional Dung-style setting, a plethora of generalizations and extensions have been proposed, such as abstract dialectical frameworks [7,6], bipolar AFs [1], preference-based AFs [2], and value-based AFs [5]. All of these new approaches would certainly benefit from a thorough study of commonalities and differences in terms of general formal properties of the diverse semantics.

Acknowledgements. As this Festschrift is dedicated to Gerd Brewka's 60 th birthday, we want to congratulate him as well as thank him for his inspiring work and the fruitful and enjoyable discussions.

References

1. Amgoud, L., Cayrol, C., Lagasquie, M.-C., Livet, P.: On bipolarity in argumentation frameworks. Int. Journal of Intelligent Systems 23, 1–32 (2008)
2. Amgoud, L., Vesic, S.: Repairing preference-based argumentation frameworks. In: Proc. IJCAI 2009, pp. 665–670 (2009)
3. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. Knowledge Eng. Review 26(4), 365–410 (2011)
4. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artif. Intell. 171(10-15), 675–700 (2007)
5. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. J. Log. Comput. 13(3), 429–448 (2003)

6. Brewka, G., Polberg, S., Woltran, S.: Generalizations of dung frameworks and their role in formal argumentation. IEEE Intel. Sys. 29(1), 30–38 (2014)
7. Brewka, G., Woltran, S.: Abstract Dialectical Frameworks. In: Proc. KR 2010, pp. 102–111 (2010)
8. Büchi, R.J.: On a decision method in restricted second order arithmetic. In: Proc. Logic, Methodology and Philosophy of Science 1960, pp. 1–11 (1962)
9. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artif. Intell. 171(5-6), 286–310 (2007)
10. Coste-Marquis, S., Devred, C., Marquis, P.: Symmetric argumentation frameworks. In: Godo, L. (ed.) ECSQARU 2005. LNCS (LNAI), vol. 3571, pp. 317–328. Springer, Heidelberg (2005)
11. Dimopoulos, Y., Torres, A.: Graph theoretical structures in logic programs and default theories. Theor. Comput. Sci. 170(1-2), 209–244 (1996)
12. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–358 (1995)
13. Dunne, P.E., Bench-Capon, T.J.M.: Coherence in finite argument systems. Artif. Intell. 141(1/2), 187–203 (2002)
14. Dunne, P.E., Wooldridge, M.: Complexity of abstract argumentation. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 85–104. Springer, US (2009)
15. Dvořák, W., Gaggl, S.A.: Stage semantics and the scc-recursive schema for argumentation semantics. J. Log. Comput. 2014 (2014)
16. Dvořák, W., Järvisalo, M., Wallner, J.P., Woltran, S.: Complexity-sensitive decision procedures for abstract argumentation. Artif. Intell. 206, 53–78 (2014)
17. Dvořák, W., Szeider, S., Woltran, S.: Abstract argumentation via monadic second order logic. In: Hüllermeier, E., Link, S., Fober, T., Seeger, B. (eds.) SUM 2012. LNCS, vol. 7520, pp. 85–98. Springer, Heidelberg (2012)
18. Egly, U., Gaggl, S., Woltran, S.: Answer-set programming encodings for argumentation frameworks. In Argument and Computation 1(2), 147–177 (2010)
19. Modgil, S., Caminada, M.: Proof theories and algorithms for abstract argumentation frameworks. In: Rahwan, I., Simari, G. (eds.) Argumentation in Artificial Intelligence, pp. 105–132. Springer (2009)
20. Oikarinen, E., Woltran, S.: Characterizing strong equivalence for argumentation frameworks. Artif. Intell. 175(14-15), 1985–2009 (2011)