



# On the Input/Output behavior of argumentation frameworks



Pietro Baroni<sup>a</sup>, Guido Boella<sup>b</sup>, Federico Cerutti<sup>c</sup>, Massimiliano Giacomin<sup>a,\*</sup>,  
Leendert van der Torre<sup>d</sup>, Serena Villata<sup>e</sup>

<sup>a</sup> Dipartimento di Ingegneria dell'Informazione, University of Brescia, via Branze, 38, 25123, Brescia, Italy

<sup>b</sup> Dipartimento di Informatica, University of Torino, Italy

<sup>c</sup> Department of Computing Science, University of Aberdeen, UK

<sup>d</sup> University of Luxembourg, Luxembourg

<sup>e</sup> INRIA Sophia Antipolis - Mediterranée, France

## ARTICLE INFO

### Article history:

Received 21 March 2014

Received in revised form 22 August 2014

Accepted 25 August 2014

Available online 28 August 2014

### Keywords:

Argumentation frameworks

Argumentation semantics

Modularity

Decomposability

Equivalence

## ABSTRACT

This paper tackles the fundamental questions arising when looking at argumentation frameworks as interacting components, characterized by an Input/Output behavior, rather than as isolated monolithic entities. This modeling stance arises naturally in some application contexts, like multi-agent systems, but, more importantly, has a crucial impact on several general application-independent issues, like argumentation dynamics, argument summarization and explanation, incremental computation, and inter-formalism translation. Pursuing this research direction, the paper introduces a general modeling approach and provides a comprehensive set of theoretical results putting the intuitive notion of Input/Output behavior of argumentation frameworks on a solid formal ground. This is achieved by combining three main ingredients. First, several novel notions are introduced at the representation level, notably those of *argumentation framework with input*, of *argumentation multipole*, and of *replacement* of multipoles within a traditional argumentation framework. Second, several relevant features of argumentation semantics are identified and formally characterized. In particular, the *canonical local function* provides an input-aware semantics characterization and a suite of *decomposability* properties are introduced, concerning the correspondences between semantics outcomes at global and local level. The third ingredient glues the former ones, as it consists of the investigation of some semantics-dependent properties of the newly introduced entities, namely *S-equivalence* of multipoles, *S-legitimacy* and *S-safeness* of replacements, and *transparency* of a semantics with respect to replacements. Altogether they provide the basis and draw the limits of sound interchangeability of multipoles within traditional frameworks. The paper develops an extensive analysis of all the concepts listed above, covering seven well-known literature semantics and taking into account various, more or less constrained, ways of partitioning an argumentation framework. Diverse examples, taken from the literature, are used to illustrate the application of the results obtained and, finally, an extensive discussion of the related literature is provided.

© 2014 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail addresses: [pietro.baroni@ing.unibs.it](mailto:pietro.baroni@ing.unibs.it) (P. Baroni), [guido@di.unito.it](mailto:guido@di.unito.it) (G. Boella), [f.cerutti@abdn.ac.uk](mailto:f.cerutti@abdn.ac.uk) (F. Cerutti), [massimiliano.giacomin@ing.unibs.it](mailto:massimiliano.giacomin@ing.unibs.it) (M. Giacomin), [leon.vandertorre@uni.lu](mailto:leon.vandertorre@uni.lu) (L. van der Torre), [serena.villata@inria.fr](mailto:serena.villata@inria.fr) (S. Villata).

<http://dx.doi.org/10.1016/j.artint.2014.08.004>

0004-3702/© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

This paper deals with modularity in abstract argumentation. The “Merriam-Webster Learner’s Dictionary” defines *modular* as “having parts that can be connected or combined in different ways” while the “Free Dictionary online” remarks that modularity is intended “for easy assembly and repair or flexible arrangement and use”. As such, modularity is a highly desirable property, often enforced by design, in any kind of either material (like the popular Lego toys) or immaterial (like programs developed according to the object-oriented paradigm) artifacts, including knowledge representation and reasoning formalisms.

Roughly speaking, modularity involves two main properties, namely *separability* and *interchangeability* of modules. As to the former, it has to be possible to describe and analyze the global behavior of an artifact in terms of the combination of the local behaviors of the modules composing it. Each local behavior can be characterized individually in a way which is independent of the internal details of the other modules (and, in a sense, of the module itself) and captures only the connections and mutual interactions between the module and the other ones. To put it in other words, each module can be described as a black-box whose Input/Output behavior fully determines its role in the global behavior of any artifact it is plugged in. As to the latter, the interest in replacing a module with another one is very common and arises from a large variety of motivations, either at the operational or design level. Interchangeability of two modules requires first of all that they are compatible as far as the connections with the rest of the artifact are concerned, i.e. that the interfaces they expose are such that wherever one of the modules can be “plugged in”, the other can too. Besides this *plug-level* interchangeability, it is of great interest to characterize the *behavior-level* interchangeability of modules, namely to identify the situations where internally different modules can be freely interchanged without affecting the global behavior of the artifact they belong to, since their Input/Output behavior is equivalent in this respect.

While the formalism of abstract argumentation frameworks [25] and the relevant argumentation semantics (see [3] for a survey) do not appear to have been designed with modularity in mind, investigating their relevant properties is an important research topic which, after having been somehow overlooked, is attracting increasing attention in recent years. An argumentation framework is basically a directed graph representing the conflicts between a set of arguments (the nodes of the graph) and an argumentation semantics can be regarded as a method to answer (typically in a non-univocal way, i.e. producing a set of alternative answers) the “justification question”: “Which is the justification status of arguments given the conflict?”

Referring to a representative set of semantics proposed in the literature, (namely admissible, complete, grounded, preferred, stable, semi-stable and ideal semantics) this paper provides a systematic and comprehensive assessment of modularity in abstract argumentation, by identifying and analyzing in this context the formal counterparts of the general notions of separability and interchangeability described above.

Given a partition of an argumentation framework into *partial* (or *local*) interacting subframeworks, analyzing separability consists in addressing the following issues:

- Is it possible to define a local counterpart of the notion of semantics? i.e. Is there a method to produce local answers to the justification question, taking into account the interactions with other subframeworks?
- Can the set of justification answers prescribed by the (global) semantics be obtained by properly combining (in a bottom-up fashion) the sets of local answers produced in the subframeworks by its local counterpart?
- symmetrically, Can the sets of local answers be obtained (in a top-down fashion) as projections onto the subframeworks of the global answers?

As to the first issue, we introduce the notion of *local function* for a subframework<sup>1</sup> and show that under very mild requirements, satisfied by all semantics considered in this paper, it is possible (and easy) to identify the *canonical local function* for a global semantics. As to the second and third issues, we introduce the formal notions of top-down and bottom-up decomposability, which, jointly, correspond to the notion of (full) decomposability of an argumentation semantics.

Strong as it may seem, full decomposability with respect to every arbitrary partition of every argumentation framework is not unattainable. Indeed, we show that it is satisfied by some of the semantics considered in this paper, while some others are able to achieve at least top-down decomposability and the remaining ones lack all decomposability properties.

As arbitrary partitions correspond to a completely free (if not anarchical) notion of modularity, we also consider a “tidier” style of partitioning, involving the graph-theoretical notion of *strongly connected components*. It turns out that, restricting the set of partitions this way, helps some, but not all, semantics to recover full decomposability.

Turning to interchangeability, we deal with both its plug-level and behavior-level aspects. As to the plug-level, borrowing some terminology from circuit theory, we introduce the notion of *argumentation multipole* as a generic replaceable argumentation component, namely a partial framework interacting through an input and output relation with an external set of invariant arguments.

Plug-level compatibility of two multipoles is a very relaxed notion, since it is only required that two multipoles refer to the same set of external arguments. This is motivated by the fact that imposing a tighter correspondence between Input/

<sup>1</sup> Technically, a subframework is captured by the formal notion of *argumentation framework with input* provided in Definition 11.

Output “terminals” of the multipoles would unnecessarily restrict the scope of the subsequent analysis on behavior-level compatibility. In fact, our analysis shows that a sensible notion of behavioral equivalence between multipoles (called *Input/Output equivalence*) can be introduced by requiring that the effect of the multipoles on the external arguments is the same: it may well be the case that multipoles with different “terminals” have the same effect in behavioral terms. Of course, Input/Output equivalence is a semantics-dependent notion since the behavior of a multipole can only be defined by referring to a specific semantics using the notion of local function mentioned above. In particular, it may be the case that two multipoles are equivalent with respect to some semantics and not equivalent with respect to another semantics.

Input/Output equivalence is the basis for the analysis of the operation of replacement within an argumentation framework. Basically, a replacement consists in substituting a part of the framework with a plug-level compatible multipole. While this notion *per se* allows for arbitrary substitutions, one is interested in analyzing those replacements which have a sound basis. In this perspective, building on multipole equivalence, it is possible to identify the semantics-dependent notions of *legitimate* and *contextually legitimate* replacement. Briefly, a replacement is legitimate if it involves “fully” equivalent multipoles, while it is contextually legitimate if it involves multipoles which are equivalent in the context where the replacement takes place, while not necessarily being equivalent in other contexts. Clearly, legitimate replacements are a (typically strict) subset of contextually legitimate replacements.

One might expect that, given a semantics, legitimate (with respect to that semantics) replacements ensure that the invariant part of the framework is unaffected (in a sense, that it does not notice the change). This property is called *semantics transparency*. A stronger expectation (since the requirement on the replacements is weaker) would be that the invariant part of the framework is unaffected for any contextually legitimate replacement: this property is called *strong transparency*.

Natural as it may seem, transparency is not achieved by all semantics and requires a detailed analysis, showing that different levels of transparency are achieved by the semantics considered in this paper, also taking into account different restrictions on the set of allowed replacements.

These results provide a reference context and fundamental answers to modularity-related issues in abstract argumentation, which, up to now, have been considered in the literature focusing on specific aspects and hence obtaining partial and problem-specific results. Moreover, while being theoretical by nature, the achievements of this paper have several significant application-oriented implications.

On the one hand, semantics decomposability properties provide a sound basis for exploiting various forms of incremental computation which may deliver important efficiency gains in two main respects. First, they enable (and characterize the limits of) the application of divide-and-conquer strategies in the design of algorithms for computational problems in abstract argumentation frameworks. As most of these problems are intractable in the worst case, facing reduced-size sub-problems separately and then combining the partial results in an efficient manner may significantly improve performances on the average. Second, there is a significant application interest in argumentation dynamics, which captures all contexts where a given framework is updated incrementally, as a consequence of the acquisition of new information and/or of the actions of the participants to a multi-agent system. Clearly, if the modification to the initial framework is limited, one is interested to partially reuse the results of previous computations in the new framework rather than redoing all computations from scratch. Again, decomposability properties enable (and characterize the limits of) the use of incremental computation techniques based on the separation between modified and unmodified parts in the updated framework.

On the other hand, the notions and properties concerning multipole equivalence and semantics transparency are applicable in all contexts where there is an interest in replacing a part of a framework with another one. As an example, the activities of summarization and explanation involved in reasoning and communicating at different levels of granularity are, basically, alternative forms of replacement. In the former, a complex part of an argumentation process (e.g. the analysis and discussion of factual evidences in a legal case) is summarized (i.e. replaced) by a more synthetic representation (e.g. focusing on the facts which turn out to have an actual impact on the case decision) which, while leaving out unnecessary details, must ensure that the global outcome is preserved. Dually, explanation can be regarded as the replacement of a synthetic representation with a more detailed/articulated one, again ensuring that this does not induce undesired side-effects outside the replaced part. Further, and more specific of the abstract argumentation field, the basic formalism of argumentation frameworks is often used as a “ground level” representation for other richer and/or more specific formalisms. For instance, formalisms involving the explicit representation of preferences, values, and attacks to attacks can be translated (or flattened) to the basic formalism through suitable procedures. As these procedures typically consist of a set of local replacement rules, multipole equivalence and semantics transparency are very effective tools to analyze their behavior, soundness and applicability under various semantics.

The paper is organized as follows. After recalling the necessary background in Section 2, the general notions concerning semantics decomposability are introduced and discussed in Section 3, while Section 4 provides decomposability results for the seven semantics considered in this paper. Section 5 deals with the key technical notion of effect-dictated semantics and Section 6 then introduces the fundamental concepts concerning interchangeability, namely argumentation multipoles, their Input/Output equivalence, the replacement operator and the properties of semantics transparency. Section 7 analyzes the relationships between decomposability and transparency at a general level, while Section 8 provides transparency results for the seven semantics considered in this paper. Application examples are given in Section 9, Section 10 discusses related works and, finally, Section 11 concludes the paper. All proofs are given in [Appendix A](#).

## 2. Background

We follow the traditional definition of argumentation framework introduced by Dung [25] and define its restriction to a subset of arguments.

**Definition 1.** An *argumentation framework* is a pair  $AF = (Ar, att)$  in which  $Ar$  is a finite set of arguments and  $att \subseteq Ar \times Ar$ . An argument  $A$  such that there is no  $B$  such that  $(B, A) \in att$  is called *initial*. An argument  $B$  such that  $(B, B) \in att$  is called *self-attacking*. Given a set  $Args \subseteq Ar$ , the restriction of  $AF$  to  $Args$ , denoted as  $AF \downarrow_{Args}$  is the argumentation framework  $(Args, att \cap (Args \times Args))$ .

In this paper we use the labelling-based approach to the definition of argumentation semantics. As shown in [21,3], for the semantics considered in this paper there is a direct correspondence with the “traditional” extension-based approach, hence the results presented in this paper are valid in both approaches. The labelling-based definitions have been adopted only because they allow simpler proofs.

A labelling assigns to each argument of an argumentation framework a label taken from a predefined set  $\Lambda$ . For technical reasons, we define labellings both for argumentation frameworks and for arbitrary sets of arguments.

**Definition 2.** Let  $\Lambda$  be a set of labels. Given a set of arguments  $Args$ , a *labelling* of  $Args$  is a total function  $Lab : Args \rightarrow \Lambda$ . The set of all *labellings* of  $Args$  is denoted as  $\mathcal{L}_{Args}$ . Given an argumentation framework  $AF = (Ar, att)$ , a *labelling* of  $AF$  is a labelling of  $Ar$ . The set of all *labellings* of  $AF$  is denoted as  $\mathcal{L}(AF)$ . For a labelling  $Lab$  of  $Args$ , the restriction of  $Lab$  to a set of arguments  $Args' \subseteq Args$ , denoted as  $Lab \downarrow_{Args'}$ , is defined as  $Lab \cap (Args' \times \Lambda)$ .

We adopt the most common choice for  $\Lambda$ , i.e.  $\{\text{in}, \text{out}, \text{undec}\}$ , where the label *in* means that the argument is accepted, the label *out* means that the argument is rejected, and the label *undec* means that the status of the argument is undecided. As explained after Definition 8, an exception is made for stable semantics, which can be more conveniently defined assuming  $\Lambda = \{\text{in}, \text{out}\}$ . Given a labelling  $Lab$ , we write  $\text{in}(Lab)$  for  $\{A \mid Lab(A) = \text{in}\}$ ,  $\text{out}(Lab)$  for  $\{A \mid Lab(A) = \text{out}\}$  and  $\text{undec}(Lab)$  for  $\{A \mid Lab(A) = \text{undec}\}$ .

A labelling-based semantics prescribes a set of labellings for each argumentation framework.

**Definition 3.** Given an argumentation framework  $AF = (Ar, att)$ , a labelling-based semantics  $\mathbf{S}$  associates with  $AF$  a subset of  $\mathcal{L}(AF)$ , denoted as  $\mathbf{L}_{\mathbf{S}}(AF)$ .

In general, a semantics encompasses a set of alternative labellings for a single argumentation framework. However, a semantics may be defined so that a unique labelling is always prescribed, i.e. for every argumentation framework  $AF$ ,  $|\mathbf{L}_{\mathbf{S}}(AF)| = 1$ . In this case the semantics is said to be *single-status*, while in the general case it is said to be *multiple-status*.

In the labelling-based approach, a semantics definition relies on some *legality* constraints relating the label of an argument to those of its attackers.

**Definition 4.** Let  $Lab$  be a labelling of the argumentation framework  $(Ar, att)$ . An *in*-labelled argument is said to be *legally in* iff all its attackers are labelled *out*. An *out*-labelled argument is said to be *legally out* iff it has at least one attacker that is labelled *in*. An *undec*-labelled argument is said to be *legally undec* iff not all its attackers are labelled *out* and it does not have an attacker that is labelled *in*.

We now introduce the definitions of labellings corresponding to traditional admissible<sup>2</sup> and complete semantics.

**Definition 5.** Let  $AF = (Ar, att)$  be an argumentation framework. An *admissible labelling* is a labelling  $Lab$  where every *in*-labelled argument is legally *in* and every *out*-labelled argument is legally *out*.

**Definition 6.** A *complete labelling* is a labelling where every *in*-labelled argument is legally *in*, every *out*-labelled argument is legally *out* and every *undec*-labelled argument is legally *undec*.

On this basis, the labelling-based definitions of several argumentation semantics can be introduced. To simplify the technical treatment in the following, grounded and preferred semantics are defined by referring to the commitment relation between labellings [3].

<sup>2</sup> It can be remarked that (unlike the other semantics) admissible labellings are not in a one-to-one correspondence to admissible sets since several admissible labellings might correspond to the same admissible set.

**Definition 7.** Let  $Lab_1$  and  $Lab_2$  be two labellings. We say that  $Lab_2$  is *more or equally committed* than  $Lab_1$  ( $Lab_1 \sqsubseteq Lab_2$ ) iff  $\text{in}(Lab_1) \subseteq \text{in}(Lab_2)$  and  $\text{out}(Lab_1) \subseteq \text{out}(Lab_2)$ .

**Definition 8.** Let  $AF = (Ar, att)$  be an argumentation framework. A *stable labelling* of  $AF$  is a complete labelling without  $\text{undec}$ -labelled arguments. The *grounded labelling* of  $AF$  is the minimal (w.r.t.  $\sqsubseteq$ ) labelling among all complete labellings. A *preferred labelling* of  $AF$  is a maximal (w.r.t.  $\sqsubseteq$ ) labelling among all complete labellings. The *ideal labelling* of  $AF$  is the maximal (under  $\sqsubseteq$ ) complete<sup>3</sup> labelling  $Lab$  that is less or equally committed than each preferred labelling of  $AF$  (i.e. for each preferred labelling  $Lab_p$  it holds that  $Lab \sqsubseteq Lab_p$ ). A *semi-stable labelling* of  $AF$  is a complete labelling  $Lab$  where  $\text{undec}(Lab)$  is minimal (w.r.t. set inclusion) among all complete labellings.

While stable semantics is defined by assuming  $\Lambda = \{\text{in}, \text{out}, \text{undec}\}$ , the definition of stable labelling entails that stable semantics can be equivalently defined with reference to the set of labels  $\Lambda = \{\text{in}, \text{out}\}$ . In this case, a stable labelling is simply a complete labelling, since the codomain  $\Lambda$  does not include  $\text{undec}$ . In the sequel we implicitly assume that, for stable semantics only,  $\Lambda = \{\text{in}, \text{out}\}$ : this allows a simpler treatment of such semantics without any loss of generality.

The uniqueness of the grounded and the ideal labelling has been proved in [22]. Accordingly, grounded and ideal semantics are single-status, the other semantics are multiple-status. Admissible, complete, stable, grounded, preferred, ideal and semi-stable semantics are denoted in the following as **AD**, **CO**, **ST**, **GR**, **PR**, **ID** and **SST**, respectively.

We also recall the traditional notions of skeptical and credulous justification of an argument with respect to a semantics.

**Definition 9.** Given a labelling-based semantics **S** and an argumentation framework  $AF$ , an argument  $A$  is *skeptically justified* under **S** if  $\forall Lab \in \mathbf{L_S}(AF) \text{ } Lab(A) = \text{in}$ ; an argument  $A$  is *credulously justified* under **S** if  $\exists Lab \in \mathbf{L_S}(AF) : Lab(A) = \text{in}$ .

Finally, a comment is in order on a special case of argumentation framework that is explicitly considered in the paper, i.e. the empty argumentation framework  $AF_\emptyset \triangleq (\emptyset, \emptyset)$ . By definition the only possible labelling of  $AF_\emptyset$  is the empty set, thus a semantics can either prescribe  $\emptyset$  for  $AF_\emptyset$  or it can prescribe no labelling at all. In this respect, for any semantics **S** introduced above it holds  $\mathbf{L_S}(AF_\emptyset) = \{\emptyset\}$ , i.e. the empty set is actually prescribed by **S**. Note in particular that  $\emptyset$  is a stable labelling, since it is complete and does not include  $\text{undec}$ -labelled arguments.

### 3. Decomposability of argumentation semantics

#### 3.1. The notion of local function

The first step to define the notion of semantics decomposability is to introduce a formal setting to express the interactions between the partial frameworks induced by an arbitrary partitioning of an argumentation framework. Intuitively, given an argumentation framework  $AF = (Ar, att)$  and a subset  $Args$  of its arguments, the elements affecting  $AF \downarrow_{Args}$  include the arguments attacking  $Args$  from the outside, called *input arguments*, and the attack relation from the input arguments to  $Args$ , called *conditioning relation*.

**Definition 10.** Given  $AF = (Ar, att)$  and a set  $Args \subseteq Ar$ , the *input* of  $Args$ , denoted as  $Args^{\text{inp}}$ , is the set  $\{B \in Ar \setminus Args \mid \exists A \in Args, (B, A) \in att\}$ , the *conditioning relation* of  $Args$ , denoted as  $Args^R$ , is defined as  $att \cap (Args^{\text{inp}} \times Args)$ .

**Example 1.** Consider  $AF = (\{A, B, C, D\}, \{(A, B), (B, C), (C, A), (A, D), (D, A)\})$  with reference to the partial frameworks induced by the sets  $\{A, B, C\}$  and  $\{D\}$  (see Fig. 1). It holds that  $\{A, B, C\}^{\text{inp}} = \{D\}$  and  $\{A, B, C\}^R = \{(D, A)\}$ , while  $\{D\}^{\text{inp}} = \{A\}$  and  $\{D\}^R = \{(A, D)\}$ .

Given a partial argumentation framework  $AF \downarrow_{Args}$  (possibly  $AF$  itself) affected by a (possibly empty) set of arguments  $Args^{\text{inp}}$  attacking  $Args$  according to  $Args^R$ , one may wonder whether fixing the labelling assigned to the input arguments allows one to determine the set of labellings of  $AF \downarrow_{Args}$ . As shown in the following, this question cannot be answered once and for all, since different semantics exhibit different behaviors in this respect, and, for some semantics, a dependency holds under specific constraints on the considered partition of the argumentation framework. In order to express such a dependency (whenever it holds), we introduce the notions of *argumentation framework with input*, consisting of an argumentation framework  $AF = (Ar, att)$  (playing the role of a partial argumentation framework), a set of external input arguments  $\mathcal{I}$ , a labelling  $L_{\mathcal{I}}$  assigned to them and an attack relation  $R_{\mathcal{I}}$  from  $\mathcal{I}$  to  $Ar$ , and of a *local function* which, given an argumentation framework with input, returns a corresponding set of labellings of  $AF$ .

<sup>3</sup> Literally, the original definition refers to an admissible labelling rather than a complete labelling. However, the definition adopted here is equivalent to the original one, since it can be shown that the ideal labelling is a complete labelling [22].

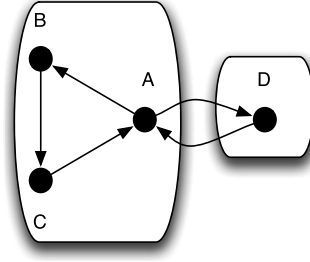


Fig. 1. Running example: a partition of a simple framework (Examples 1–5).

**Definition 11.** An *argumentation framework with input* is a tuple  $(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}})$ , including an argumentation framework<sup>4</sup>  $AF = (Ar, att)$ , a set of arguments  $\mathcal{I}$  such that  $\mathcal{I} \cap Ar = \emptyset$ , a labelling  $L_{\mathcal{I}} \in \mathcal{L}_{\mathcal{I}}$  and a relation  $R_{\mathcal{I}} \subseteq \mathcal{I} \times Ar$ . A *local function* assigns to any argumentation framework with input a (possibly empty) set of labellings of  $AF$ , i.e.  $F(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}}) \in 2^{\mathcal{L}(AF)}$ .

For any semantics, a “sensible” local function, called *canonical local function*, is the one that describes the labellings of the so-called standard argumentation frameworks.

**Definition 12.** Given an argumentation framework with input  $(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}})$ , the standard argumentation framework w.r.t.  $(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}})$  is defined as  $AF' = (Ar \cup \mathcal{I}', att \cup R'_{\mathcal{I}})$ , where  $\mathcal{I}' = \mathcal{I} \cup \{A' \mid A \in out(L_{\mathcal{I}})\}$  and  $R'_{\mathcal{I}} = R_{\mathcal{I}} \cup \{(A', A) \mid A \in out(L_{\mathcal{I}})\} \cup \{(A, A) \mid A \in undec(L_{\mathcal{I}})\}$ .

Roughly, the standard argumentation framework puts  $AF$  under the influence of  $(\mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}})$ , by adding  $\mathcal{I}$  to  $Ar$  and  $R_{\mathcal{I}}$  to  $att$ , and by enforcing<sup>5</sup> the label  $L_{\mathcal{I}}$  for the arguments of  $\mathcal{I}$  in this way:

- for each argument  $A \in \mathcal{I}$  such that  $L_{\mathcal{I}}(A) = out$ , an unattacked argument  $A'$  is included which attacks  $A$ , in order to get  $A$  labelled *out* by all labellings of  $AF'$ ;
- for each argument  $A \in \mathcal{I}$  such that  $L_{\mathcal{I}}(A) = undec$ , a self-attack is added to  $A$  in order to get it labelled *undec* by all labellings of  $AF'$ ;
- each argument  $A \in \mathcal{I}$  such that  $L_{\mathcal{I}}(A) = in$  is left unattacked, so that it is labelled *in* by all labellings of  $AF'$ .

**Definition 13.** Given a semantics  $S$ , the *canonical local function* of  $S$  (also called *local function* of  $S$ ) is defined as  $F_S(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}}) = \{Lab \downarrow_{Ar} \mid Lab \in \mathbf{L}_S(AF')\}$ , where  $AF = (Ar, att)$  and  $AF'$  is the standard argumentation framework w.r.t.  $(AF, \mathcal{I}, L_{\mathcal{I}}, R_{\mathcal{I}})$ .

Note that in the case of stable semantics  $undec \notin \Lambda$ , thus  $R'_{\mathcal{I}}$  does not include self-attacks.

In case  $\mathcal{I} = \emptyset$  (entailing  $L_{\mathcal{I}} = \emptyset$  and  $R_{\mathcal{I}} = \emptyset$ ) the canonical local function returns the labellings of  $AF$ , as shown by Proposition 1.

**Proposition 1.** Given a semantics  $S$  and an argumentation framework  $AF$ ,  $F_S(AF, \emptyset, \emptyset, \emptyset) = \mathbf{L}_S(AF)$ .

While the canonical local function is defined for any semantics, its definition is best suited for *complete-compatible* semantics, i.e. semantics satisfying a number of intuitive constraints.

**Definition 14.** A semantics  $S$  is *complete-compatible* iff the following conditions hold:

1. For any argumentation framework  $AF = (Ar, att)$ , every labelling  $\mathbf{L} \in \mathbf{L}_S(AF)$  satisfies the following conditions:
  - if  $A \in Ar$  is initial, then  $\mathbf{L}(A) = in$
  - if  $B \in Ar$  and there is an initial argument  $A$  which attacks  $B$ , then  $\mathbf{L}(B) = out$
  - if  $C \in Ar$  is self-attacking, and there are no attackers of  $C$  besides  $C$  itself, then  $\mathbf{L}(C) = undec$
2. for any set of arguments  $\mathcal{I}$  and any labelling  $L_{\mathcal{I}} \in \mathcal{L}_{\mathcal{I}}$ , the argumentation framework  $AF' = (\mathcal{I}', att')$ , where  $\mathcal{I}' = \mathcal{I} \cup \{A' \mid A \in out(L_{\mathcal{I}})\}$  and  $att' = \{(A', A) \mid A \in out(L_{\mathcal{I}})\} \cup \{(A, A) \mid A \in undec(L_{\mathcal{I}})\}$ , admits a (unique) labelling, i.e.  $|\mathbf{L}_S(AF')| = 1$ .

<sup>4</sup> In the following, unless otherwise specified, we will implicitly assume  $AF = (Ar, att)$ .

<sup>5</sup> Actually, the enforcement is a bit different for admissible semantics. This exception has no consequences on the technical development of the paper.



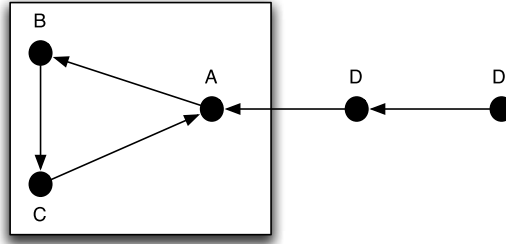


Fig. 2. The standard argumentation framework w.r.t.  $(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{out})\}, \{(D, A)\})$  (Example 2).

It should be noted that, in case  $\text{undec} \notin A$ , the third bullet of condition 1 entails that there is no labelling if a self-attacking argument  $C$  is attacked by  $C$  only, and in condition 2 it necessarily holds that  $\text{undec}(L_{\mathcal{S}}) = \emptyset$ .

As shown by Proposition 2, the requirements of the previous definition guarantee that the construction of the standard argumentation framework makes sense, i.e. given a standard argumentation framework w.r.t.  $(AF, \mathcal{S}, L_{\mathcal{S}}, R_{\mathcal{S}})$ , a complete-compatible semantics enforces the labelling  $L_{\mathcal{S}}$  for the arguments of  $\mathcal{S}$  as described above.

**Proposition 2.** Let  $\mathbf{S}$  be a complete-compatible semantics and let  $AF' = (Ar \cup \mathcal{S}', att \cup R'_{\mathcal{S}})$  be the standard argumentation framework w.r.t. an argumentation framework with input  $(AF, \mathcal{S}, L_{\mathcal{S}}, R_{\mathcal{S}})$ . Then for any  $Lab \in \mathbf{L}_{\mathbf{S}}(AF')$  it holds that  $Lab \downarrow_{\mathcal{S}'} = \{(A', \text{in}) \mid A \in \text{out}(L_{\mathcal{S}})\} \cup L_{\mathcal{S}}$  and  $Lab \downarrow_{\mathcal{S}} = L_{\mathcal{S}}$ .

Moreover, when applied to the empty argumentation framework (which by definition does not receive attacks from  $\mathcal{S}$ ) the canonical local function of a complete-compatible semantics always returns the empty set as a unique labelling.

**Proposition 3.** Given a complete-compatible semantics  $\mathbf{S}$ , a set of arguments  $\mathcal{S}$  and a labelling  $L_{\mathcal{S}} \in \mathbf{L}_{\mathcal{S}}$ , it holds that  $F_{\mathbf{S}}(AF_{\emptyset}, \mathcal{S}, L_{\mathcal{S}}, \emptyset) = \{\emptyset\}$ .

Taking into account Proposition 1 this result entails that  $\mathbf{L}_{\mathbf{S}}(AF_{\emptyset}) = \{\emptyset\}$ , corresponding to the second requirement of Definition 14 with  $\mathcal{S} = \emptyset$ .

All the semantics considered in the paper are complete-compatible, with the exception of admissible semantics.

**Proposition 4.** GR, CO, ST, PR, SST, ID are all complete-compatible semantics.

Admissible semantics is not complete-compatible, as it can be seen by considering e.g. the argumentation framework  $AF = (\{A\}, \emptyset)$ , where  $\mathbf{L}_{\text{AD}}(AF) = \{(A, \text{undec}), (A, \text{in})\}$ .

The following example clarifies the notion of canonical local function, considering in particular complete semantics.

**Example 2.** Let us refer again to the argumentation framework  $AF$  of Fig. 1. For the canonical local function of complete semantics it holds that  $F_{\text{CO}}(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{out})\}, \{(D, A)\}) = \{(A, \text{undec}), (B, \text{undec}), (C, \text{undec})\}$ , due to the fact that the standard argumentation framework w.r.t.  $(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{out})\}, \{(D, A)\})$ , shown in Fig. 2, admits as the unique complete labelling  $\{(D', \text{in}), (D, \text{out}), (A, \text{undec}), (B, \text{undec}), (C, \text{undec})\}$ . In a similar way, it is easy to show that  $F_{\text{CO}}(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{in})\}, \{(D, A)\}) = \{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$  and  $F_{\text{CO}}(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{undec})\}, \{(D, A)\}) = \{(A, \text{undec}), (B, \text{undec}), (C, \text{undec})\}$ . Considering the application of  $F_{\text{CO}}$  to  $AF \downarrow_{\{D\}}$ ,  $F_{\text{CO}}(AF \downarrow_{\{D\}}, \{A\}, \{(A, \text{out})\}, \{(A, D)\}) = \{(D, \text{in})\}$ ,  $F_{\text{CO}}(AF \downarrow_{\{D\}}, \{A\}, \{(A, \text{in})\}, \{(A, D)\}) = \{(D, \text{out})\}$  and  $F_{\text{CO}}(AF \downarrow_{\{D\}}, \{A\}, \{(A, \text{undec})\}, \{(A, D)\}) = \{(D, \text{undec})\}$ .

As shown in Section 4, for any semantics considered in this paper the local function admits a compact representation, without the need to refer to standard argumentation frameworks.

### 3.2. Decomposability properties of argumentation semantics

We now aim at introducing a formal notion of semantics decomposability. To this purpose, consider a generic argumentation framework  $AF = (Ar, att)$  and an arbitrary partition of  $Ar$ , i.e. a set  $\{P_1, \dots, P_n\}$  such that  $\forall i \in \{1, \dots, n\} P_i \subseteq Ar$  and  $P_i \neq \emptyset$ ,  $\bigcup_{i=1}^n P_i = Ar$  and  $P_i \cap P_j = \emptyset$  for  $i \neq j$ . Such a partition identifies the restricted argumentation frameworks  $AF \downarrow_{P_1}, \dots, AF \downarrow_{P_n}$ , that affect each other with the relevant input arguments and conditioning relations as stated in Definition 10. Intuitively a semantics  $\mathbf{S}$  is decomposable if  $\mathbf{S}$  can be put in correspondence with a local function  $F$  such that:

- every labelling prescribed by  $\mathbf{S}$  on  $AF$ , namely every element of  $\mathbf{L}_{\mathbf{S}}(AF)$ , corresponds to the union of  $n$  “compatible” labellings  $L_{P_1}, \dots, L_{P_n}$  of the restricted argumentation frameworks, all of them obtained applying  $F$ ;

- in turn, each union of  $n$  “compatible” labellings  $L_{P_1}, \dots, L_{P_n}$  obtained applying  $F$  to the restricted frameworks gives rise to a labelling of  $AF$ .

The “compatibility” constraint mentioned above reflects the fact that any labelling of a restricted framework is used by  $F$  for computing the other ones:  $L_{P_i}$  plays a role in determining  $L_{P_1}, \dots, L_{P_{i-1}}, L_{P_{i+1}}, \dots, L_{P_n}$  and vice versa. This means that  $L_{P_1}, \dots, L_{P_n}$  are “compatible” if each  $L_{P_i}$  is produced by  $F$  for  $AF \downarrow_{P_i}$  with the input arguments  $P_i^{\text{inp}}$  labelled according to  $L_{P_1}, \dots, L_{P_{i-1}}, L_{P_{i+1}}, \dots, L_{P_n}$ . **Definition 15** synthesizes all these considerations.

**Definition 15.** A semantics  $\mathbf{S}$  is *fully decomposable* (or simply *decomposable*) iff there is a local function  $F$  such that for every argumentation framework  $AF = (Ar, att)$  and every partition  $\mathcal{P} = \{P_1, \dots, P_n\}$  of  $Ar$ ,  $\mathbf{L}_{\mathbf{S}}(AF) = \mathcal{U}(\mathcal{P}, AF, F)$  where  $\mathcal{U}(\mathcal{P}, AF, F) \triangleq \{L_{P_1} \cup \dots \cup L_{P_n} \mid L_{P_i} \in F(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}, P_i^{\text{R}}})\}$ .

**Example 3.** Considering again the argumentation framework  $AF$  of Fig. 1 and the partition  $\{\{A, B, C\}, \{D\}\}$ , full decomposability of complete semantics requires a local function such that the labellings of  $AF$  are exactly those obtained by the union of the compatible labellings of  $AF \downarrow_{\{A, B, C\}}$  and  $AF \downarrow_{\{D\}}$  given by the local function itself. Let us consider the canonical local function<sup>6</sup> of **CO** (refer to **Example 2**). The labelling  $\{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$  is compatible with  $\{(D, \text{in})\}$ , since the first is obtained by  $F_{\text{CO}}$  with  $D$  labelled  $\text{in}$ , and the latter is obtained by  $F_{\text{CO}}$  with  $A$  labelled  $\text{out}$ . On the other hand, the labelling  $\{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$  is not compatible e.g. with  $\{(D, \text{out})\}$ . Overall, exactly two global labellings arise from the combinations of the compatible outcomes of  $F_{\text{CO}}$ , namely  $\{(A, \text{undec}), (B, \text{undec}), (C, \text{undec}), (D, \text{undec})\}$  and  $\{(A, \text{out}), (B, \text{in}), (C, \text{out}), (D, \text{in})\}$ , corresponding to the complete labellings of  $AF$ .

The behavior of complete semantics in this example is not incidental: we will prove in Section 4 that complete semantics is fully decomposable.

**Proposition 5** shows that, if a complete-compatible semantics  $\mathbf{S}$  is fully decomposable, then the local function appearing in **Definition 15** coincides with the canonical local function  $F_{\mathbf{S}}$ .

**Proposition 5.** Given a complete-compatible semantics  $\mathbf{S}$ , if  $\mathbf{S}$  is fully decomposable then there is a unique local function satisfying the conditions of **Definition 15**, coinciding with the canonical local function  $F_{\mathbf{S}}$ .

Full decomposability can be viewed as the conjunction of two partial decomposability properties, namely *top-down decomposability* and *bottom-up decomposability*.

In words, a semantics is top-down decomposable if the procedure to compute the global labellings identified by **Definition 15** is complete, i.e. all of the global labellings can be obtained by combining the labellings prescribed by  $F_{\mathbf{S}}$  for the restricted subframeworks, even if putting together labellings of the restricted subframeworks may give rise to some “spurious” labellings besides the correct ones. The following definition formalizes this intuition.

**Definition 16.** A complete-compatible semantics  $\mathbf{S}$  is *top-down decomposable* iff for any argumentation framework  $AF = (Ar, att)$  and any partition  $\mathcal{P} = \{P_1, \dots, P_n\}$  of  $Ar$ , it holds that  $\mathbf{L}_{\mathbf{S}}(AF) \subseteq \mathcal{U}(\mathcal{P}, AF, F_{\mathbf{S}})$ .

While top-down decomposability corresponds to completeness of the procedure identified by **Definition 15**, bottom-up decomposability requires its soundness, i.e. that any combination of local labellings is a global labelling, while it is not guaranteed that all global labellings can be obtained in this way.

**Definition 17.** A complete-compatible semantics  $\mathbf{S}$  is *bottom-up decomposable* iff for any argumentation framework  $AF = (Ar, att)$  and any partition  $\mathcal{P} = \{P_1, \dots, P_n\}$  of  $Ar$ , it holds that  $\mathbf{L}_{\mathbf{S}}(AF) \supseteq \mathcal{U}(\mathcal{P}, AF, F_{\mathbf{S}})$ .

A comment on the two definitions above is in order. While the definition of full decomposability applies to any kind of semantics and requires the existence of a local function satisfying the decomposability property, **Definitions 16 and 17** are restricted to complete-compatible semantics and refer to the canonical local function  $F_{\mathbf{S}}$  to avoid triviality: the local function returning all the possible labellings of  $AF$  trivially satisfies the inclusion condition of **Definition 16** for any semantics, while the local function always returning the empty set trivially satisfies the condition of **Definition 17**. This is the reason why both definitions refer to the specific canonical local function, which makes sense for complete-compatible semantics in the light of **Proposition 5**. If a semantics is not complete-compatible<sup>7</sup> then the notion of canonical local function is meaningless, since the labelling  $L_{\mathcal{J}}$  would not be in general enforced for the arguments of  $\mathcal{J}$  in the standard argumentation framework w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  (see **Proposition 2**).

<sup>6</sup> It is shown in **Proposition 5** that considering the canonical local function is without loss of generality.

<sup>7</sup> Besides admissible semantics, in the literature there are a few examples of non-complete-compatible semantics, like stage semantics [42] and various forms of prudent semantics [24].



As shown in Section 4, some semantics that do not satisfy full decomposability are still able to satisfy top-down decomposability. Moreover, there are semantics that do not satisfy either of them: in this case it is interesting to investigate whether decomposability holds by restricting the possible partitions of the argumentation frameworks to those satisfying a given set of constraints. To express this restriction, we first introduce the notion of partition selector.

**Definition 18.** A *partition selector*  $\mathcal{F}$  is a function receiving as input an argumentation framework  $AF = (Ar, att)$  and returning a set of partitions of  $Ar$ .

A partition selector is defined as a function of argumentation frameworks, since different argumentation frameworks with the same set of arguments may allow different sets of partitions, depending on the attack relation.

The decomposability notions introduced so far can then be extended to take into account a specific restriction on the considered partitions.

**Definition 19.** Let  $\mathcal{F}$  be a partition selector. A complete-compatible semantics  $S$  is *top-down decomposable* w.r.t.  $\mathcal{F}$  iff for any argumentation framework  $AF$  and any partition  $\mathcal{P} = \{P_1, \dots, P_n\} \in \mathcal{F}(AF)$ , it holds that  $L_S(AF) \subseteq \mathcal{U}(\mathcal{P}, AF, F_S)$ . A complete-compatible semantics  $S$  is *bottom-up decomposable* w.r.t.  $\mathcal{F}$  iff for any argumentation framework  $AF$  and any partition  $\{P_1, \dots, P_n\} \in \mathcal{F}(AF)$ ,  $L_S(AF) \supseteq \mathcal{U}(\mathcal{P}, AF, F_S)$ . A complete-compatible semantics is *fully decomposable* (or simply *decomposable*) w.r.t. a partition selector  $\mathcal{F}$  iff it is both top-down and bottom-up decomposable w.r.t.  $\mathcal{F}$ .

Of course, full decomposability, top-down decomposability and bottom-up decomposability as introduced in Definitions 15, 16 and 17, respectively, are equivalent to the corresponding decomposability properties w.r.t.  $\mathcal{F}_{ALL}$ , i.e. the selector returning all possible partitions.

**Definition 20.** For any argumentation framework  $AF = (Ar, att)$ ,  $\mathcal{F}_{ALL}(AF) \triangleq \{\{P_1, \dots, P_n\} \mid \{P_1, \dots, P_n\} \text{ is a partition of } Ar\}$ .

Apart from this limit case, a particular partition selector that has received attention in the literature and will be considered in this paper is the one based on the notion of strongly connected component (SCC) of an argumentation framework. Its importance is due to the fact that most argumentation semantics in the literature are *SCC-recursive* [10], which, briefly, means that the semantics can be defined in terms of a *base function* operating at the level of single strongly connected components. Roughly, this also implies that an incremental computation procedure based on the decomposition of the framework into its strongly connected components can be defined, a property exploited in several subsequent works [33,40,23]. Here we introduce the necessary basic definitions, leaving further discussion on this subject to Section 10.

**Definition 21.** Given an argumentation framework  $AF = (Ar, att)$ , the set of strongly connected components of  $AF$ , denoted as  $SCCS_{AF}$ , consists of the equivalence classes of arguments induced by the binary relation of path-equivalence, i.e. the relation  $\rho(A, B)$  defined over  $Ar \times Ar$  such that  $\rho(A, B)$  holds if and only if  $A = B$  or there are directed paths from  $A$  to  $B$  and from  $B$  to  $A$  in  $AF$ .

For instance, the argumentation framework of Fig. 1 has a unique strongly connected component including all of the arguments, while for the argumentation framework  $AF$  of Fig. 2 it holds that  $SCCS_{AF} = \{\{D'\}, \{D\}, \{A, B, C\}\}$ .

At least two partition selectors based on strongly connected components can be considered. The simplest selector, denoted as  $\mathcal{F}_{SCC}$ , includes for each argumentation framework  $AF$  the unique partition consisting of the strongly connected components  $SCCS_{AF}$ . A second selector, denoted as  $\mathcal{F}_{USCC}$ , includes all the partitions such that every element is the union of some (possibly unconnected) strongly connected components.

**Definition 22.** For any argumentation framework  $AF = (Ar, att)$ ,  $\mathcal{F}_{SCC}(AF) \triangleq \{SCCS_{AF}\} \setminus \{\emptyset\}$ ,  $\mathcal{F}_{USCC}(AF) \triangleq \{\{P_1, \dots, P_n\} \mid \{P_1, \dots, P_n\} \text{ is a partition of } Ar \text{ and } \forall i ((S \in SCCS_{AF} \wedge P_i \cap S \neq \emptyset) \rightarrow S \subseteq P_i)\}$ .

It is immediate to see that, for any  $AF$ ,  $\mathcal{F}_{SCC}(AF) \subseteq \mathcal{F}_{USCC}(AF)$ . As to the first part of the definition, note that the set  $SCCS_{AF}$  includes  $\emptyset$  only in case  $AF = AF_\emptyset$ , which does not admit any partition (since all the elements of a partition must be nonempty), thus  $\mathcal{F}_{SCC}(AF_\emptyset) = \emptyset$ .

#### 4. Analyzing semantics decomposability

In this section we discuss the decomposability properties of the semantics reviewed in Section 2. A synthetic view of the results is given in Table 1 (note that for all semantics full, top-down and bottom-up decomposability w.r.t.  $\mathcal{F}_{USCC}$  turn out to be satisfied if and only if full, top-down and bottom-up decomposability w.r.t.  $\mathcal{F}_{SCC}$  are satisfied, respectively). Since admissible semantics is not complete-compatible, only the notion of full decomposability is applicable to it.

**Table 1**

Decomposability properties of argumentation semantics.

	AD	CO	ST	GR	PR	ID	SST
Full decomposability (Definition 15)	Yes	Yes	Yes	No	No	No	No
Top-down decomposability (Definition 16)	–	Yes	Yes	Yes	Yes	No	No
Bottom-up decomposability (Definition 17)	–	Yes	Yes	No	No	No	No
Full decomposability w.r.t. $\mathcal{F}_{USCC}$ and $\mathcal{F}_{SCC}$ (Definition 19)	Yes	Yes	Yes	Yes	Yes	No	No
Top-down decomposability w.r.t. $\mathcal{F}_{USCC}$ and $\mathcal{F}_{SCC}$ (Definition 19)	–	Yes	Yes	Yes	Yes	No	No
Bottom-up decomposability w.r.t. $\mathcal{F}_{USCC}$ and $\mathcal{F}_{SCC}$ (Definition 19)	–	Yes	Yes	Yes	Yes	No	No

#### 4.1. Admissible and complete semantics

We first analyze admissible and complete semantics, since they are the basis for the other ones considered in this paper: according to Definition 8, stable, grounded, preferred, ideal, and semi-stable semantics select labellings among the complete ones, which are admissible by definition. Given this, it would be very unpleasant if complete (and thus admissible) semantics would not be decomposable. As shown by Theorems 1 and 3, luckily both admissible and complete semantics turn out to be fully decomposable.

The following definition introduces the canonical local function of admissible semantics, by extending the definition of admissible labelling in order to account for “external” input arguments in the obvious way. The proof that the definition is correct is provided by Theorem 2.

**Definition 23.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ ,  $F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{Lab \in \mathcal{L}(AF) \mid Lab(A) = in \rightarrow ((\forall B \in Ar : (B, A) \in att, Lab(B) = out) \wedge (\forall B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}(B) = out)), Lab(A) = out \rightarrow ((\exists B \in Ar : (B, A) \in att \wedge Lab(B) = in) \vee (\exists B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}(B) = in))\}$ .

Theorem 1 proves that admissible semantics is fully decomposable, showing that the local function  $F_{AD}$  introduced in Definition 23 satisfies the conditions of Definition 15.

**Theorem 1.** Admissible semantics **AD** is fully decomposable, with  $F_{AD}$  satisfying the conditions of Definition 15.

The following theorem confirms that Definition 23 actually corresponds to the canonical local function of admissible semantics.

**Theorem 2.** The canonical local function of admissible semantics is  $F_{AD}$ , as defined in Definition 23.

Also the canonical local function of complete semantics can be guessed on the basis of the definition of complete labelling.

**Definition 24.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ ,  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{Lab \in \mathcal{L}(AF) \mid Lab(A) = in \rightarrow ((\forall B \in Ar : (B, A) \in att, Lab(B) = out) \wedge (\forall B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}(B) = out)), Lab(A) = out \rightarrow ((\exists B \in Ar : (B, A) \in att \wedge Lab(B) = in) \vee (\exists B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}(B) = in)), Lab(A) = undec \rightarrow (((\forall B \in Ar : (B, A) \in att, Lab(B) \neq in) \wedge (\forall B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}(B) \neq in)) \wedge ((\exists B \in Ar : (B, A) \in att \wedge Lab(B) = undec) \vee (\exists B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}(B) = undec)))\}$ .

It is easy to see that  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \subseteq F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , i.e. every “locally complete” labelling is also “locally admissible”.

Theorem 3 shows that also complete semantics is fully decomposable.<sup>8</sup> Since the proof adopts  $F_{CO}$  as the local function and **CO** is complete-compatible, by Proposition 5 it holds that  $F_{CO}$  is actually the canonical local function of complete semantics.

**Theorem 3.** Complete semantics **CO** is fully decomposable and  $F_{CO}$  is its canonical local function.

#### 4.2. Stable semantics

Stable semantics inherits full decomposability from complete semantics: the reason is that the definition of stable labelling corresponds to that of complete labelling with the additional requirement that no argument is labelled undec, and

<sup>8</sup> Proposition 3 of [40] proves a weaker property of complete semantics, corresponding to bottom-up decomposability in the extension-based approach.

this requirement holds at the level of the whole argumentation framework iff it holds in any of its subframeworks. The relevant local function can easily be identified by taking into account this requirement (again, the fact that such local function is the canonical one holds in virtue of [Proposition 5](#)).

**Definition 25.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ ,  $F_{ST}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{Lab \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mid \forall A \in Ar, Lab(A) \neq \text{undec}\}$ .

**Theorem 4.** Stable semantics **ST** is fully decomposable and  $F_{ST}$  is its canonical local function.

**Example 4.** Consider again the running example of [Fig. 1](#). Taking into account the results provided in [Example 2](#) for the local function of complete semantics, it is easy to see that  $F_{ST}(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{in})\}, \{(D, A)\}) = \emptyset$ , that  $F_{ST}(AF \downarrow_{\{A,B,C\}}, \{D\}, \{(D, \text{in})\}, \{(D, A)\}) = \{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$ , and for  $AF \downarrow_{\{D\}}$  that  $F_{ST}(AF \downarrow_{\{D\}}, \{A\}, \{(A, \text{out})\}, \{(A, D)\}) = \{(D, \text{in})\}$ ,  $F_{ST}(AF \downarrow_{\{D\}}, \{A\}, \{(A, \text{in})\}, \{(A, D)\}) = \{(D, \text{out})\}$ . Accordingly, there is just a pair of compatible local labellings, namely  $\{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$  and  $\{(D, \text{in})\}$ , giving rise to the unique stable labelling  $\{(A, \text{out}), (B, \text{in}), (C, \text{out}), (D, \text{in})\}$ .

#### 4.3. Grounded and preferred semantics

As in the previous cases, the canonical local functions of grounded and preferred semantics can be obtained by extending the definition of grounded and preferred labelling, respectively. [Proposition 6](#) identifies these functions, also showing that the relevant definitions are well-founded, in particular, that there is always a unique minimal labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and that  $F_{PR}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  is nonempty.

**Proposition 6.** The canonical local function of grounded and preferred semantics are defined as

- $F_{GR}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L}^*\}$ , where  $\mathbf{L}^*$  is the minimal (w.r.t.  $\sqsubseteq$ ) labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$
- $F_{PR}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L} \mid \mathbf{L} \text{ is a maximal (w.r.t. } \sqsubseteq) \text{ labelling in } F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})\}$ .

Differently from stable semantics, grounded semantics and preferred semantics do not inherit decomposability from complete semantics. The reason is that the definition of grounded/preferred labelling includes a minimization/maximization requirement, and satisfying this requirement in all of the subframeworks does not entail satisfying it at the level of the whole framework. To show this, consider the following counterexample.<sup>9</sup>

**Example 5.** We have shown in [Example 2](#) that in the running example of [Fig. 1](#) the outcome of  $F_{CO}$  is a unique labelling in all cases, thus by definition it coincides with the outcome of  $F_{GR}$  and  $F_{PR}$ . Given the compatibility constraint, exactly two global labellings arise from the combinations of the outcomes of  $F_{CO}$ , namely  $\{(A, \text{undec}), (B, \text{undec}), (C, \text{undec}), (D, \text{undec})\}$  and  $\{(A, \text{out}), (B, \text{in}), (C, \text{out}), (D, \text{in})\}$ . The former is the grounded labelling, the latter is the preferred labelling: it turns out that the combination of two “locally grounded” labellings gives rise not just to the “global” grounded labelling but also to the preferred labelling, and analogously that the combination of two “locally preferred” labellings gives rise not just to the “global” preferred labelling but also to the grounded one. This shows that grounded and preferred semantics are not bottom-up decomposable.

Now, a question arises as to whether satisfying the minimization/maximization requirement at the level of the whole argumentation framework entails that such requirement is satisfied at the local level, i.e. whether grounded and preferred semantics are top-down decomposable. This result turns out to be true and is achieved through some intermediate steps.

First, [Lemma 1](#) shows that if a labelling produced by  $F_{AD}$  does not belong to  $F_{CO}$  then there is an  $\text{undec}$ -labelled argument which can be labelled  $\text{in}$  or  $\text{out}$  obtaining a labelling still in  $F_{AD}$ .

**Lemma 1.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where  $AF = (Ar, att)$ , let  $\mathbf{L}$  be a labelling such that  $\mathbf{L} \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $\mathbf{L} \notin F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Then there is an argument  $A \in Ar$  such that  $\mathbf{L}(A) = \text{undec}$  and a labelling  $\mathbf{L}^A \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}^A(A) \in \{\text{in}, \text{out}\}$  and  $\forall B \in Ar : B \neq A, \mathbf{L}^A(B) = \mathbf{L}(B)$ .

[Lemma 2](#) shows that for every labelling produced by  $F_{AD}$  there is a more or equally committed labelling produced by  $F_{CO}$ .

**Lemma 2.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , for every labelling  $\mathbf{L}_1 \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  there exists a labelling  $\mathbf{L}_2 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ .

<sup>9</sup> A counterexample to decomposability of grounded semantics is provided also in [\[40\]](#).

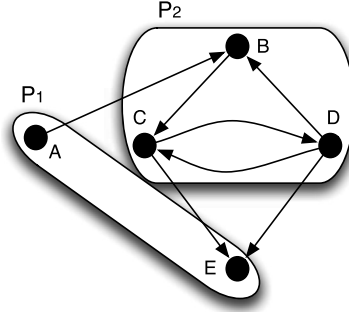


Fig. 3. A partition belonging to  $\mathcal{F}_{USCC}$  (Example 6).

Proposition 7 shows a sort of monotonicity property of  $F_{CO}$  with respect to the  $\sqsubseteq$  relation.

**Proposition 7.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , let  $L_{\mathcal{J}}^1, L_{\mathcal{J}}^2 \in \mathcal{L}_{\mathcal{J}}$  be two labellings of  $\mathcal{J}$  such that  $L_{\mathcal{J}}^1 \sqsubseteq L_{\mathcal{J}}^2$ . Then it holds that

1.  $\forall \mathbf{L}_1 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}}), \exists \mathbf{L}_2 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ ; and
2.  $\forall \mathbf{L}_2 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}}), \exists \mathbf{L}_1 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ .

Building on the above results (more specifically, using the first point for Theorem 6 and the second point for Theorem 5) we are now in a position to prove that grounded and preferred semantics are top-down decomposable.

**Theorem 5.** Given an argumentation framework  $AF = (Ar, att)$ , let  $\mathbf{L}$  be the grounded labelling of  $AF$ . For any set  $P \subseteq Ar$ ,  $\mathbf{L}_{\downarrow P} \in F_{GR}(AF_{\downarrow P}, P^{inp}, \mathbf{L}_{\downarrow P^{inp}}, P^R)$ .

**Theorem 6.** Given an argumentation framework  $AF = (Ar, att)$ , let  $\mathbf{L}$  be a preferred labelling of  $AF$ . For any set  $P \subseteq Ar$ ,  $\mathbf{L}_{\downarrow P} \in F_{PR}(AF_{\downarrow P}, P^{inp}, \mathbf{L}_{\downarrow P^{inp}}, P^R)$ .

While preferred and complete semantics fail to achieve bottom-up decomposability for arbitrary partitions, they turn out to be bottom-up decomposable (thus fully decomposable) w.r.t.  $\mathcal{F}_{USCC}$ . The result, proved in Theorem 7, is based on a preliminary lemma, which roughly states that if a semantics  $\mathbf{S}$  is top-down decomposable then a kind of top-down decomposability relation holds for any labelling  $\mathbf{L} \in F_{\mathbf{S}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  w.r.t. any set of arguments  $P$  in  $AF$ . More specifically, given such a labelling  $\mathbf{L}$  and  $P$ , it is possible to refer to a “restricted” argumentation framework with input based on  $P$ , namely  $(AF_{\downarrow P}, P^{F-inp}, (\mathbf{L} \cup L_{\mathcal{J}})_{\downarrow P^{F-inp}}, P_F^R)$ , where intuitively  $P^{F-inp}$ ,  $(\mathbf{L} \cup L_{\mathcal{J}})_{\downarrow P^{F-inp}}$ , and  $P_F^R$  are obtained by considering both  $\mathcal{J}$  and  $AF$  (outside  $P$ ). Then, the restriction of  $\mathbf{L}$  to  $P$  is produced by  $F_{\mathbf{S}}$  when applied to the restricted argumentation framework with input mentioned above.

**Lemma 3.** Let  $\mathbf{S}$  be a complete-compatible semantics which is top-down decomposable, with the canonical local function  $F_{\mathbf{S}}$ . Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , consider a labelling  $\mathbf{L} \in F_{\mathbf{S}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and let  $P \subseteq Ar$  be an arbitrary set of arguments of  $AF$ . Then, letting  $P^{F-inp} \triangleq P^{inp} \cup \{A \in \mathcal{J} \mid \exists B \in P, (A, B) \in R_{\mathcal{J}}\}$  and  $P_F^R \triangleq P^R \cup (R_{\mathcal{J}} \cap (\mathcal{J} \times P))$ , it holds that  $\mathbf{L}_{\downarrow P} \in F_{\mathbf{S}}(AF_{\downarrow P}, P^{F-inp}, (\mathbf{L} \cup L_{\mathcal{J}})_{\downarrow P^{F-inp}}, P_F^R)$ .

**Theorem 7.** Grounded and preferred semantics are decomposable w.r.t.  $\mathcal{F}_{USCC}$ .

**Example 6.** Consider  $AF = (\{A, B, C, D, E\}, \{(A, B), (B, C), (C, D), (D, C), (D, B), (C, E), (D, E)\})$  and the partition  $\{P_1, P_2\} \in \mathcal{F}_{USCC}(AF)$  where  $P_1 = \{A, E\}$  and  $P_2 = \{B, C, D\}$  (see Fig. 3). It holds that  $P_1^{inp} = \{C, D\}$ ,  $P_1^R = \{(C, E), (D, E)\}$ ,  $P_2^{inp} = \{A\}$ ,  $P_2^R = \{(A, B)\}$ . Note that the partition is not “acyclic”, in that  $P_1$  attacks  $P_2$  and  $P_2$  attacks  $P_1$ . We show that both in the case of grounded semantics and of preferred semantics the union of compatible local labellings gives rise to the grounded labelling or a preferred labelling, respectively. First, note that any labelling returned by  $F_{GR}$  and  $F_{PR}$  applied to  $AF_{\downarrow P_1}$  prescribes that  $A$  is labelled *in*, therefore it suffices to consider the labelling  $\{(A, in)\}$  for the unique input argument of  $P_2$ . As to grounded semantics, it turns out that  $F_{GR}(AF_{\downarrow P_2}, \{A\}, \{(A, in)\}, \{(A, B)\}) = \{(B, out), (C, undec), (D, undec)\}$ , while  $F_{GR}(AF_{\downarrow P_1}, \{C, D\}, \{(C, undec), (D, undec)\}, \{(C, E), (D, E)\}) = \{(A, in), (E, undec)\}$ . We have a unique pair of compatible local labellings which give rise to the global labelling  $\{(A, in), (B, out), (C, undec), (D, undec), (E, undec)\}$ , i.e. the grounded labelling of  $AF$ . As to preferred semantics,  $F_{PR}(AF_{\downarrow P_2}, \{A\}, \{(A, in)\}, \{(A, B)\})$  returns two labellings, i.e.  $\{(B, out), (C, in), (D, out)\}$  and  $\{(B, out), (C, out), (D, in)\}$ , while  $F_{PR}(AF_{\downarrow P_1}, \{C, D\},$

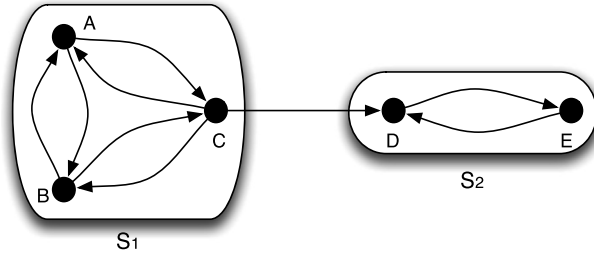


Fig. 4. Ideal semantics is neither top-down nor bottom-up decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$  (Example 7).

$\{(C, \text{in}), (D, \text{out})\}, \{(C, E), (D, E)\} = F_{\text{PR}}(AF \downarrow_{P_1}, \{C, D\}, \{(C, \text{out}), (D, \text{in})\}, \{(C, E), (D, E)\}) = \{(A, \text{in}), (E, \text{out})\}$ . Accordingly, the union of compatible local labellings gives rise to  $\{(A, \text{in}), (B, \text{out}), (C, \text{in}), (D, \text{out}), (E, \text{out})\}$  and  $\{(A, \text{in}), (B, \text{out}), (C, \text{out}), (D, \text{in}), (E, \text{out})\}$ , i.e. the preferred labellings of AF.

#### 4.4. Ideal semantics

Similarly to the cases analyzed in the previous sections, the canonical local function of ideal semantics corresponds to an extension of the definition of ideal labelling. The following proposition identifies the relevant definition, also showing that it is well founded (in particular, that  $F_{\text{ID}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  always returns a unique labelling).

**Proposition 8.** *The canonical local function of ideal semantics is defined as  $F_{\text{ID}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L}^*\}$ , where  $\mathbf{L}^*$  is the maximal (w.r.t.  $\sqsubseteq$ ) labelling in  $F_{\text{CO}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that for each  $\mathbf{L}_P \in F_{\text{PR}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  it holds that  $\mathbf{L}^* \sqsubseteq \mathbf{L}_P$ .*

Ideal semantics has some common features both with preferred and with grounded semantics: on the one hand, its definition is based on the preferred labellings, on the other hand it yields a unique labelling, like the grounded semantics. As a matter of fact, a formal skepticism comparison between semantics shows that ideal semantics lies between grounded and preferred semantics [9]. Ideal semantics does not inherit any decomposability property from them: the following example shows that ideal semantics is neither top-down nor bottom-up decomposable even w.r.t.  $\mathcal{F}_{\text{SCC}}$ .

**Example 7.**  $AF = (\{A, B, C, D, E\}, \{(A, B), (B, A), (A, C), (C, A), (B, C), (C, B), (C, D), (D, E), (E, D)\})$  has the unique partition  $\{S_1, S_2\} \in \mathcal{F}_{\text{SCC}}(AF)$ , where  $S_1 = \{A, B, C\}$  and  $S_2 = \{D, E\}$  are the strongly connected components of AF (see Fig. 4). There are 5 preferred labellings of AF and there is no argument which is labelled *in* in all of them, thus the ideal labelling  $\mathbf{L}^*$  leaves all of the arguments undecided. To show that ideal semantics is not top-down decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$ , it is sufficient to note that  $\mathbf{L}^* \downarrow_{S_2} = \{(D, \text{undec}), (E, \text{undec})\}$ , while it turns out that  $F_{\text{ID}}(AF \downarrow_{S_2}, \{C\}, \{(C, \text{undec})\}, \{(C, D)\}) = \{(D, \text{out}), (E, \text{in})\}$ .

To show that ideal semantics is not bottom-up decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$ , consider first the application of  $F_{\text{ID}}$  to  $AF \downarrow_{S_1}$ : it is easy to see that  $F_{\text{ID}}(AF \downarrow_{S_1}, \emptyset, \emptyset, \emptyset) = \{(A, \text{undec}), (B, \text{undec}), (C, \text{undec})\}$ , since  $AF \downarrow_{S_1}$  admits the three preferred labellings where one of the three arguments  $\{A, B, C\}$  is *in* and the others are *out*. Moreover, we already know that  $F_{\text{ID}}(AF \downarrow_{S_2}, \{C\}, \{(C, \text{undec})\}, \{(C, D)\}) = \{(D, \text{out}), (E, \text{in})\}$ , thus the labellings  $\{(A, \text{undec}), (B, \text{undec}), (C, \text{undec})\}$  and  $\{(D, \text{out}), (E, \text{in})\}$  are compatible. However, the union of these two labellings does not coincide with the ideal labelling  $\mathbf{L}^*$ .

The previous example contradicts<sup>10</sup> a result presented in [33], according to which ideal semantics is decomposable w.r.t. partitions including two elements one of which is unattacked (i.e. does not receive attacks from outside,  $S_1$  in Fig. 4). The reason why ideal semantics is not decomposable is that, considering a strongly connected component  $P$ , the restriction of the ideal labelling to the input arguments of  $P$  does not always carry enough information to compute the restriction of the ideal labelling to  $P$ . In the previous example, argument  $C$  is labelled *undec* by the ideal labelling while it is labelled *in* or *out* by the preferred labellings, i.e. those which actually determine the ideal labelling according to Definition 8.

#### 4.5. Semi-stable semantics

The definition of semi-stable semantics somewhat resembles that of preferred semantics, in that semi-stable labellings correspond to those preferred labellings which satisfy the additional requirement of minimizing the set of arguments labelled *undec*. The following proposition shows that the canonical local function is defined accordingly.

<sup>10</sup> A detailed discussion of this matter is given in [11].

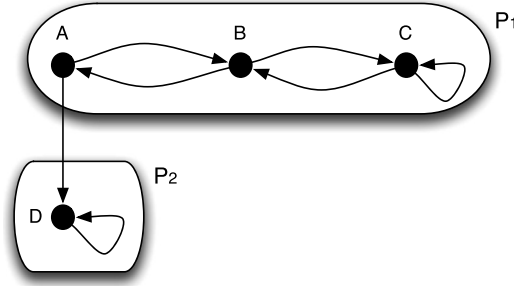


Fig. 5. Semi-stable semantics is not top-down decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$  (Example 8).

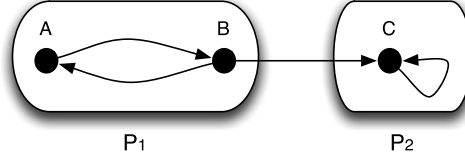


Fig. 6. Semi-stable semantics is not bottom-up decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$  (Example 9).

**Proposition 9.** The canonical local function of semi-stable semantics is defined as  $F_{\text{SST}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L} \mid \mathbf{L} \in F_{\text{CO}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \text{ such that } \text{undec}(\mathbf{L}) \text{ is minimal w.r.t. set inclusion}\}$ .

Differently from all semantics considered above, semi-stable semantics is not directional [8], i.e. given an unattacked set of arguments  $S$  the labellings computed in  $\text{AF} \downarrow_S$  do not correspond to the restrictions of the labellings of  $\text{AF}$  in  $S$ . As shown in the following two examples, this behavior prevents the satisfaction of top-down and bottom-up decomposability even w.r.t.  $\mathcal{F}_{\text{SCC}}$ .

**Example 8.** To show that semi-stable semantics is not top-down decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$ , consider  $\text{AF} = (\{A, B, C, D\}, \{(A, B), (B, A), (B, C), (C, B), (C, C), (A, D), (D, D)\})$ , where  $\text{SCCS}_{\text{AF}} = \{P_1, P_2\}$  with  $P_1 = \{A, B, C\}$  and  $P_2 = \{D\}$  (see Fig. 5). There are two semi-stable labellings in  $\text{AF}$ , namely  $\mathbf{L}_1 = \{(A, \text{in}), (B, \text{out}), (C, \text{undec}), (D, \text{out})\}$  and  $\mathbf{L}_2 = \{(A, \text{out}), (B, \text{in}), (C, \text{out}), (D, \text{undec})\}$ . Consider then the partition  $\{P_1, P_2\} \in \mathcal{F}_{\text{SCC}}(\text{AF})$  where  $P_1$  is unattacked. Note in particular that  $\mathbf{L}_1 \downarrow_{P_1} = \{(A, \text{in}), (B, \text{out}), (C, \text{undec})\}$ , which however does not belong to  $F_{\text{SST}}(\text{AF} \downarrow_{P_1}, \emptyset, \emptyset, \emptyset)$ , since the only semi-stable labelling in  $\text{AF} \downarrow_{P_1}$  is  $\{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$ .

**Example 9.** To show that semi-stable semantics is not bottom-up decomposable w.r.t.  $\mathcal{F}_{\text{SCC}}$ , consider the argumentation framework  $\text{AF} = (\{A, B, C\}, \{(A, B), (B, A), (B, C), (C, C)\})$  and the partition  $\{P_1, P_2\} \in \mathcal{F}_{\text{SCC}}(\text{AF})$  with  $P_1 = \{A, B\}$  and  $P_2 = \{C\}$  (see Fig. 6). It is easy to see that  $\{(A, \text{in}), (B, \text{out})\} \in F_{\text{SST}}(\text{AF} \downarrow_{P_1}, \emptyset, \emptyset, \emptyset)$ , and that  $F_{\text{SST}}(\text{AF} \downarrow_{P_2}, \{B\}, \{(B, \text{out})\}, \{(B, C)\}) = \{(C, \text{undec})\}$ . Now, the union of these compatible labellings, i.e.  $\{(A, \text{in}), (B, \text{out}), (C, \text{undec})\}$ , is not a semi-stable labelling of  $\text{AF}$ , since the unique semi-stable labelling of  $\text{AF}$  is  $\{(A, \text{out}), (B, \text{in}), (C, \text{out})\}$ .

## 5. Effect-dictated semantics

This short section introduces the simple concept of *effect-dictated* semantics, which is crucial for the analysis to be carried out in the next section. For every semantics  $\mathbf{S}$  analyzed in Section 4, it can be noted that  $F_{\mathbf{S}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  may return the same result given different  $\mathcal{J}$ ,  $L_{\mathcal{J}}$  and  $R_{\mathcal{J}}$ . For instance, if an argument  $A$  of  $\text{AF}$  is attacked by an argument of  $\mathcal{J}$  which is labelled *in*, then  $F_{\mathbf{S}}$  returns the same set of labellings independently of the presence and the number of additional attackers of  $A$  in  $\mathcal{J}$ . The *effect* of  $(\mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  on the arguments  $\text{Args}$  of  $\text{AF}$  can be modelled as the labelling that would be induced on  $\text{Args}$  by neglecting the attacks inside  $\text{AF}$ . For instance, if an argument  $A$  of  $\text{AF}$  is only attacked through  $R_{\mathcal{J}}$  by *out*-labelled arguments according to  $L_{\mathcal{J}}$ , then  $A$  would be *in* in the case that it does not receive other attacks inside  $\text{AF}$ .<sup>11</sup> The following definition formalizes this intuition.

**Definition 26.** Given a set of arguments  $\mathcal{J}$ , a labelling  $L_{\mathcal{J}} \in \mathcal{L}_{\mathcal{J}}$ , a set of arguments  $\text{Args}$  such that  $\mathcal{J} \cap \text{Args} = \emptyset$  and a relation  $R_{\text{INP}} \subseteq \mathcal{J} \times \text{Args}$ , the effect of  $(\mathcal{J}, L_{\mathcal{J}}, R_{\text{INP}})$  on  $\text{Args}$ , denoted as  $\text{eff}_{\text{Args}}(\mathcal{J}, L_{\mathcal{J}}, R_{\text{INP}})$ , is defined as

<sup>11</sup> The effect is a bit different for admissible semantics, but this does not affect its technical treatment, as well as the subsequent results.



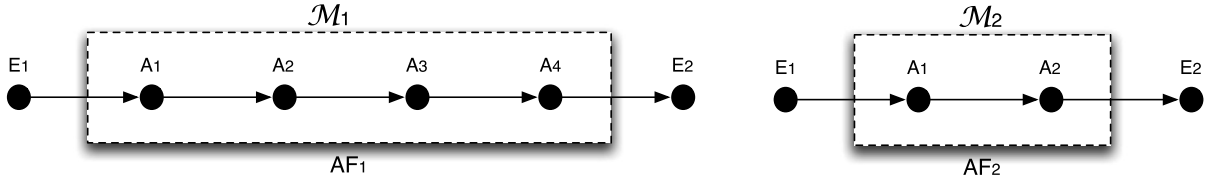


Fig. 7. Summarizing a chain of arguments (Example 10).

$$\begin{aligned} & \{(A, \text{out}) \mid A \in \text{Args}, \exists B \in \mathcal{J} : (B, A) \in R_{\text{INP}} \wedge L_{\mathcal{J}}(B) = \text{in}\} \cup \\ & \{(A, \text{undec}) \mid A \in \text{Args}, \exists B \in \mathcal{J} : (B, A) \in R_{\text{INP}} \wedge L_{\mathcal{J}}(B) = \text{undec}, \nexists C \in \mathcal{J} : (C, A) \in R_{\text{INP}} \wedge L_{\mathcal{J}}(C) = \text{in}\} \cup \\ & \{(A, \text{in}) \mid A \in \text{Args}, \nexists B \in \mathcal{J} : (B, A) \in R_{\text{INP}} \wedge L_{\mathcal{J}}(B) \in \{\text{in}, \text{undec}\}\} \end{aligned}$$

By definition,  $\text{eff}_{\text{Args}}(\mathcal{J}, L_{\mathcal{J}}, R_{\text{INP}})$  only depends on the labelling of the arguments in  $\mathcal{J}$  that attack  $\text{Args}$  through  $R_{\text{INP}}$ . Moreover each argument in  $\text{Args}$  not receiving attacks from  $\mathcal{J}$  is labelled *in* according to  $\text{eff}_{\text{Args}}(\mathcal{J}, L_{\mathcal{J}}, R_{\text{INP}})$ . Thus, in the particular case where  $\mathcal{J} = \emptyset$  (thus also  $L_{\mathcal{J}}$  and  $R_{\text{INP}}$  are empty), it turns out that  $\text{eff}_{\text{Args}}(\emptyset, \emptyset, \emptyset) = \{(A, \text{in}) \mid A \in \text{Args}\}$ .

The following lemma proves a monotonic relation between labellings and effects.

**Lemma 4.** Given a set of arguments  $\mathcal{J}$ , two labellings  $L_{\mathcal{J}}^1, L_{\mathcal{J}}^2 \in \mathcal{L}_{\mathcal{J}}$ , a set of arguments  $\text{Args}$  such that  $\mathcal{J} \cap \text{Args} = \emptyset$  and a relation  $R_{\text{INP}} \subseteq \mathcal{J} \times \text{Args}$ , if  $L_{\mathcal{J}}^1 \subseteq L_{\mathcal{J}}^2$  then  $\text{eff}_{\text{Args}}(\mathcal{J}, L_{\mathcal{J}}^1, R_{\text{INP}}) \subseteq \text{eff}_{\text{Args}}(\mathcal{J}, L_{\mathcal{J}}^2, R_{\text{INP}})$ .

A semantics **S** is said to be *effect-dictated* if, given  $AF = (Ar, att)$ ,  $F_{\mathbf{S}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  only depends on  $\text{eff}_{Ar}(\mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , rather than on the whole labelling  $L_{\mathcal{J}}$  and the specific relation  $R_{\mathcal{J}}$ .

**Definition 27.** A semantics **S** is effect-dictated if  $(\text{eff}_{Ar}(\mathcal{J}_1, L_{\mathcal{J}_1}, R_{\mathcal{J}_1}) = \text{eff}_{Ar}(\mathcal{J}_2, L_{\mathcal{J}_2}, R_{\mathcal{J}_2})) \Rightarrow F_{\mathbf{S}}(AF, \mathcal{J}_1, L_{\mathcal{J}_1}, R_{\mathcal{J}_1}) = F_{\mathbf{S}}(AF, \mathcal{J}_2, L_{\mathcal{J}_2}, R_{\mathcal{J}_2})$  for every  $AF$ ,  $\mathcal{J}_1, \mathcal{J}_2, L_{\mathcal{J}_1}, L_{\mathcal{J}_2}, R_{\mathcal{J}_1}$  and  $R_{\mathcal{J}_2}$ , where  $AF = (Ar, att)$  is an argumentation framework,  $\mathcal{J}_1$  and  $\mathcal{J}_2$  are two sets of arguments such that  $\mathcal{J}_1 \cap Ar = \emptyset$  and  $\mathcal{J}_2 \cap Ar = \emptyset$ ,  $L_{\mathcal{J}_1} \in \mathcal{L}_{\mathcal{J}_1}$  and  $L_{\mathcal{J}_2} \in \mathcal{L}_{\mathcal{J}_2}$  two labellings of  $\mathcal{J}_1$  and  $\mathcal{J}_2$  respectively, and  $R_{\mathcal{J}_1} \subseteq \mathcal{J}_1 \times Ar$  and  $R_{\mathcal{J}_2} \subseteq \mathcal{J}_2 \times Ar$  two relations.

All the semantics considered in this paper are effect-dictated as shown by the following proposition.

**Proposition 10.** Every semantics **S**  $\in \{\text{AD}, \text{CO}, \text{ST}, \text{GR}, \text{PR}, \text{ID}, \text{SST}\}$  is effect-dictated.

## 6. Argumentation multipoles and their interchangeability

In this section, we introduce argumentation multipoles, that are conceived as modular components equipped with a well-defined interface to connect with each other and may play the role of “partial” frameworks in the context of a global one. This yields the possibility of replacing a component with another one which is equivalent as far as the Input/Output behavior is concerned.

### 6.1. The notion of argumentation multipole

The first step to provide a systematic treatment of argumentation multipoles is to identify a definition to capture their structure in the most general way. To this aim, we consider a number of examples, starting from a common component, i.e. a chain of arguments.

**Example 10.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 7.  $AF_2$  can be obtained from  $AF_1$  by “summarizing” the component  $\mathcal{M}_1$ , including the arguments  $A_1, A_2, A_3, A_4$ , with the component  $\mathcal{M}_2$ , including the arguments  $A_1$  and  $A_2$ : according to any complete-compatible semantics considered in this paper, the labellings restricted to  $E_1$  and  $E_2$ , i.e. the arguments common to  $AF_1$  and  $AF_2$ , are the same in the two frameworks, i.e.  $E_1$  is labelled *in* and  $E_2$  is labelled *out*. More generally, consider a finite sequence of  $n$  arguments  $A_1, \dots, A_n$  such that each argument attacks the subsequent one, i.e.  $A_i$  attacks  $A_{i+1}$  with  $1 \leq i < n$  and suppose that only  $A_1$  can receive further attacks from other arguments and only  $A_n$  can attack other arguments. Then it is intuitive to see that the “black-box behavior” of a sequence of arguments of this kind, whose external “terminals” are  $A_1$  and  $A_n$ , only depends on whether  $n$  is even or odd. In fact, the behavior of any even-length sequence is the same as in the case  $n = 2$  (if  $A_1$  is *in* then  $A_n$  is *out*, if  $A_1$  is *out* then  $A_n$  is *in*, if  $A_1$  is *undec* then  $A_n$  is *undec*), while for any odd-length sequence the behavior is the same as the one of  $A_1$  alone (with  $n$  odd,  $A_n$  gets necessarily the same label as  $A_1$ ).

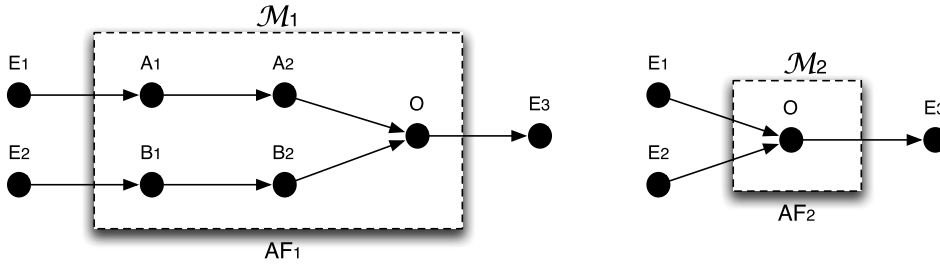


Fig. 8. Summarizing two chains of arguments attacking an argument  $O$  (Example 11).

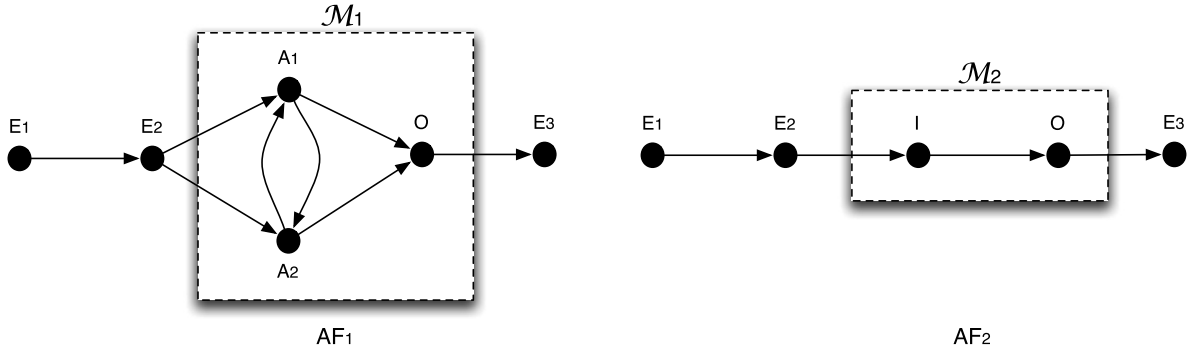


Fig. 9. Summarizing two contradicting arguments attacking an argument  $O$  (Example 12).

On the basis of the previous example, a modular component may tentatively be defined as an argumentation framework<sup>12</sup> where the “input terminals” and the “output terminals” are explicitly identified (e.g.  $AF_1 \downarrow \{A_1, \dots, A_4\}$  in the example, where  $A_1$  is the unique input terminal and  $A_4$  in the unique output terminal). Two components can be interchanged only if they have the same input and output terminals, and this interchange does not modify the attacks relating these terminals with the unchanged arguments ( $E_1$  and  $E_2$  in the example). However, the following two examples show that this approach is too restrictive, since there are cases where it is useful to modify both the set of input and output terminals as well as the relevant attack relation.

**Example 11.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 8.  $AF_2$  can be obtained from  $AF_1$  by summarizing the component  $\mathcal{M}_1$ , including the arguments  $A_1, A_2, B_1, B_2, O$ , with the component  $\mathcal{M}_2$  including the argument  $O$  only: according to all complete-compatible semantics considered in this paper the arguments  $E_1, E_2$  and  $E_3$  are labelled *in* both in  $AF_1$  and  $AF_2$ . More generally, the black-box behavior of  $\mathcal{M}_1$  is the same as the one of  $\mathcal{M}_2$ , since in  $\mathcal{M}_1$   $A_2$  gets the same label as  $E_1$  and  $B_2$  gets the same label as  $E_2$ , thus the label of  $O$  is the same as in  $\mathcal{M}_2$ . As a consequence, one may expect that  $\mathcal{M}_1$  can be interchanged with  $\mathcal{M}_2$  also in more articulated examples. Note that while  $\mathcal{M}_1$  has two input terminals,  $\mathcal{M}_2$  has only one input terminal coinciding with the unique output one.

**Example 12.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 9 and assume preferred semantics is adopted.  $AF_2$  can be obtained from  $AF_1$  by summarizing the component  $\mathcal{M}_1$ , including the arguments  $A_1, A_2$  and  $O$ , with the component  $\mathcal{M}_2$  including the arguments  $I$  and  $O$ : both in  $AF_1$  and  $AF_2$  the argument  $E_1$  is labelled *in*,  $E_2$  is labelled *out* and  $E_3$  is labelled *in*. More generally, under preferred semantics the black-box behavior of  $\mathcal{M}_1$  is the same as the one of  $\mathcal{M}_2$ : if  $E_2$  is *in* then  $O$  is *in*, if  $E_2$  is *out* then  $O$  is *out* (in particular  $\mathcal{M}_1$  admits a labelling where  $A_1$  is *in*,  $A_2$  is *out* and  $O$  is *out*, and a labelling where  $A_1$  is *out*,  $A_2$  is *in* and  $O$  is *out*), if  $E_2$  is *undec* then  $O$  is *undec*. As a consequence, one may expect that  $\mathcal{M}_1$  can be interchanged with  $\mathcal{M}_2$  also in more articulated examples. Note that while  $\mathcal{M}_1$  receives two attacks from  $E_2$  in  $AF_1$ ,  $\mathcal{M}_2$  receives one attack only in  $AF_2$ .

The previous examples show that the definition of a modular component should include the *input attack relation*  $R_{INP}$ , consisting of the attacks from the arguments that are not part of the component to the arguments that belong to the component itself: this way, the definition leaves room for replacements of modular components that lead to changes in the input attack relation, as in the previous example. A similar reasoning concerns the *output attack relation*  $R_{OUTP}$ , including the attacks from a modular component towards the outside arguments. In any case, there is no need to explicitly model the input and output terminals, since they can easily be derived from the input and output attack relations. Inspired by the

<sup>12</sup> This approach has been followed in our paper [2], leading to the notion of *Input/Output Argumentation Framework*.

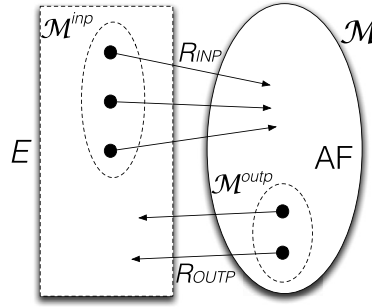


Fig. 10. A graphical representation of the notion of argumentation multipole.

digital logic field, we call the resulting structure an *Argumentation Multipole*. In order to express  $R_{INP}$  and  $R_{OUTP}$ , without loss of generality we define an Argumentation Multipole w.r.t. a set  $E$ , i.e. w.r.t. the set of arguments that are not part of the multipole and thus remain unchanged if the multipole is replaced.

**Definition 28.** An Argumentation Multipole (or, briefly, multipole)  $\mathcal{M}$  w.r.t. a set  $E$  is a tuple  $(AF, R_{INP}, R_{OUTP})$ , where letting  $AF = (Ar, att)$  it holds that  $Ar \cap E = \emptyset$ ,  $R_{INP} \subseteq E \times Ar$ , and  $R_{OUTP} \subseteq Ar \times E$ . Extending the notation introduced in Definition 10, we denote as  $\mathcal{M}^{inp}$  the set  $\{A \in E \mid \exists B \in Ar, (A, B) \in R_{INP}\}$ , i.e. including the arguments of  $E$  which attack  $Ar$  through  $R_{INP}$ . Moreover, we denote as  $\mathcal{M}^{outp}$  the set  $\{A \in Ar \mid \exists B \in E, (A, B) \in R_{OUTP}\}$ , i.e. including the arguments of  $AF$  attacking  $E$  through  $R_{OUTP}$ .

Fig. 10 provides a graphical representation of the definition. For instance, in Example 10  $\mathcal{M}_1 = (AF_1 \downarrow_{\{A_1, A_2, A_3, A_4\}}, \{(E_1, A_1)\}, \{(A_4, E_2)\})$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{\{A_1, A_2\}}, \{(E_1, A_1)\}, \{(A_2, E_2)\})$ , in Example 11 it holds that  $\mathcal{M}_1 = (AF_1 \downarrow_{\{A_1, A_2, B_1, B_2, O\}}, \{(E_1, A_1), (E_2, B_1)\}, \{(O, E_3)\})$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{\{O\}}, \{(E_1, O), (E_2, O)\}, \{(O, E_3)\})$ , in Example 12  $\mathcal{M}_1 = (AF_1 \downarrow_{\{A_1, A_2, O\}}, \{(E_2, A_1), (E_2, A_2)\}, \{(O, E_3)\})$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{\{I, O\}}, \{(E_2, I)\}, \{(O, E_3)\})$ .

A particular multipole which is useful to consider in some practical examples is the *empty multipole*  $\mathcal{M}_\emptyset \triangleq (AF_\emptyset, \emptyset, \emptyset)$ , i.e. including the empty argumentation framework  $AF_\emptyset$ . It is easy to see that  $\mathcal{M}_\emptyset^{inp} = \mathcal{M}_\emptyset^{outp} = \emptyset$ .

## 6.2. Input/Output equivalence of argumentation multipoles

After having introduced the definition of argumentation multipole, the next step is to formally characterize the relevant “black-box behavior”: this way, the Input/Output equivalence relation between multipoles can be identified as the one relating the multipoles having the same behavior.

When a multipole w.r.t. a set  $E$  is “connected to the external world” it “receives” some input from outside through the relation  $R_{INP}$  and “produces” an output which is induced by the labellings of the multipole and transferred to the set  $E$  through the relation  $R_{OUTP}$ . Technically speaking, the labellings and thus the relation between input and output are determined by a (semantics specific) local function, thus the equivalence relation between argumentation multipoles depends on the considered semantics  $S$ , and is called  $S$ -equivalence to reflect this dependency. For instance, in Example 12  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent (i.e. equivalent under preferred semantics), while they are not **GR**-equivalent, since under grounded semantics if  $E_2$  is labelled out then  $O$  in  $\mathcal{M}_1$  is labelled undec, while  $O$  in  $\mathcal{M}_2$  is labelled out. Intuitively,  $\mathcal{M}_1$  is **GR**-equivalent e.g. to a multipole  $\mathcal{M}'_2$  obtained from  $\mathcal{M}_2$  by adding a self-attack from  $I$  to  $I$  itself.

According to the above examples, two argumentation multipoles w.r.t. the same set  $E$  may be tentatively defined as  $S$ -equivalent if for any possible input, i.e. any labelling of  $E$ ,  $F_S$  produces the same labellings of the output terminals in the two argumentation multipoles. For instance, in Example 12 under preferred semantics  $O$  is in for any labelling where  $E_2$  is in, it is out for any labelling where  $E_2$  is out and it is undec for any labelling where  $E_2$  is undec. However, this approach works only in case the two multipoles have the same output terminals. Moreover, as the following example shows, the way  $E$  is affected by the labellings of an argumentation multipole  $(AF, R_{INP}, R_{OUTP})$  also depends on the attack relation  $R_{OUTP}$ .

**Example 13.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 11 and the application of preferred semantics. The multipole  $\mathcal{M}_1 = (AF_1 \downarrow_{\{O_1, O_2\}}, \emptyset, \{(O_1, E), (O_2, E)\})$  w.r.t.  $\{E\}$  in  $AF_1$  affects the argument  $E$  by means of the two arguments  $O_1$  and  $O_2$ , while  $\mathcal{M}_2 = (AF_2 \downarrow_{\{O\}}, \emptyset, \{(O, E)\})$  in  $AF_2$  affects  $E$  by means of the argument  $O$ . Intuitively, under preferred semantics  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are equivalent: in  $\mathcal{M}_1$  there are two preferred labellings, i.e.  $\{(O_1, in), (O_2, out)\}$  and  $\{(O_1, out), (O_2, in)\}$ , thus in any case an argument labelled in attacks  $E$  making it out, and similarly  $\mathcal{M}_2$  interacts with  $E$  making it out, since  $\mathcal{M}_2$  admits the unique labelling  $\{(O, in)\}$ .

We can formalize these intuitions by extending the notion of effect to multipoles (see Definition 26). Let us consider a semantics  $S$ . Given a multipole  $\mathcal{M}$  w.r.t. a set  $E$ , for any “input” labelling  $L_E \in \mathcal{L}_E$  the local function  $F_S$  prescribes a set of

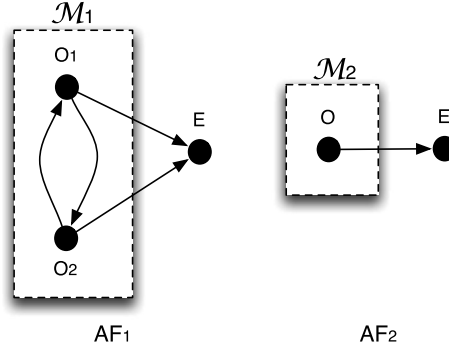


Fig. 11. Summarizing two contradicting arguments (Example 13).

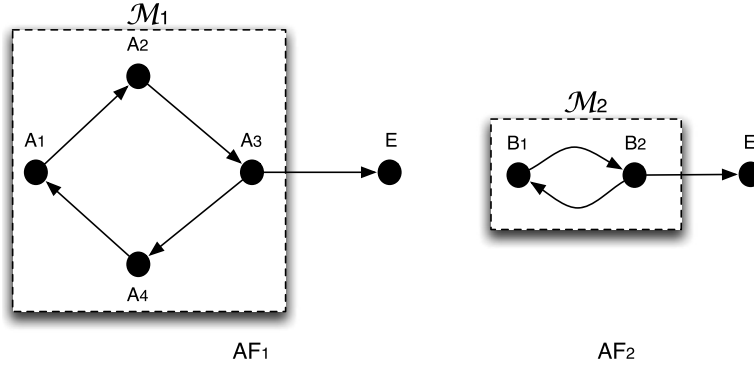


Fig. 12. Summarizing a 4-length cycle of arguments (Example 14).

labellings for  $\mathcal{M}$ . Each of these labellings has its own effect on  $E$ , therefore the global effect of the multipole receiving an input  $\mathbf{L}_E$  is a set of labellings of  $E$  whose members are all the single effects.

**Definition 29.** Let  $\mathcal{M} = (AF, R_{INP}, R_{OUTP})$  a multipole w.r.t. a set  $E$  and  $\mathbf{S}$  an argumentation semantics. Given a labelling  $\mathbf{L}_E \in \mathcal{L}_E$ , the  $\mathbf{S}$ -effect of  $(\mathcal{M}, \mathbf{L}_E)$  on  $E$ , denoted as  $\mathbf{S}\text{-eff}_E(\mathcal{M}, \mathbf{L}_E)$ , is defined as  $\{\text{eff}_E(\mathcal{M}^{\text{outp}}, \mathbf{L} \downarrow_{\mathcal{M}^{\text{outp}}}, R_{OUTP}) \mid \mathbf{L} \in F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{INP})\}$ .

Note that if  $F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{INP}) = \emptyset$ , i.e. the local function prescribes no labelling, then  $\mathbf{S}\text{-eff}_E(\mathcal{M}, \mathbf{L}_E) = \emptyset$ .

**Example 14.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 12, and the multipoles  $\mathcal{M}_1 = (AF_1 \downarrow_{\{A_1, A_2, A_3, A_4\}}, \emptyset, \{(A_3, E)\})$  w.r.t.  $\{E\}$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{\{B_1, B_2\}}, \emptyset, \{(B_2, E)\})$  w.r.t.  $\{E\}$ .  $\mathcal{M}_1$  has two preferred labellings, one where  $A_3$  is in and another where  $A_3$  is out, hence  $\mathbf{PR}\text{-eff}_{\{E\}}(\mathcal{M}_1, \emptyset) = \{\{(E_1, \text{in})\}, \{(E_1, \text{out})\}\}$ . Similarly,  $\mathcal{M}_2$  has two preferred labellings, one where  $B_2$  is in and another where  $B_2$  is out, leading to  $\mathbf{PR}\text{-eff}_{\{E\}}(\mathcal{M}_1, \emptyset) = \mathbf{PR}\text{-eff}_{\{E\}}(\mathcal{M}_2, \emptyset)$ .

It is worth considering the effect of the empty multipole  $\mathcal{M}_\emptyset$ . Intuitively,  $\mathcal{M}_\emptyset$  should have no effect on the arguments of  $E$ , i.e. all of them should be assigned the label in according to the effect itself. Technically, this is guaranteed if the semantics is defined in such a way as to prescribe the unique possible labelling  $\emptyset$  to the empty argumentation framework  $AF_\emptyset$ , as it happens for any semantics considered in this paper. Intuitively, if this were not the case the empty multipole would prevent the identification of any labelling for the whole argumentation framework, yielding to a pathological behavior. Accordingly, the condition  $\mathbf{L}_{\mathbf{S}}(AF_\emptyset) = \{\emptyset\}$  is required in all the following propositions and theorems<sup>13</sup> referring to a generic semantics  $\mathbf{S}$ .

**Proposition 11.** Consider a semantics  $\mathbf{S}$  such that  $\mathbf{L}_{\mathbf{S}}(AF_\emptyset) = \{\emptyset\}$ . Given a set of arguments  $E$  and a labelling  $\mathbf{L}_E \in \mathcal{L}_E$ , it holds that  $\mathbf{S}\text{-eff}_E(\mathcal{M}_\emptyset, \mathbf{L}_E) = \{\{(A, \text{in}) \mid A \in E\}\}$ .

<sup>13</sup> The reader may wonder why this condition has never been considered in the context of decomposability properties. The reason is that decomposability refers to partitions of the argumentation framework, which by definition include nonempty sets only.

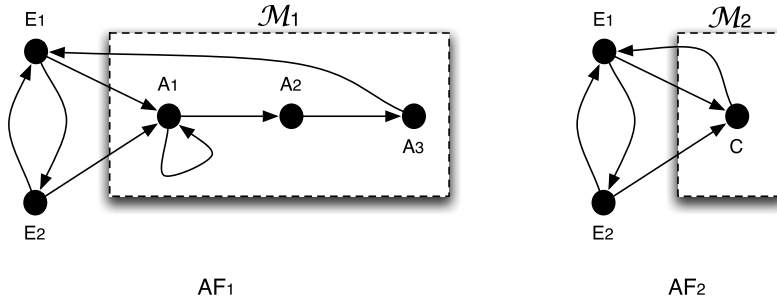


Fig. 13. A contextually PR-legitimate replacement (Examples 15 and 17).

Two multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  w.r.t.  $E$  can be considered **S-equivalent** if, for any possible labelling  $\mathbf{L}_E \in \mathcal{L}_E$ ,  $\mathbf{S}\text{-eff}_E(\mathcal{M}_1, \mathbf{L}_E) = \mathbf{S}\text{-eff}_E(\mathcal{M}_2, \mathbf{L}_E)$ . For reasons that will be clear later, it is also useful to identify multipoles that have the same effect only for a subset of input labellings: in order to capture this possibility, we define equivalence under a set of labellings of  $E$ .

**Definition 30.** Two multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  w.r.t. a set  $E$  are **Input/Output S-equivalent** (or simply **S-equivalent**) under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$  iff for any labelling  $\mathbf{L}_E \in \mathcal{L}'$  it holds that  $\mathbf{S}\text{-eff}_E(\mathcal{M}_1, \mathbf{L}_E) = \mathbf{S}\text{-eff}_E(\mathcal{M}_2, \mathbf{L}_E)$ . The multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **S-equivalent** iff they are **S-equivalent** under  $\mathcal{L}_E$ .

It is easy to see that if two multipoles w.r.t.  $E$  are **S-equivalent** then they are **S-equivalent** under any set  $\mathcal{L}' \subseteq \mathcal{L}_E$ .

In Example 10, Example 11 and Example 14  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **GR-equivalent** and **PR-equivalent**, while in Example 12 and Example 13  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR-equivalent** but not **GR-equivalent**.

### 6.3. Replacements and transparent argumentation semantics

As anticipated by previous examples, an argumentation multipole can be viewed as a component of an argumentation framework that can be *replaced* with another multipole giving rise to a (possibly) different argumentation framework. In particular, given an argumentation framework  $AF = (Ar, att)$ , one may partition the set of arguments  $Args$  into two sets, i.e. a set  $E$  which is not involved in the replacement and the set  $D_1 = Ar \setminus E$  which is replaced along with the relevant attacks: the set  $D_1$  identifies the multipole  $\mathcal{M}_1 = (AF \downarrow_{D_1}, att \cap (E \times D_1), att \cap (D_1 \times E))$  w.r.t.  $E$ , which can be replaced with another multipole  $\mathcal{M}_2$  w.r.t. the same set  $E$ . For later use in the paper, it is worth identifying those replacements such that a partition belonging to the set returned by a selector  $\mathcal{F}$  is enforced both before and after the replacement.

**Definition 31.** Let  $AF = (Ar, att)$  be an argumentation framework, and  $E \subseteq Ar$  be a subset of its arguments. Let  $D_1 = Ar \setminus E$ ,  $R_{INP}^1 = att \cap (E \times D_1)$  and  $R_{OUTP}^1 = att \cap (D_1 \times E)$ . A *replacement*  $\mathcal{R}$  is a tuple  $(AF, \mathcal{M}_1, \mathcal{M}_2)$  where  $\mathcal{M}_1 = (AF \downarrow_{D_1}, R_{INP}^1, R_{OUTP}^1)$  and  $\mathcal{M}_2$  is an argumentation multipole w.r.t.  $E$ . The set  $E$  is called the *invariant set* of the replacement  $\mathcal{R}$ . Assuming  $\mathcal{M}_2 = ((D_2, R_{D_2}), R_{INP}^2, R_{OUTP}^2)$ , the result of the replacement  $\mathcal{R}$ , denoted as  $T(\mathcal{R})$ , is the argumentation framework  $AF_2 = (E \cup D_2, (att \cap (E \times E) \cup R_{INP}^2 \cup R_{D_2} \cup R_{OUTP}^2))$ . Given a partition selector  $\mathcal{F}$ , a replacement  $(AF, \mathcal{M}_1, \mathcal{M}_2)$  is  $\mathcal{F}$ -preserving if both  $(\{E, D_1\} \setminus \emptyset) \in \mathcal{F}(AF)$  and  $(\{E, D_2\} \setminus \emptyset) \in \mathcal{F}(T(AF, \mathcal{M}_1, \mathcal{M}_2))$ .

It is easy to see that  $T(AF, \mathcal{M}_1, \mathcal{M}_1) = AF$ . Moreover, letting  $AF_2 = T(AF, \mathcal{M}_1, \mathcal{M}_2)$  it holds that  $T(AF_2, \mathcal{M}_2, \mathcal{M}_1) = AF$ . Note that, in the definition of  $\mathcal{F}$ -preserving replacement, the empty set is excluded from the requirement of belonging to  $\mathcal{F}(AF)$ . The reason is that by definition the empty set does not belong to any partition, however in case one of the sets in  $\{E, D_1\}$  or  $\{E, D_2\}$  is empty then it is sensible to require only the nonempty set to belong to  $\mathcal{F}(AF)$ .

In Examples 10–14, the result of the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is the argumentation framework  $AF_2$ .

While Definition 31 leaves room for any possible replacement, not all of them can be considered legitimate. In particular, we seek for replacements involving multipoles having the same Input/Output behavior, otherwise in most cases the labellings of the resulting frameworks would be different in the invariant set  $E$ , leading to changes in the status assignment of the relevant arguments. For instance, in Example 10 replacing  $\mathcal{M}_1$  in  $AF_1$  with a multipole including a single argument (or an odd-length chain of arguments) would change the label assigned to  $E_2$  from *out* to *in*. In order to explore the notion of legitimate replacements, let us consider an issue arising e.g. in the following example.

**Example 15.** Consider the application of preferred semantics on the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 13, where  $\mathcal{M}_1 = (AF_1 \downarrow_{\{A_1, A_2, A_3\}}, \{(E_1, A_1), (E_2, A_1)\}, \{(A_3, E_1)\})$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{\{C\}}, \{(E_1, C), (E_2, C)\}, \{(C, E_1)\})$  are two argumentation multipoles w.r.t.  $\{E_1, E_2\}$ ,  $AF_2 = T(\mathcal{R})$  with  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and the invariant set of the replacement  $\mathcal{R}$  is  $E = \{E_1, E_2\}$ . The multipole  $\mathcal{M}_1$  is not **PR-equivalent** to  $\mathcal{M}_2$ : considering the labelling  $\{(E_1, \text{out}), (E_2, \text{out})\}$   $F_{PR}$  prescribes for  $\mathcal{M}_1$  the unique labelling  $\{(A_1, \text{undec}), (A_2, \text{undec}), (A_3, \text{undec})\}$ , whose effect on  $\{E_1, E_2\}$  is  $\{(E_1, \text{undec}), (E_2, \text{in})\}$ , while  $F_{PR}$  prescribes for  $\mathcal{M}_2$  the unique labelling  $\{(C, \text{in})\}$ , whose effect on  $\{E_1, E_2\}$  is

$\{(E_1, \text{out}), (E_2, \text{in})\}$ . However, taking into account the possible labellings of  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , it can be noted that the labelling  $\{(E_1, \text{out}), (E_2, \text{out})\}$  is impossible both in  $AF_1$  and in  $AF_2$ . As to  $AF_1$ , if  $A_3$  is *in* then  $F_{\mathbf{PR}}$  prescribes for  $\{E_1, E_2\}$  the labelling  $\{(E_1, \text{out}), (E_2, \text{in})\}$ , if  $A_3$  is *out* then it prescribes the labellings  $\{(E_1, \text{out}), (E_2, \text{in})\}$  and  $\{(E_1, \text{in}), (E_2, \text{out})\}$ , if  $A_3$  is *undec* then it prescribes the labelling  $\{(E_1, \text{out}), (E_2, \text{in})\}$ . As to  $AF_2$ , the situation is the same. Summing up, the set of labellings that can be “seen” by  $\mathcal{M}_1$  and  $\mathcal{M}_2$  is  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}} = \{\{(E_1, \text{out}), (E_2, \text{in})\}, \{(E_1, \text{in}), (E_2, \text{out})\}\}$ , under which  $\mathcal{M}_1$  and  $\mathcal{M}_2$  turn out to be **PR**-equivalent. In fact, for each of the labellings in  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ ,  $F_{\mathbf{PR}}$  prescribes for  $\mathcal{M}_1$  the unique labelling  $\{(A_1, \text{out}), (A_2, \text{in}), (A_3, \text{out})\}$ , whose effect on  $\{E_1, E_2\}$  is  $\{(E_1, \text{in}), (E_2, \text{in})\}$ , and  $F_{\mathbf{PR}}$  prescribes for  $\mathcal{M}_2$  the unique labelling  $\{(C, \text{out})\}$ , whose effect on  $\{E_1, E_2\}$  is again  $\{(E_1, \text{in}), (E_2, \text{in})\}$ .

Thus, a replacement may be considered as legitimate even if the involved multipoles are not equivalent under all labellings, provided that they are equivalent under the *possible* ones (in a sense, input labellings that never occur are neglected as the “don’t care terms” in digital logic). Of course, one may accept to replace a multipole only with an equivalent one, since in this case equivalence holds independently of the context (in particular, the multipoles would remain equivalent even modifying the attack relations between arguments of the invariant set  $E$ ). In order to distinguish between the two cases, a replacement is called *contextually legitimate* in the first case, and simply *legitimate* in the latter. Independently of its legitimacy properties, we call *safe* a replacement that does not yield modifications of the labellings in  $E$ .

**Definition 32.** Let  $\mathbf{S}$  be an argumentation semantics and  $AF = (Ar, att)$  be an argumentation framework. A replacement  $\mathcal{R} = (AF, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$  is **S**-legitimate if  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **S**-equivalent, it is *contextually S-legitimate* if  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **S**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{S}}$ , where  $\mathcal{L}_{\mathcal{R}}^{\mathbf{S}} \triangleq \{F_{\mathbf{S}}(AF \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1, R_{\text{OUTP}}^1) \mid \mathbf{L}_1 \in \mathcal{L}_{\mathcal{M}_1^{\text{outp}}}^{\mathbf{S}}\} \cup \{F_{\mathbf{S}}(AF \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2, R_{\text{OUTP}}^2) \mid \mathbf{L}_2 \in \mathcal{L}_{\mathcal{M}_2^{\text{outp}}}^{\mathbf{S}}\}$ . Moreover,  $\mathcal{R}$  is **S**-safe if  $\{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(AF)\} = \{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(T(AF, \mathcal{M}_1, \mathcal{M}_2))\}$ .

It is easy to see that every legitimate replacement is also contextually legitimate. For instance, in [Example 12](#) the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is **PR**-legitimate and **PR**-safe, it is not contextually **GR**-legitimate nor **GR**-safe. In [Example 15](#) the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is contextually **PR**-legitimate (but not **PR**-legitimate) and **PR**-safe, and the same holds according to grounded semantics.

The examples presented so far may give the impression that for any semantics  $\mathbf{S}$  a (possibly contextually) **S**-legitimate replacement is always **S**-safe, i.e. replacing a multipole with an equivalent multipole preserves the labellings in the invariant set of the replacement. This property may seem natural and easy to prove, however it is shown in [Section 8](#) that it does not hold for all semantics: we denote as *transparent* the semantics such that legitimate replacements are always safe, *strongly transparent* the semantics such that contextually legitimate replacements are always safe. Similarly to decomposability, also transparency may hold under a restriction on the partition identified by the multipoles that are replaced: accordingly, we introduce the concept of transparency w.r.t. a partition selector  $\mathcal{F}$ .

**Definition 33.** A semantics  $\mathbf{S}$  is *transparent* if any **S**-legitimate replacement is **S**-safe, it is *strongly transparent* if any contextually **S**-legitimate replacement is **S**-safe. Given a partition selector  $\mathcal{F}$ , a semantics  $\mathbf{S}$  is transparent w.r.t.  $\mathcal{F}$  if any  $\mathcal{F}$ -preserving and **S**-legitimate replacement is **S**-safe, it is strongly transparent w.r.t.  $\mathcal{F}$  if any  $\mathcal{F}$ -preserving and contextually **S**-legitimate replacement is **S**-safe.

Since any ( $\mathcal{F}$ -preserving) legitimate replacement is also contextually legitimate, any strongly transparent semantics (w.r.t.  $\mathcal{F}$ ) is also transparent (w.r.t.  $\mathcal{F}$ ).

A limit case which is theoretically interesting to consider is a replacement  $(AF, \mathcal{M}_1, \mathcal{M}_2)$  with the invariant set  $E$  equal to the empty set, i.e. when an entire argumentation framework is replaced by another one.

**Proposition 12.** Consider a semantics  $\mathbf{S}$  such that  $\mathbf{L}_{\mathbf{S}}(AF_{\emptyset}) = \{\emptyset\}$  and a replacement  $\mathcal{R} = (AF, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E = \emptyset$ . Letting  $AF_2 = T(\mathcal{R})$ , the following conditions are equivalent:

- $\mathcal{R}$  is **S**-legitimate
- $\mathcal{R}$  is contextually **S**-legitimate
- $|\mathbf{L}_{\mathbf{S}}(AF)| > 0 \wedge |\mathbf{L}_{\mathbf{S}}(AF_2)| > 0$ , or  $\mathbf{L}_{\mathbf{S}}(AF) = \mathbf{L}_{\mathbf{S}}(AF_2) = \emptyset$
- $\mathcal{R}$  is **S**-safe.

Intuitively, there are no preserved arguments, thus the effect of any labelling of  $AF$  on the outside empty set is the same as the effect of any labelling of  $AF_2$ . The only difference arises in the case that  $AF$  “crashes” (i.e. admits no labellings) while  $AF_2$  does not exhibit such pathological behavior, or vice versa.

Note that the notions of replacement and transparent semantics refer to partitions of argumentation frameworks into just two subframeworks, i.e. one corresponding to the replaced multipole  $\mathcal{M}_1$  (or the replacing one  $\mathcal{M}_2$ ) and the other identified by the invariant set  $E$ . This is not restrictive, since one can treat a multiple replacement of several multipoles as a sequence of replacements each involving just one multipole. The following proposition shows that safeness is preserved



**Table 2**

Transparency properties of argumentation semantics.

	AD	CO	ST	GR	PR	ID	SST
(Strong) transparency	Yes	Yes	Yes	Yes	No <sup>a</sup>	No	No
(Strong) transparency w.r.t. $\mathcal{F}_{USCC}$ and $\mathcal{F}_{SCC}$	Yes	Yes	Yes	Yes	Yes	No	No
(Strong) transparency in case of acyclic multipoles	Yes	Yes	Yes	Yes	Yes	Yes	No

<sup>a</sup> holds under additional conditions (see Definitions 34 and 35).

by a sequence of safe replacements, and the same holds for skeptical and credulous justification of those arguments that are not replaced.

**Proposition 13.** Let  $AF = (Ar, att)$  be an argumentation framework. Consider a sequence of replacements  $(R_1, R_2, \dots, R_n)$  where  $R_i = (AF_i, \mathcal{M}_{i,1}, \mathcal{M}_{i,2})$ ,  $E_i$  is the invariant set of  $R_i$ ,  $AF_1 = AF$  and, for any  $1 < i \leq n$ ,  $AF_i = T(AF_{i-1}, \mathcal{M}_{i-1,1}, \mathcal{M}_{i-1,2})$ . Let  $AF_*$  be the result of the sequence of replacements, i.e.  $AF_* = T(AF_n, \mathcal{M}_{n,1}, \mathcal{M}_{n,2})$ . If all replacements  $R_i$  are **S**-safe, then letting  $E = E_1 \cap \dots \cap E_n$  it holds that  $\{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_S(AF)\} = \{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_S(AF_*)\}$ . Moreover, any argument  $A \in E$  is skeptically/credulously justified according to **S** in  $AF$  if and only if it is skeptically/credulously justified according to **S** in  $AF_*$ .

## 7. The relationship between decomposability and transparency

Intuitively, there is a close relationship between decomposability and transparency: if a semantics is decomposable, i.e. the labellings prescribed for an argumentation framework are completely determined by applying the canonical local function to the elements of a partition, then one may expect that replacing a multipole with another one having the same Input/Output behavior has no impact on the invariant set of the replacement. This intuition is confirmed by Theorem 8, showing that decomposability of a semantics **S** is a sufficient condition for strong transparency.

**Theorem 8.** Consider an effect-dictated semantics **S** such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ . If **S** is decomposable w.r.t. a partition selector  $\mathcal{F}$  then **S** is strongly transparent w.r.t.  $\mathcal{F}$ .

While full decomposability is a sufficient condition for strong transparency, it is not necessary. In particular, for a single-status semantics which is top-down decomposable a relaxed form of bottom-up decomposability is sufficient to ensure strong transparency. More specifically, in this case bottom-up decomposability requires the union of local labellings to coincide with the (unique) global labelling. However, just requiring the union of local labellings to be more or equally committed than the global labelling is enough to achieve strong transparency, as shown by Theorem 9.

**Theorem 9.** Let **S** be an effect-dictated single-status semantics such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ . Suppose that **S** is top-down decomposable w.r.t. a partition selector  $\mathcal{F}$  and satisfies the following property: for any argumentation framework  $AF$  and any partition  $\{E, D\} \in \mathcal{F}(AF)$ , letting  $\mathbf{L}$  be the labelling prescribed by **S** for  $AF$ , if  $\mathbf{L}^E \in \mathcal{L}_E$  and  $\mathbf{L}^D \in \mathcal{L}_D$  are two labellings such that  $\mathbf{L}^E \in \mathbf{F}_S(AF \downarrow_E, E^{\text{inp}}, \mathbf{L}^D \downarrow_{E^{\text{inp}}}, E^R)$  and  $\mathbf{L}^D \in \mathbf{F}_S(AF \downarrow_D, D^{\text{inp}}, \mathbf{L}^E \downarrow_{D^{\text{inp}}}, D^R)$ , then  $\mathbf{L} \subseteq \mathbf{L}^E \cup \mathbf{L}^D$ . Then **S** is strongly transparent w.r.t.  $\mathcal{F}$ .

## 8. Analyzing transparency of argumentation semantics

In this section we discuss the transparency properties of the semantics reviewed in Section 2. A synthetic view of the results is given in Table 2 (for all semantics strong transparency turns out to be equivalent to transparency, and any transparency property w.r.t.  $\mathcal{F}_{USCC}$  holds if and only if the same property holds w.r.t.  $\mathcal{F}_{SCC}$ ).

### 8.1. Admissible, complete and stable semantics

As shown in Section 4, admissible, complete and stable semantics satisfy full decomposability: this easily yields strong transparency for such semantics.

**Theorem 10.** Admissible semantics **AD**, complete semantics **CO** and stable semantics **ST** are strongly transparent.

For instance, in Examples 10 and 11 the replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is **S**-legitimate, where  $\mathbf{S} \in \{\mathbf{AD}, \mathbf{CO}, \mathbf{ST}\}$ , therefore it is also **S**-safe, i.e.  $\{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_S(AF_1)\} = \{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_S(AF_2)\}$ . In Examples 12, 13 and 14  $\mathcal{R}$  is **ST**-legitimate, in Example 15 it is contextually **ST**-legitimate, therefore in all cases  $\mathcal{R}$  is **ST**-safe. In particular, in Example 15  $\mathbf{L}_{ST}(AF_1) = \{(E_1, \text{in}), (E_2, \text{out}), (A_1, \text{out}), (A_2, \text{in}), (A_3, \text{out})\}$ ,  $\{(E_1, \text{out}), (E_2, \text{in}), (A_1, \text{out}), (A_2, \text{in}), (A_3, \text{out})\}$  and  $\mathbf{L}_{ST}(AF_2) = \{(E_1, \text{in}), (E_2, \text{out}), (C, \text{out})\}$ ,  $\{(E_1, \text{out}), (E_2, \text{in}), (C, \text{out})\}$ , thus the stable labellings restricted to  $\{E_1, E_2\}$  are  $\{(E_1, \text{out}), (E_2, \text{in})\}$  and  $\{(E_1, \text{in}), (E_2, \text{out})\}$  both in  $AF_1$  and in  $AF_2$ .

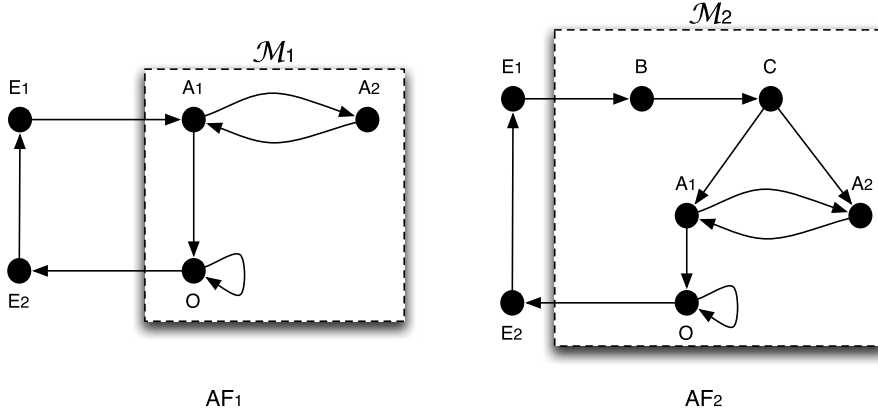


Fig. 14. Preferred semantics is not transparent (Examples 16 and 18).

## 8.2. Grounded semantics

As shown in Section 4.3, grounded semantics is not fully decomposable but only top-down decomposable.

**Theorem 11** shows however that grounded semantics is strongly transparent, building on the result proved in **Theorem 9**.

**Theorem 11.** *Grounded semantics GR is strongly transparent.*

For instance, in Examples 10, 11 and 14 the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is **GR**-legitimate, therefore it is also **GR**-safe. In Example 15 the replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is contextually **GR**-legitimate, since  $\mathcal{L}_{\mathcal{R}}^{\text{GR}} = \{(E_1, \text{undec}), (E_2, \text{undec})\}, \{(E_1, \text{out}), (E_2, \text{in})\}$  and  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **GR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{GR}}$ : as a consequence,  $\mathcal{R}$  is **GR**-safe, as it can be seen by considering that both the grounded labelling of  $AF_1$  and the grounded labelling of  $AF_2$  assign to all arguments the label undec.

## 8.3. Preferred semantics

Like grounded semantics, preferred semantics is top-down decomposable but not fully decomposable. However, differently from grounded semantics, preferred semantics is not transparent, as shown by the following counterexample.

**Example 16.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 14, where  $AF_2 = T(\mathcal{R})$  with  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and the invariant set of the replacement  $\mathcal{R}$  is  $E = \{E_1, E_2\}$ . It turns out that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent, thus  $\mathcal{R}$  is **PR**-legitimate. In fact, for any label  $\mathbf{L}^{\text{in}} \in \mathcal{L}_E$  such that  $E_1$  is labelled in the local function  $F_{\text{PR}}$  prescribes for  $\mathcal{M}_1$  the unique labelling  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , therefore  $\text{PR-eff}_E(\mathcal{M}_1, \mathbf{L}^{\text{in}}) = \{(E_2, \text{undec}), (E_1, \text{in})\}$ , and it prescribes for  $\mathcal{M}_2$  the unique labelling  $\{(B, \text{out}), (C, \text{in}), (A_1, \text{out}), (A_2, \text{out}), (O, \text{undec})\}$ , therefore also  $\text{PR-eff}_E(\mathcal{M}_2, \mathbf{L}^{\text{in}}) = \{(E_2, \text{undec}), (E_1, \text{in})\}$ . For any label  $\mathbf{L}^{\text{out}} \in \mathcal{L}_E$  such that  $E_1$  is labelled out  $F_{\text{PR}}$  prescribes for  $\mathcal{M}_1$  the labellings  $\{(A_1, \text{in}), (A_2, \text{out}), (O, \text{out})\}$  and  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , for  $\mathcal{M}_2$  the labellings  $\{(B, \text{in}), (C, \text{out}), (A_1, \text{in}), (A_2, \text{out}), (O, \text{out})\}$  and  $\{(B, \text{in}), (C, \text{out}), (A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , thus  $\text{PR-eff}_E(\mathcal{M}_1, \mathbf{L}^{\text{out}}) = \text{PR-eff}_E(\mathcal{M}_2, \mathbf{L}^{\text{out}}) = \{(E_2, \text{in}), (E_1, \text{in})\}, \{(E_2, \text{undec}), (E_1, \text{in})\}$ . For any label  $\mathbf{L}^{\text{undec}} \in \mathcal{L}_E$  such that  $E_1$  is labelled undec,  $F_{\text{PR}}$  prescribes for  $\mathcal{M}_1$  the unique labelling  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , and it prescribes for  $\mathcal{M}_2$  the unique labelling  $\{(B, \text{undec}), (C, \text{undec}), (A_1, \text{undec}), (A_2, \text{undec}), (O, \text{undec})\}$ , therefore  $\text{PR-eff}_E(\mathcal{M}_1, \mathbf{L}^{\text{undec}}) = \text{PR-eff}_E(\mathcal{M}_2, \mathbf{L}^{\text{undec}}) = \{(E_2, \text{undec}), (E_1, \text{in})\}$ . However, the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is not **PR**-safe. In fact, the preferred labellings of  $AF_1$  are  $\{(A_1, \text{in}), (A_2, \text{out}), (O, \text{out}), (E_2, \text{in}), (E_1, \text{out})\}$  and  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec}), (E_2, \text{undec}), (E_1, \text{undec})\}$ , while  $\{(B, \text{in}), (C, \text{out}), (A_1, \text{in}), (A_2, \text{out}), (O, \text{out}), (E_2, \text{in}), (E_1, \text{out})\}$  is the only preferred labelling of  $AF_2$ . Note in particular that  $E_2$  is skeptically justified in  $AF_2$  but not in  $AF_1$ . Interestingly enough, considering the application of stable semantics it can be checked that the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is **ST**-legitimate, therefore according to **Theorem 10** it is also **ST**-safe. In fact,  $\mathbf{L}_{\text{ST}}(AF_1) = \{(A_1, \text{in}), (A_2, \text{out}), (O, \text{out}), (E_2, \text{in}), (E_1, \text{out})\}$  and  $\mathbf{L}_{\text{ST}}(AF_2) = \{(B, \text{in}), (C, \text{out}), (A_1, \text{in}), (A_2, \text{out}), (O, \text{out}), (E_2, \text{in}), (E_1, \text{out})\}$ , therefore both in  $AF_1$  and in  $AF_2$  the argument  $E_1$  is labelled out and  $E_2$  is labelled in by all stable labellings.

In the previous example a **PR**-legitimate replacement yields a change in the status assignment of arguments belonging to the invariant set  $E$ , however it can be noted that their credulous justification is preserved, i.e.  $E_2$  is credulously justified both in  $AF_1$  and  $AF_2$ ,  $E_1$  is not credulously justified either in  $AF_1$  or in  $AF_2$ . **Theorem 12** proves that this result holds in general.

**Theorem 12.** For any contextually **PR**-legitimate replacement  $\mathcal{R} = (AF, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , any argument  $A \in E$  is credulously justified according to **PR** in  $AF$  if and only if it is credulously justified according to **PR** in  $T(AF, \mathcal{M}_1, \mathcal{M}_2)$ .

While the obtained result is somewhat weak, as it concerns credulous justification only, it has to be acknowledged that the counterexample against transparency of **PR** (Example 16) is rather tricky. In particular,  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent, but they differ in the following aspect. On the one hand, in  $\mathcal{M}_1$ , the local function  $F_{\mathbf{PR}}$  prescribes for any input labelling  $\mathbf{L}^{undec}$  the unique labelling  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , and with the “more committed” input labelling  $\mathbf{L}^{out} \in \mathcal{L}_E$  it returns (among others) the labelling  $\{(A_1, \text{in}), (A_2, \text{out}), (O, \text{out})\}$  which is not “more committed” than  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ , i.e. it is not the case that  $\{(A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\} \subseteq \{(A_1, \text{in}), (A_2, \text{out}), (O, \text{out})\}$ . On the other hand, in  $\mathcal{M}_2$  both the labellings returned by the local function  $F_{\mathbf{PR}}$  with the input labelling  $\mathbf{L}^{out}$  are “more committed” than the labelling returned by  $F_{\mathbf{PR}}$  with the input labelling  $\mathbf{L}^{undec}$ , i.e. it holds that  $\{(B, \text{undec}), (C, \text{undec}), (A_1, \text{undec}), (A_2, \text{undec}), (O, \text{undec})\} \subseteq \{(B, \text{in}), (C, \text{out}), (A_1, \text{in}), (A_2, \text{out}), (O, \text{out})\}$ ,  $\{(B, \text{undec}), (C, \text{undec}), (A_1, \text{undec}), (A_2, \text{undec}), (O, \text{undec})\} \subseteq \{(B, \text{in}), (C, \text{out}), (A_1, \text{out}), (A_2, \text{in}), (O, \text{undec})\}$ . More generally, we define the notion of *homogeneously equivalent* argumentation multipoles, corresponding to equivalent multipoles that exhibit a sort of mutually regular behavior.

**Definition 34.** Two multipoles  $\mathcal{M}_1 = (AF_1, R_{\text{INP}}^1, R_{\text{OUTP}}^1)$  and  $\mathcal{M}_2 = (AF_2, R_{\text{INP}}^2, R_{\text{OUTP}}^2)$  w.r.t. a set  $E$  are homogeneously **S**-equivalent under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$  iff they are **S**-equivalent under  $\mathcal{L}'$  and the following two symmetric conditions hold:

1. Given  $\mathbf{L}_E^1, \mathbf{L}_E^2 \in \mathcal{L}'$  such that  $\mathbf{L}_E^1 \subseteq \mathbf{L}_E^2$ , if there are two labellings  $\mathbf{L}_1^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  and  $\mathbf{L}_2^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  such that  $\mathbf{L}_1^{D_1} \subseteq \mathbf{L}_2^{D_1}$ , then  $\forall \mathbf{L}_1^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ , there is a labelling  $\mathbf{L}_2^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$  and  $\mathbf{L}_1^{D_2} \subseteq \mathbf{L}_2^{D_2}$ .
2. Given  $\mathbf{L}_E^1, \mathbf{L}_E^2 \in \mathcal{L}'$  such that  $\mathbf{L}_E^1 \subseteq \mathbf{L}_E^2$ , if there are two labellings  $\mathbf{L}_1^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  and  $\mathbf{L}_2^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\mathbf{L}_1^{D_2} \subseteq \mathbf{L}_2^{D_2}$ , then  $\forall \mathbf{L}_1^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  such that  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) = \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ , there is a labelling  $\mathbf{L}_2^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  such that  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) = \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$  and  $\mathbf{L}_1^{D_1} \subseteq \mathbf{L}_2^{D_1}$ .

In Example 16, it can be seen that the argumentation multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , while being **PR**-equivalent, are not homogeneously **PR**-equivalent.

It turns out that strong transparency of preferred semantics is recovered in case of replacements involving homogeneously **PR**-equivalent multipoles.

**Theorem 13.** Any replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are homogeneously **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ , is **PR**-safe.

Given two equivalent multipoles, a sufficient condition for their homogeneous equivalence is that each multipole is “internally homogeneous”, i.e. the labellings prescribed by the local function are related by set-inclusion in a regular way w.r.t. the commitment relation between the input labellings. Definition 35 formalizes this intuition, while the sufficiency result is proved by Lemma 5 and Corollary 1.

**Definition 35.** Consider an argumentation semantics **S**. An argumentation multipole  $\mathcal{M} = (AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  is *internally S-homogeneous* under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$  iff for all labellings  $\mathbf{L}_E^1, \mathbf{L}_E^2 \in \mathcal{L}'$  such that  $\mathbf{L}_E^1 \subseteq \mathbf{L}_E^2$ , it holds that  $\forall \mathbf{L}_1 \in F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}})$ ,  $\forall \mathbf{L}_2 \in F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}})$  such that  $\text{eff}_E(\mathcal{M}^{\text{outp}}, \mathbf{L}_1 \downarrow_{\mathcal{M}^{\text{outp}}}, R_{\text{OUTP}}) \subseteq \text{eff}_E(\mathcal{M}^{\text{outp}}, \mathbf{L}_2 \downarrow_{\mathcal{M}^{\text{outp}}}, R_{\text{OUTP}})$ , there is a labelling  $\mathbf{L}'_2 \in F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}})$  such that  $\text{eff}_E(\mathcal{M}^{\text{outp}}, \mathbf{L}_2 \downarrow_{\mathcal{M}^{\text{outp}}}, R_{\text{OUTP}}) = \text{eff}_E(\mathcal{M}^{\text{outp}}, \mathbf{L}'_2 \downarrow_{\mathcal{M}^{\text{outp}}}, R_{\text{OUTP}})$  and  $\mathbf{L}_1 \subseteq \mathbf{L}'_2$ .

**Lemma 5.** Consider two multipoles  $\mathcal{M}_1 = (AF_1, R_{\text{INP}}^1, R_{\text{OUTP}}^1)$  and  $\mathcal{M}_2 = (AF_2, R_{\text{INP}}^2, R_{\text{OUTP}}^2)$  w.r.t. a set  $E$  which are internally **S**-homogeneous under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . If  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **S**-equivalent under  $\mathcal{L}'$ , then they are homogeneously **S**-equivalent under  $\mathcal{L}'$ .

**Corollary 1.** Any replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$  and both  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are internally **PR**-homogeneous under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ , is **PR**-safe.

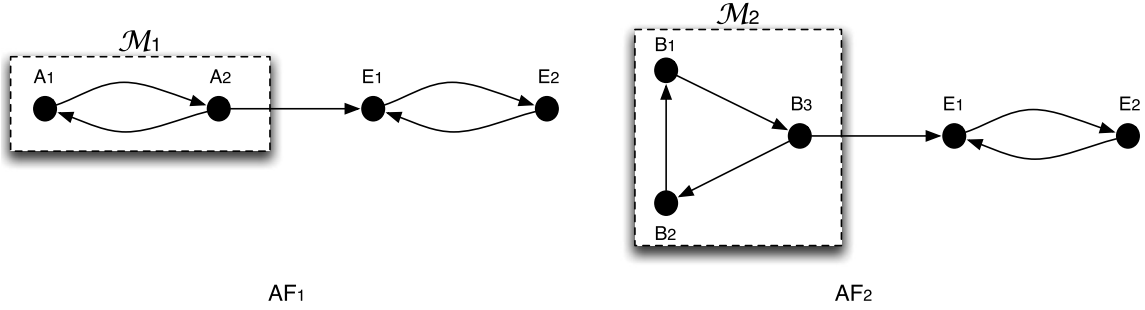


Fig. 15. Ideal semantics is not transparent w.r.t.  $\mathcal{F}_{\text{SCC}}$  (Example 19).

**Example 17.** Consider again the replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  depicted in Fig. 13. As shown in Example 15,  $\mathcal{L}_{\mathcal{R}}^{\text{PR}} = \{(E_1, \text{out}), (E_2, \text{in}), (E_1, \text{in}), (E_2, \text{out})\}$  and  $\mathcal{M}_1, \mathcal{M}_2$  are **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ . Since there are no distinct labellings  $\mathbf{L}_E^1, \mathbf{L}_E^2 \in \mathcal{L}_{\mathcal{R}}^{\text{PR}}$  such that  $\mathbf{L}_E^1 \sqsubseteq \mathbf{L}_E^2$ ,  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are trivially internally **PR**-homogeneous under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ . As a consequence, by Corollary 1 the replacement  $\mathcal{R}$  is **PR**-safe. In fact, there are two preferred labellings in  $AF_1$ , namely  $\{(E_1, \text{in}), (E_2, \text{out}), (A_1, \text{out}), (A_2, \text{in}), (A_3, \text{out})\}$  and  $\{(E_1, \text{out}), (E_2, \text{in}), (A_1, \text{out}), (A_2, \text{in}), (A_3, \text{out})\}$ , while in  $AF_2$  the preferred labellings are  $\{(E_1, \text{in}), (E_2, \text{out}), (C, \text{out})\}$  and  $\{(E_1, \text{out}), (E_2, \text{in}), (C, \text{out})\}$ . Thus, the restriction of the preferred labellings to  $\{E_1, E_2\}$  are  $\{(E_1, \text{out}), (E_2, \text{in})\}$  and  $\{(E_1, \text{in}), (E_2, \text{out})\}$  both in  $AF_1$  and in  $AF_2$ .

Turning to non-arbitrary partitionings, strong transparency of preferred semantics is recovered without additional conditions for replacements involving the union of strongly connected components.

**Theorem 14.** Preferred semantics **PR** is strongly transparent w.r.t.  $\mathcal{F}_{\text{USCC}}$ .

**Example 18.** The multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  shown in Fig. 14 can be safely interchanged if they correspond to the union of strongly connected components. For instance, removing the attack from  $E_2$  to  $E_1$  makes the replacement  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$   $\mathcal{F}_{\text{USCC}}$ -preserving, thus such replacement is safe. In fact, in this case there is a unique preferred labelling in  $AF_1$  and a unique preferred labelling in  $AF_2$ , and in both cases  $E_1$  is labelled *in* and  $E_2$  is labelled *undec*.

It is easy to see that in Examples 10, 11, 12, 13 and 14 the replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is  $\mathcal{F}_{\text{USCC}}$ -preserving and **PR**-legitimate. As a consequence, in all cases the replacement  $\mathcal{R}$  is safe, i.e.  $\{\mathbf{L}_{\downarrow E} \mid \mathbf{L} \in \mathbf{L}_{\text{PR}}(AF_1)\} = \{\mathbf{L}_{\downarrow E} \mid \mathbf{L} \in \mathbf{L}_{\text{PR}}(AF_2)\}$ . Moreover, it can be seen that in all cases the multipoles are internally **PR**-homogeneous, therefore they could be safely interchanged also in the context of non- $\mathcal{F}_{\text{USCC}}$ -preserving replacements.

#### 8.4. Ideal semantics

The transparency properties of ideal semantics mirror the discouraging decomposability properties analyzed in Section 4.4: the following example, inspired by Example 7, shows that ideal semantics is not transparent even w.r.t.  $\mathcal{F}_{\text{SCC}}$ .

**Example 19.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 15, where  $AF_2 = T(\mathcal{R})$  with  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and the invariant set of the replacement  $\mathcal{R}$  is  $E = \{E_1, E_2\}$ . It is easy to see that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **ID**-equivalent, since  $F_{\text{ID}}$  prescribes for  $\mathcal{M}_1$  the labelling  $\{(A_1, \text{undec}), (A_2, \text{undec})\}$  and for  $\mathcal{M}_2$  the labelling  $\{(B_1, \text{undec}), (B_2, \text{undec}), (B_3, \text{undec})\}$ . As a consequence, the replacement  $\mathcal{R}$  is **ID**-legitimate, and it is also easy to see that it is  $\mathcal{F}_{\text{SCC}}$ -preserving. However,  $\mathcal{R}$  is not **ID**-safe, since the ideal labelling of  $AF_1$  leaves all the arguments *undec*, while the ideal labelling of  $AF_2$  is  $\{(B_1, \text{undec}), (B_2, \text{undec}), (B_3, \text{undec}), (E_1, \text{out}), (E_2, \text{in})\}$ .

Transparency is recovered in the (somewhat specific) case of replacements involving multipoles for which  $F_{\text{CO}}$  always prescribes a unique labelling.

**Definition 36.** Consider an argumentation semantics **S**. An argumentation multipole  $\mathcal{M} = (AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  is **S**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$  iff  $\forall \mathbf{L}_E \in \mathcal{L}' \mid F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}}) = 1$ .

The following lemmas prove some specific results holding in the case of **CO**-univocal argumentation multipoles.

**Lemma 6.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . Then  $\forall \mathbf{L}_E \in \mathcal{L}', F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}}) = F_{\mathbf{S}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}})$  for any  $\mathbf{S} \in \{\text{GR}, \text{PR}, \text{ID}, \text{SST}\}$ .

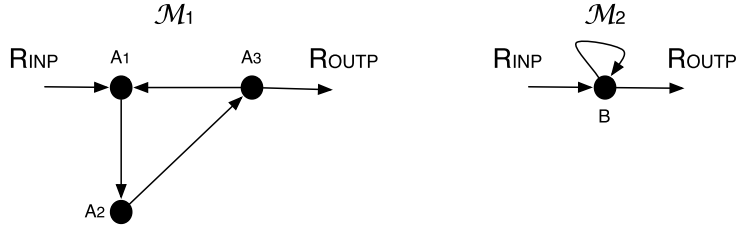


Fig. 16. Two multipoles that can be safely interchanged under ideal semantics (Example 20).

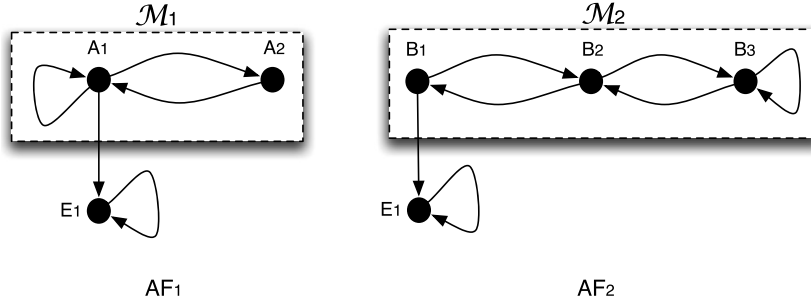


Fig. 17. Semi-stable semantics is not transparent w.r.t.  $\mathcal{F}_{\text{SCC}}$  (Example 21).

**Lemma 7.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ , and let  $\mathbf{L}_E^1, \mathbf{L}_E^2$  be two labellings of  $\mathcal{L}'$  such that  $\mathbf{L}_E^1 \subseteq \mathbf{L}_E^2$ . Then, for any two labellings  $\mathbf{L}_1, \mathbf{L}_2$  such that  $\mathbf{L}_1 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^1 \downarrow \mathcal{M}^{\text{inp}}, R_{\text{INP}})$  and  $\mathbf{L}_2 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow \mathcal{M}^{\text{inp}}, R_{\text{INP}})$ , it holds that  $\mathbf{L}_1 \subseteq \mathbf{L}_2$ .

**Lemma 8.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . Then  $\mathcal{M}$  is internally **PR**-homogeneous under  $\mathcal{L}'$ .

On this basis, Theorem 15 shows that contextually **CO**-legitimate replacements are **ID**-safe if they involve **CO**-univocal multipoles. Note that the theorem requires the involved multipoles to be **CO**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$ . In the light of Lemma 6, this is tantamount to requiring them to be **S**-equivalent for any  $\mathbf{S} \in \{\text{GR}, \text{PR}, \text{ID}, \text{SST}\}$ . We cannot, however, replace  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$  with e.g.  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ , since  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$  may be a strict superset of  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ .

**Theorem 15.** Any contextually **CO**-legitimate replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **CO**-univocal under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$ , is **ID**-safe.

As shown in Section 8.6, the previous theorem applies in particular to acyclic argumentation multipoles, while the next example shows that there are cases of equivalent multipoles containing cycles that can be safely interchanged under ideal semantics.

**Example 20.** It is easy to see that the multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  shown in Fig. 16 are **CO**-equivalent and both of them are **CO**-univocal under any set. Thus, by Theorem 15 they can be safely replaced each other under the ideal semantics, i.e. the replacement maintains the labels assigned by the ideal labelling to the arguments of the invariant set. It is also easy to see that the same holds by replacing the three-length cycles in  $\mathcal{M}_1$  with any odd-length cycle.

### 8.5. Semi-stable semantics

As in the case of ideal semantics, semi-stable semantics inherits from its lack of decomposability properties the inability of guaranteeing safeness of legitimate replacements: the following example, inspired by Examples 8 and 9, shows that semi-stable semantics is not transparent even w.r.t.  $\mathcal{F}_{\text{SCC}}$ .

**Example 21.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 17, where  $AF_2 = T(\mathcal{R})$  with  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and the invariant set of the replacement  $\mathcal{R}$  is  $\{E_1\}$ . It is easy to see that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **SST**-equivalent, since  $F_{\text{SST}}$  prescribes for  $\mathcal{M}_1$  the unique labelling  $\{(A_1, \text{out}), (A_2, \text{in})\}$  and for  $\mathcal{M}_2$  the unique labelling  $\{(B_1, \text{out}), (B_2, \text{in}), (B_3, \text{out})\}$ , thus the effect on  $\{E_1\}$  is  $\{(E_1, \text{in})\}$  in both cases. As a consequence, the replacement  $\mathcal{R}$  is **SST**-legitimate, and it is also easy to see that it is  $\mathcal{F}_{\text{SCC}}$ -preserving. However,  $\mathcal{R}$  is not **SST**-safe, since in  $AF_1$  there is only one semi-stable

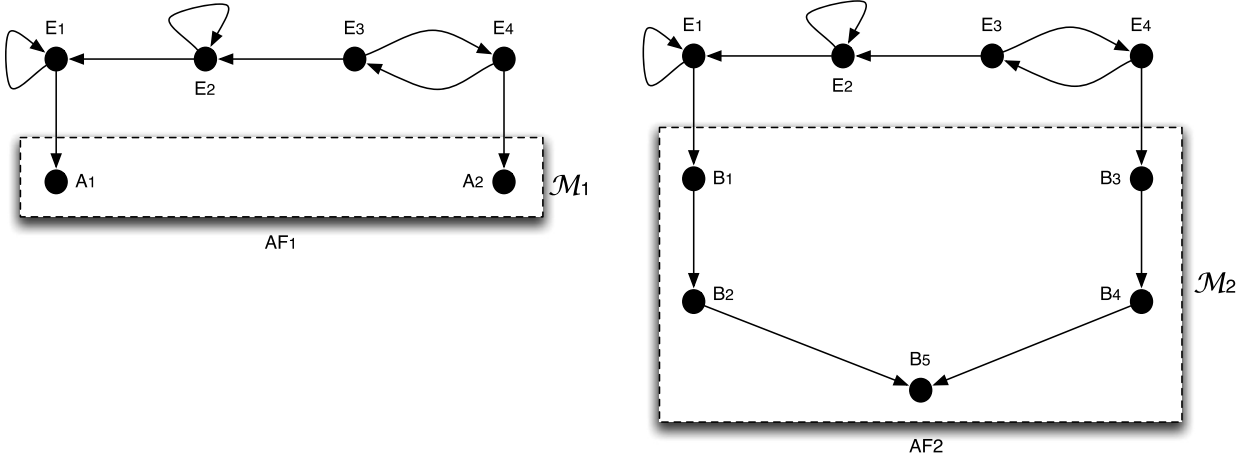


Fig. 18. Semi-stable semantics is not transparent even considering acyclic multipoles (Example 22).

labelling, namely  $\{(A_1, \text{out}), (A_2, \text{in}), (E_1, \text{undec})\}$ , which assigns to  $E_1$  the label *undec*, while there are two semi-stable labellings in  $AF_2$ , namely  $\{(B_1, \text{in}), (B_2, \text{out}), (B_3, \text{undec}), (E_1, \text{out})\}$  and  $\{(B_1, \text{out}), (B_2, \text{in}), (B_3, \text{out}), (E_1, \text{undec})\}$ , which assign to  $E_1$  the label *out* and *undec*, respectively.

### 8.6. The case of acyclic multipoles

It is well-known that an argumentation framework with an acyclic attack relation admits a unique complete labelling which is thus also grounded, preferred, ideal, stable and semi-stable. It is then interesting to specifically consider acyclic multipoles, and to investigate whether they benefit of specific properties as far as replaceability is concerned.

**Definition 37.** A multipole  $\mathcal{M} = (AF, R_{INP}, R_{OUTP})$ , where  $AF = (Ar, att)$ , is *acyclic* if there is no sequence  $A_1, \dots, A_n$  of distinct arguments with  $A_i \in Ar$  such that  $n > 1$ ,  $(A_i, A_{i+1}) \in att$  for  $1 \leq i < n$ , and  $(A_n, A_1) \in att$ .

Note that this definition does not prevent an acyclic multipole to contain self-attacking arguments, i.e. arguments attacking themselves.

The following proposition shows that the property of acyclic frameworks mentioned above can be extended to acyclic multipoles.

**Proposition 14.** An acyclic argumentation multipole  $\mathcal{M} = (AF, R_{INP}, R_{OUTP})$  w.r.t. a set  $E$  is **CO-univocal** under any set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ .

The above result entails that all semantics considered in this paper, with the exception of semi-stable semantics, become strongly transparent in case replacements involve acyclic multipoles. Since admissible, complete, stable and grounded semantics are strongly transparent, it suffices to consider preferred and ideal semantics.

**Proposition 15.** Any contextually **PR-legitimate** (**ID-legitimate**) replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are acyclic, is **PR-safe** (**ID-safe**).

The following example shows that this result cannot be extended to semi-stable semantics, i.e. there are acyclic **SST**-equivalent multipoles that cannot be safely interchanged.

**Example 22.** Consider the argumentation frameworks  $AF_1$  and  $AF_2$  shown in Fig. 18, where  $AF_2 = T(\mathcal{R})$  with  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and the invariant set of  $\mathcal{R}$  is  $\{E_1, E_2, E_3, E_4\}$ . The acyclic multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are trivially **SST**-equivalent, since they do not attack  $E$  (for both of them, the effect on  $E$  includes a unique labelling which assigns to all arguments the label *in*). However, the replacement  $\mathcal{R}$  is not **SST**-safe, since there is a unique semi-stable labelling in  $AF_1$ , namely  $\{(E_1, \text{undec}), (E_2, \text{out}), (E_3, \text{in}), (E_4, \text{out}), (A_1, \text{undec}), (A_2, \text{in})\}$ , while  $AF_2$  admits  $\{(E_1, \text{undec}), (E_2, \text{out}), (E_3, \text{in}), (E_4, \text{out}), (B_1, \text{undec}), (B_2, \text{undec}), (B_3, \text{in}), (B_4, \text{out}), (B_5, \text{undec})\}$  and  $\{(E_1, \text{undec}), (E_2, \text{undec}), (E_3, \text{out}), (E_4, \text{in}), (B_1, \text{undec}), (B_2, \text{undec}), (B_3, \text{out}), (B_4, \text{in}), (B_5, \text{out})\}$  as the two semi-stable labellings. For instance, argument  $E_4$  is assigned the unique label *out* in  $AF_1$  and the labels *in* and *out* in  $AF_2$ .



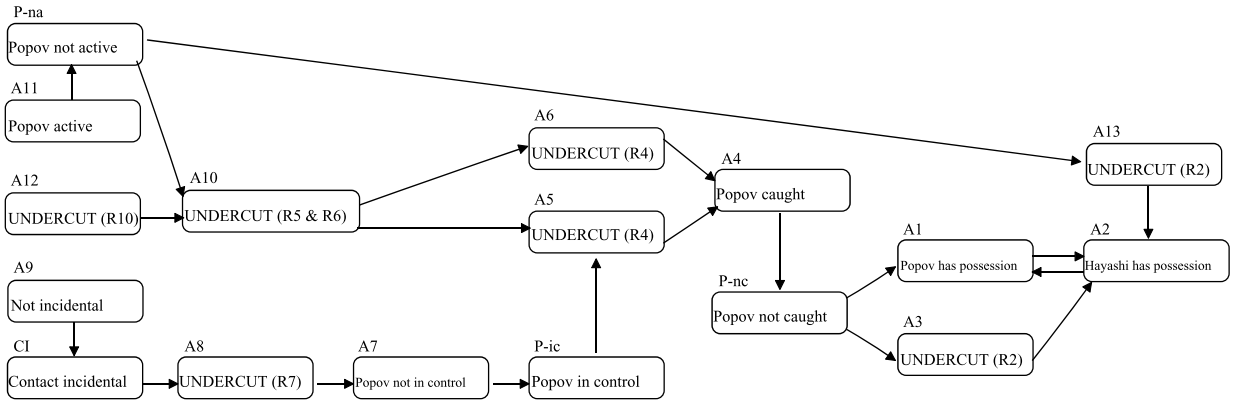


Fig. 19. The argumentation framework  $AF_J$  for the Popov v. Hayashi case from [44].

## 9. Putting modularity at work

As modularity is a very useful and pervasive property, the notions and results introduced in this paper have an open-ended range of applications. In fact, they can be exploited in all contexts, either theoretical or practical, where a non-monolithic approach is appropriate, ranging from the management of dynamics in argumentation to the study of efficient divide-and-conquer algorithms. While an extensive discussion of related works with pointers to future research directions is given in Section 10, in this section we use, as sample case-studies, the tasks of summarization and translation of argumentation frameworks and develop in detail some relevant application examples.

### 9.1. Summarizing argumentation frameworks

In this subsection we illustrate an example of application of the notion of equivalence between argumentation multipoles for the purpose of summarization of argumentation frameworks. In particular we take from the literature two argument-based reconstructions of the court's decision of the Popov v. Hayashi case and show that, in spite of many differences in the details, they can be reduced to a comparable basic structure through considerations based on multipole equivalence.

We borrow a synthetic description of the facts originating the case from [44]. "The case concerned the possession of the baseball which Barry Bonds hit for his record breaking 73rd home run in the 2001 season. Such a ball is very valuable (Mark McGwire's 1998 70th home run ball sold at auction for \$3,000,000). When the ball was struck into the crowd, Popov caught it in the upper part of the webbing of his baseball glove. Such a catch, a snowcone catch because the ball is not fully in the mitt, does not give certainty of retaining control of the ball, particularly since Popov was stretching and may have fallen. However, Popov was not given the chance to complete his catch since, as it entered his glove, he was tackled and thrown to the ground by others trying to secure the ball, which became dislodged from his glove. Hayashi (himself innocent of the attack on Popov), then picked up the ball and put it in his pocket, so securing possession."

Popov then claimed possession of the ball and sued Hayashi. The court finally decided that the ball should be sold and the proceeds divided between the two.

The rather articulated motivations underlying the decision have attracted the attention of researchers and have been the subject of several papers, culminating in a special issue of the *Artificial Intelligence and Law* journal devoted to the modelling of this case [1]. In the following subsections we present the argument-based formalizations provided by Wyner and Bench-Capon [44] and by Prakken [39] respectively. Then we show how the notions and results presented in previous sections can be used to summarize the two formalizations and simplify their comparison.

For the sake of uniformity with the original formalizations, in the following we will sometimes refer to the extension-based rather than the labelling-based approach. In particular, an **S** extension (e.g. the grounded extension) is the set of arguments labelled **in** by an **S** labelling (e.g. the grounded labelling).

#### 9.1.1. The formalization by Wyner and Bench-Capon

In [44] the legal analysis of the case is synthesized by the argumentation framework presented in Fig. 19 (the paper also presents an analysis of the values underlying the final decision using the formalism of value-based argumentation frameworks, which is beyond the scope of the present paper). In the original figure of [44] the boxes representing arguments are labeled with an identifier  $A_x$ , where  $x$  is a number, while a few other boxes have no label and contain a statement corresponding to the conclusion of the argument. In Fig. 19 all arguments have both a label (on top of the box and corresponding to the original one where present) and a text synthesizing their conclusion. Each argument labeled as  $A_x$  derives from the application of a rule with some premises and a conclusion, while the other four arguments are intended to represent default answers to some questions: quoting [44], "if the argument is not defeated, the contrary has not been shown". The conclusion of an argument may correspond to the undercut of some rule. An argument attacks another argument if the

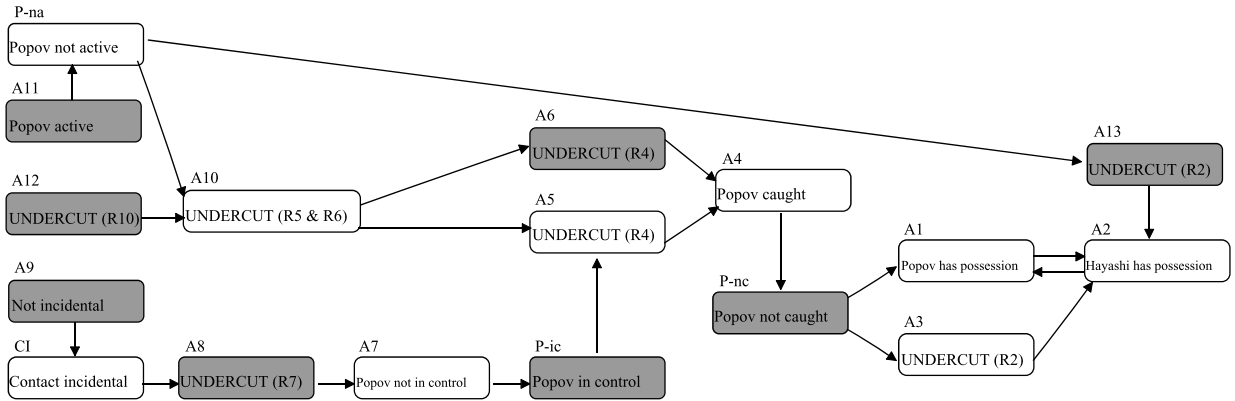


Fig. 20. The only extension of  $AF_J$ .

conclusion of the former contradicts the conclusion or some premise of the latter or undercuts the rule used for its construction. Default arguments can only attack another argument on its premises. Turning to a quick explanation of Fig. 19, we can proceed backwards starting from the mutually attacking arguments A1 and A2, concerning who has possession of the ball. A2 is undercut by A13: the rule that Hayashi has possession of the ball because he retrieved it is not applicable given that Popov was active in catching the ball before Hayashi retrieved it. A13 is attacked by the default argument P-na, which is in turn attacked by A11 based on factual evidence of the snowcone catch. A2 is also undercut by A3, whose premise (by the way, the same as of A1) is that Popov caught the ball before Hayashi. However both A1 and A3 are attacked by the default argument that the ball was not caught by Popov. This is in turn attacked by A4, based on the fact that the ball was in Popov's glove. A4 is undercut by A5 and A6, the former based on the fact that the ball was still in motion, the latter on the fact that Popov was not in control of the ball. Both A5 and A6 are undercut by A10 based on the fact that Popov was active. A10 is hence attacked by the default argument P-na and is also undercut by A12, based on the custom and practice of the stands in baseball. Moreover A5 is attacked by the default argument P-ic, which is attacked by A7 based on the fact that Popov did not retain the ball in the glove. A7 is undercut by A8, based on the fact that Popov lost the ball due to an intentional contact of other people. Finally, A8 is attacked by the default argument C1 which is in turn attacked by A9 based on factual evidence that Popov was assaulted.

It can be seen that for the argumentation framework represented in Fig. 19 the grounded extension is also the only complete, stable, semi-stable, ideal and preferred extension. It consists of the arguments A9, A11, A12, A13, A6, A8, P-ic, P-nc, which are evidenced in grey in Fig. 20. We note that both A1 and A2 are rejected according to any semantics, leaving the issue of the possession of the ball unresolved.

### 9.1.2. The formalization by Prakken

The reconstruction of the case given in [39] adopts ASPIC+, which is essentially a rule-based formalism for the construction of arguments and the identification of their subargument and attack relations. It is worth remarking that the latter takes into account the former: if an argument attacks another argument then it attacks also all its superarguments. In ASPIC+ argument status evaluation follows Dung's approach: an argumentation framework consisting only of the arguments and their attack relations can be derived and then the semantics deemed most appropriate can be applied.

Coming back to Popov and Hayashi, the reconstruction of [39] covers a lot of details concerning argument construction and, as such, is much more articulated than the one of [44] as shown by Figs. 21 and 22 which correspond to the aggregation of five distinct but linked figures included in [39]. Direct subargument relationships are represented by dashed lines ending with a solid dot on the superargument, attack relationships are represented by solid arrows ending on the attacked argument. The text in an argument box essentially gives an idea on its conclusion. Fig. 21 is referred to as the upper part, while Fig. 22 is referred to as the lower part, they are linked only by two subargument relations: VR-MC8 and VR-r1 in Fig. 22 are direct subarguments respectively of EQ and H-hr in Fig. 21.

For a detailed description of the whole reconstruction, which is clearly beyond the scope of the present paper, the reader is referred to [39]. At a general level we can observe that:

- a lot of attention is reserved to issues concerning the validity of rules (sometimes based in turn on the validity of other rules), their adoption and their applicability to the case into question;
- the lower part (Fig. 22) essentially concerns the question whether Popov gained possession of the ball. There are two alternative reasoning lines leading to this conclusion, composed respectively by arguments VR-cs4, P-cc(1), P-ca(1), P-ph(1), P-hp(1), and P-wit, P-cb, VR-cs2, P-cc(2), P-ca(2), P-ph(2), P-hp(2). Both lines are defeated, the former by argument NV-cs4 stating the invalidity of the rule cs4 which is the starting point of the whole line, the latter by argument P-inc stating that Popov's testimony, on which the whole line is based, is not credible.



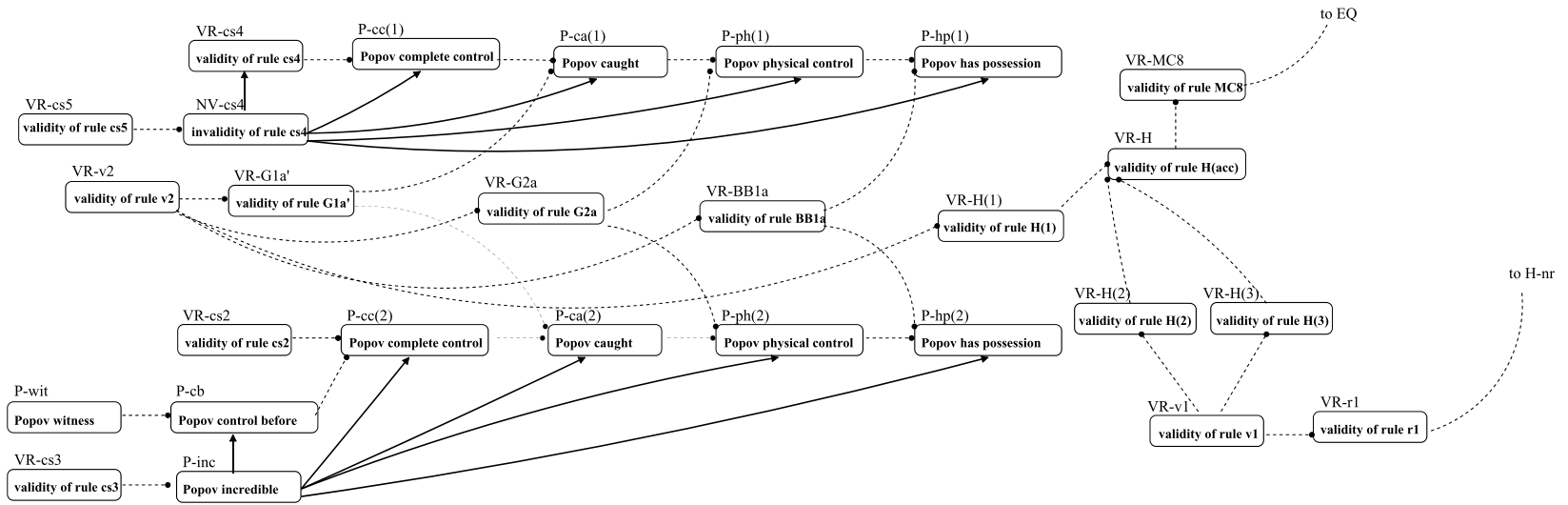


Fig. 22. Lower part of the representation of the Popov v. Hayashi case from [39].

- the upper part (Fig. 21) essentially concerns the action to be taken: three mutually exclusive alternatives (corresponding to the three mutually attacking arguments H-hr, H-nr, and EQ) are considered: Hayashi has to return the ball, Hayashi has not to return the ball, the ball is equally shared. Each of the three arguments is derived through a quite articulated reasoning line. Both H-hr and H-nr are defeated, the former by argument NV-rp, stating that the rule rp is not valid, the latter by argument NA-r4, stating that the rule r4 is not applicable.

If one considers the attack relations only (i.e. focuses on the argumentation framework to be used for argument status evaluation) the picture is simplified, as shown in Fig. 23, since a large number of arguments are neither attacking nor attacked by others. It can be seen that for the argumentation framework represented in Fig. 23 the grounded extension is also the only complete, stable, semi-stable, ideal and preferred extension and consists of the arguments evidenced in grey. We note that of the three arguments corresponding to the possible final decisions both H-hr and H-nr are rejected, while EQ is accepted.

### 9.1.3. Summarizing and comparing the two formalizations

We can now use considerations based on the equivalence properties examined in the previous sections to identify some fundamental similarities between the two reconstructions of the case.

As to the argumentation framework  $AF_J = (Ar, att)$  of Fig. 19, let us start by considering the argumentation multipole  $\mathcal{M}_1 = (AF_J \downarrow_{\{A11, P-na\}}, \emptyset, \{(P-na, A10), (P-na, A13)\})$  with respect to  $E_1 = Ar \setminus \{A11, P-na\}$ . It is rather easy to see that for any labelling  $\mathbf{L}_{E_1}$  of  $E_1$  (actually irrelevant since the multipole does not receive attacks) and for any semantics  $\mathbf{S}$  (all behave the same on such a simple subframework) it holds that  $\mathbf{S}\text{-eff}_{E_1}(\mathcal{M}_1, \mathbf{L}_{E_1}) = \{(A, in) \mid A \in E_1\} = \mathbf{S}\text{-eff}_{E_1}(\mathcal{M}_\emptyset, \mathbf{L}_{E_1})$ .

In other words, the multipole  $\mathcal{M}_1$  is  $\mathbf{S}$ -equivalent to the empty multipole for any semantics  $\mathbf{S}$ . It follows that the replacement  $\mathcal{R} = (AF_J, \mathcal{M}_1, \mathcal{M}_\emptyset)$  is  $\mathbf{S}$ -legitimate. Intuitively this means that the arguments A11 and  $P-na$  can be canceled from  $AF_J$  without any consequence on the evaluation of other arguments, provided that a suitable transparency property holds for  $\mathbf{S}$ . Since both multipoles  $\mathcal{M}_1$  and  $\mathcal{M}_\emptyset$  are acyclic, the results summarized in Table 2 ensure that the replacement is safe for any semantics considered in this paper except semi-stable semantics (by the way, the replacement is safe also for semi-stable semantics, given that in this case its labellings coincide with stable labellings).

Iterating the same kind of reasoning, it can be seen that the following pairs of arguments can progressively (and safely) be cancelled:  $\{A12, A10\}$ ,  $\{A9, CI\}$ ,  $\{A8, A7\}$ ,  $\{P-ic, A5\}$ ,  $\{A6, A4\}$ . In virtue of Proposition 13 we have that we can safely restrict  $AF_J$  to the set of arguments  $E^* = \{A1, A2, A3, A13, P-nc\}$  without affecting the labellings of the arguments in  $E^*$ . This could have been done (in a single, more laborious, step) also showing that the big multipole consisting of the set of arguments  $\{A11, P-na, A12, A10, A9, CI, A8, A7, P-ic, A5, A6, A4\}$  is  $\mathbf{S}$ -equivalent to the empty multipole.

Assuming that the main focus concerns the evaluation of arguments A1 and A2, we can also see that A3 can be suppressed in  $AF_J^* = AF_J \downarrow_{E^*}$ : given the multipole  $\mathcal{M}_2 = (AF_J^* \downarrow_{\{A3, P-nc\}}, \emptyset, \{(P-nc, A1), (A3, A2)\})$  with respect to  $E_2 = \{A1, A2, A13\}$ , it is again easy to see that for any (actually irrelevant) labelling  $\mathbf{L}_{E_2}$  of  $E_2$  and for any semantics  $\mathbf{S}$  it holds that  $\mathbf{S}\text{-eff}_{E_2}(\mathcal{M}_2, \mathbf{L}_{E_2}) = \{(A1, out), (A2, in), (A13, in)\} = \mathbf{S}\text{-eff}_{E_2}(\mathcal{M}_3, \mathbf{L}_{E_2})$  where  $\mathcal{M}_3 \triangleq ((\{P-nc\}, \emptyset), \emptyset, \{(P-nc, A1)\})$ . Using again the fact that both  $\mathcal{M}_2$  and  $\mathcal{M}_3$  are acyclic we get that the replacement is safe, i.e. that A3 can be cancelled.

Summing up, we get the simplified argumentation framework  $AF_J^-$  shown in Fig. 24 which, for any semantics considered in this paper, is equivalent to the original one as far as the evaluation of the remaining arguments is concerned.

Turning now to the argumentation framework  $AF_K$  of Fig. 23, we first note that all the isolated (i.e. both unattacking and unattacked) arguments can be suppressed. This follows from the fact that, for any semantics  $\mathbf{S}$  and for any argumentation framework  $AF_U$  such that  $\mathbf{L}_\mathbf{S}(AF_U) \neq \emptyset$ , given the multipole  $\mathcal{M}_U = (AF_U, \emptyset, \emptyset)$  with respect to any (actually irrelevant) set  $E$ , for any labelling  $\mathbf{L}_E$  of  $E$  it holds that  $\mathbf{S}\text{-eff}_E(\mathcal{M}_U, \mathbf{L}_E) = \{(A, in) \mid A \in E\} = \mathbf{S}\text{-eff}_E(\mathcal{M}_\emptyset, \mathbf{L}_E)$ .

Supposing that the main interest concerns the final decision, i.e. the evaluation of the arguments H-hr, H-nr and EQ, and using the same reasoning as above we can also see that all the arguments concerning the issue of Popov's possession, not attacking nor being attacked by arguments outside the set, can be suppressed.

Then, using a reasoning which is completely analogous to the one applied to the multipole  $\mathcal{M}_2$  above, we can also suppress the arguments H-hp, P-nhp, and Vr-rp, getting finally the argumentation framework  $AF_K^-$  represented in Fig. 25.

Comparing now Figs. 24 and 25 we observe that:

- arguments A1 and A2 in  $AF_J^-$  correspond respectively to arguments H-hr and H-nr in  $AF_K^-$  and have the same status of rejected;
- similarly, we can also say that arguments P-nc and A13 in  $AF_J^-$  correspond respectively to arguments NV-rp and NA-r4 in  $AF_K^-$ ;
- the argument EQ of  $AF_K^-$  has no counterpart in  $AF_J^-$  due to the fact that in [44] the final decision is represented only in the context of the value-based formalization.

Leaving apart EQ, we note therefore a basic structural similarity between the two simplified frameworks: in both reconstructions the arguments corresponding to giving the ball to one of the contendants are rejected due to one main reason. One may then wonder whether the reasons for these rejections are actually the same in the two reconstructions.

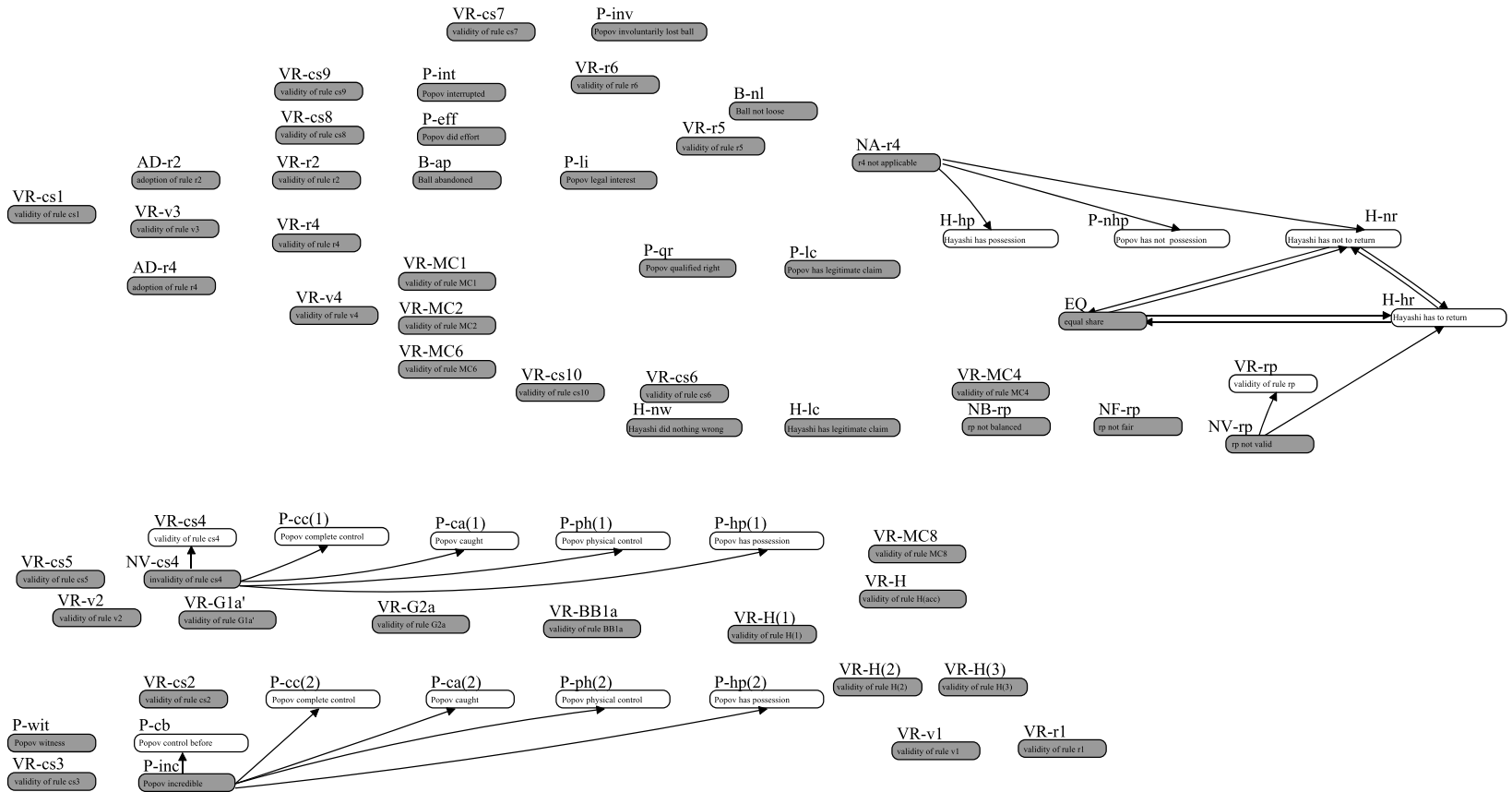


Fig. 23. The representation of the Popov v. Hayashi case from [39] without the subargument relation.



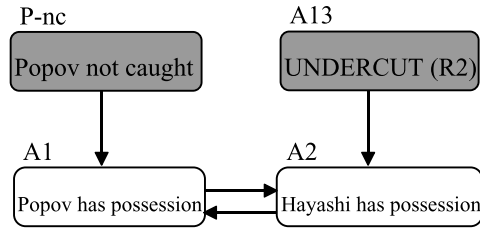


Fig. 24. The argumentation framework  $AF_J^-$  summarizing the reconstruction from [44].

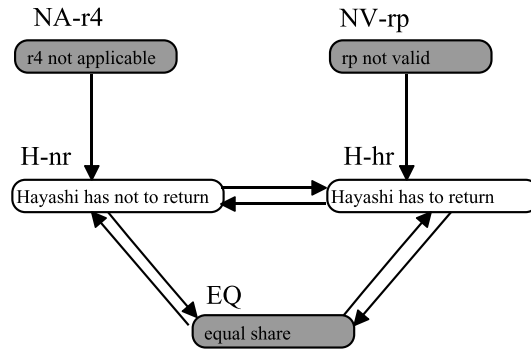


Fig. 25. The argumentation framework  $AF_K^-$  summarizing the reconstruction from [39].

As to the rejection of the decision in favor of Hayashi, in  $AF_J^-$  it is due to the undercut of A2 by A13, which is based on the fact that Popov was “ably and actively engaged in establishing control” of the ball. Similarly, in  $AF_K^-$  the rejection of H-nr is due to the fact that a rule used to derive that Hayashi has possession of the ball, is shown not to be applicable in this case through argument NA-r4, based on the fact that the ball was not loose (due to the previous attempt of Popov) when Hayashi retrieved it.

While basically similar as far as the previous point is concerned, the two reconstructions turn out to be different as to the rejection of the decision in favor of Popov: in  $AF_J^-$  A1 is attacked by P-nc which corresponds to the conclusion that Popov did not catch the ball, thus denying the premise of A1, while in  $AF_K^-$  H-hr is attacked by NV-rp, which concerns the validity of the rule rp. It is interesting to note that in [39] the argument NV-rp is essentially based on the fact that “rule rp does not promote fundamental fairness as regards Popov’s claim” and that, indeed, fairness is the primary value considered in the value-based part of [44] as a justification of the final decision.

Thus the difference arises from the fact that in the formalism adopted in [39] reasoning about values is embedded into arguments that are at the same level as other arguments, while in [44] reasoning about values is carried out in a separate layer. A discussion about the pros and cons of either approach to deal with values is clearly out of the scope of this paper.

To conclude this section we remark that the identification of some basic commonalities and differences between two argument-based reconstructions of a real law case has been greatly simplified by the possibility to summarize frameworks in a general and technically sound way. In this perspective the notion of argumentation multipole and the decomposability and equivalence properties investigated in this paper can be regarded as enabling techniques for the investigation of methods for (possibly automated) analysis, synthesis and comparison of argumentation frameworks.

## 9.2. Translations of argumentation frameworks

Translating an argumentation framework  $AF_1$  into another framework  $AF_2$  such that  $AF_2$  has some desirable features and, at the same time, preserves some specific properties of  $AF_1$  is a generic problem with significant theoretical and practical implications. In particular in [27] the problem of *intertranslatability* is considered, which is defined as follows: “Given an argumentation framework  $F$  and argumentation semantics  $\sigma$  and  $\sigma'$ , find a function  $Tr$  such that the  $\sigma$ -extensions of  $F$  are in certain correspondence to the  $\sigma'$ -extensions of  $Tr(F)$ .” As a matter of fact, in [27] *modularity* is one of the general requirements of a translation procedure, informally stated as “the translation can be done independently for certain parts of the framework”. While this generic notion may have different technical counterparts depending on the kind of translation addressed, our results provide a systematic and sound basis for ensuring modularity in any context where there is an interest in replacing a subframework with a translated counterpart. A broad investigation of this issue is clearly a matter for future work, here we provide two specific examples taken from the literature: the former concerns a subframework replacement considered in the context of the analysis of the properties of weighted argument systems, while the latter concerns the translation (also called *flattening*) of argumentation frameworks with attacks to attacks into “traditional” Dung’s frameworks.

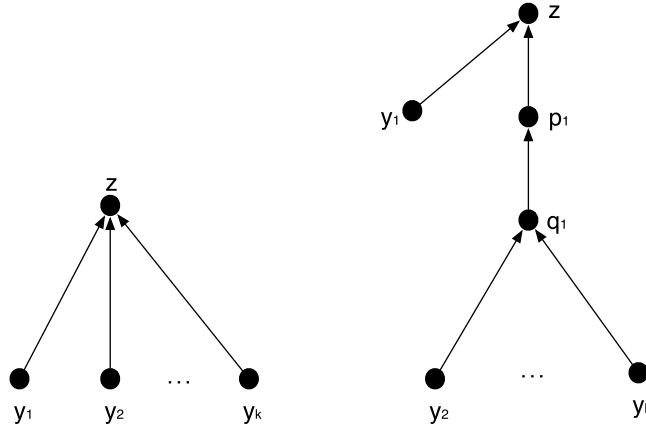


Fig. 26. A translation in the context of WAS.

### 9.2.1. Reducing the attacks involving single arguments under grounded semantics

A weighted argument system (WAS in the following), as defined in [26], is basically an argumentation framework with a numerical weight (actually a non-negative real number) attached to each attack. In the analysis of the computational properties of WASs, it turns out to be convenient to consider a translation from a WAS into another one such that no argument attacks or is attacked by more than 2 arguments and some conditions are satisfied. Leaving apart the aspects of the translation and the conditions involving weights, which are not relevant to the present paper, basically the translation described in [26] involves replacing the subframework consisting of an argument  $z$  receiving more than two attacks (from arguments  $y_1, \dots, y_k$ ) with a subframework with additional arguments  $p_1$  and  $q_1$  where  $z$  receives only two attacks (from  $p_1$  and  $y_1$ ), while  $p_1$  is attacked by  $q_1$  and  $q_1$  is attacked by the arguments  $y_2, \dots, y_k$  (see Fig. 26). The replacement can then be iterated focusing on  $q_1$  and adding  $p_2$  and  $q_2$  until  $q_{k-2}$  is only attacked by  $y_{k-1}$  and  $y_k$ . The claim (proved in [26] as part of Lemma 1) is that the grounded extension of the original framework is the same as the grounded extension of the framework resulting from the replacements mentioned above. Note that Lemma 1 of [26] concerns an arbitrary WAS, i.e. its hypotheses do not put any restriction on other attacks present in the original framework. In particular, as explicitly remarked in [26], there can be attacks between some of the attackers of  $z$ , but also (not explicitly remarked in [26])  $z$  might counterattack some of its attackers or there could be longer loops involving  $z$ , some of its attackers and possibly other arguments in the framework.

Given these remarks, the proof of Lemma 1 provided in [26] is not completely satisfactory: it consists in local considerations on the arguments involved in the replacement described above without dealing with possible effects involving other arguments in the framework. The absence of these effects, however, cannot be taken for granted. To give an example, when considering (to contradiction) a generic argument  $x$  included in the grounded extension of the original framework but not in the grounded extension of the translated framework it is stated that this implies that there is an attacker  $u$  of  $x$  in the translated framework such that  $(u, x)$  was not an attack in the original framework. This immediately leads to identify  $x$  as  $z$  and  $u$  as  $p_1$  and to apply only local considerations. However, in general, an argument might be excluded from the grounded extension not just because it has an additional attacker but also because one of its attackers has a different justification state in the new framework. In a sense, the proof of Lemma 1 of [26] seems to implicitly assume the property of transparency of grounded semantics (which, of course is not obvious *per se*) and (partially) shows a sort of local equivalence of the original fragment and of its translated counterpart.

Actually, the result of Lemma 1 of [26] is valid and this can be shown in a relatively straightforward way using the results of the present paper. First, given that grounded semantics is strongly transparent, to obtain the result it is sufficient to show that the translation step depicted in Fig. 26 involves the replacement of an argumentation multipole with another one which is Input/Output **GR**-equivalent. The fact that the translation may involve several such steps is then covered by the result of Proposition 13.

As to the identification of the equivalent multipoles  $\mathcal{M}_1 = (AF^1, R_{INP}^1, R_{OUTP}^1)$  and  $\mathcal{M}_2 = (AF^2, R_{INP}^2, R_{OUTP}^2)$ , observe that the basic idea consists in replacing the argument  $z$  with the attack chain composed by the three arguments  $p_1$ ,  $q_1$ , and  $z$  itself within an arbitrary argumentation framework  $AF = (Ar, att)$  where  $\{y_1, \dots, y_k\}$  is the set of attackers of  $z$  with  $k > 2$ . Then  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are defined with respect to the same invariant set  $E = Ar \setminus \{z\}$ , and the relevant frameworks are  $AF_1 = (\{z\}, \emptyset)$  and  $AF_2 = (\{p_1, q_1, z\}, \{(q_1, p_1), (p_1, z)\})$ . Moreover  $\mathcal{M}_1$  and  $\mathcal{M}_2$  have the same output relation:  $R_{OUTP}^1 = R_{OUTP}^2 = att \cap (\{z\} \times Ar)$ , while they differ in the input relation:  $R_{INP}^1 = \{(y_i, z) \mid 1 \leq i \leq k\}$ ;  $R_{INP}^2 = \{(y_1, z)\} \cup \{(y_i, q_1) \mid 2 \leq i \leq k\}$ . We have now to show that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **GR**-equivalent, i.e. that for any labelling  $\mathbf{L}_E \in \mathfrak{L}_E$ ,  $\mathbf{GR}\text{-eff}_E(\mathcal{M}_1, \mathbf{L}_E) = \mathbf{GR}\text{-eff}_E(\mathcal{M}_2, \mathbf{L}_E)$ , which, recalling Definition 29, amounts to show that  $\{\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{OUTP}^1) \mid \mathbf{L} \in F_{\mathbf{GR}}(AF^1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{INP}^1)\} = \{\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{OUTP}^2) \mid \mathbf{L} \in F_{\mathbf{GR}}(AF^2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{INP}^2)\}$ .

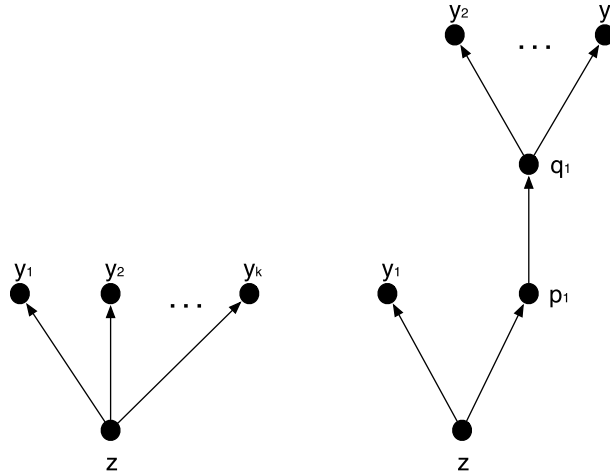


Fig. 27. Another translation in the context of WAS.

First note that, since  $\mathcal{M}_1^{\text{outp}} = \mathcal{M}_2^{\text{outp}} = \{z\}$  and  $R_{OUTP}^1 = R_{OUTP}^2$ , both  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{OUTP}^1)$  and  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{OUTP}^2)$  are totally determined by the label assigned to  $z$  by  $F_{GR}$  given the labelling of the arguments in the input set  $\{y_1, \dots, y_k\}$  (which is the same for both multipoles).

Now it is easy to see that the label assigned to  $z$  is the same for any labelling of the arguments  $\{y_1, \dots, y_k\}$  considering three basic cases: i)  $\exists y_i \in \{y_1, \dots, y_k\} : \text{Lab}(y_i) = \text{in}$ ; ii)  $\forall y_i \in \{y_1, \dots, y_k\} : \text{Lab}(y_i) = \text{out}$ ; iii)  $\nexists y_i \in \{y_1, \dots, y_k\} : \text{Lab}(y_i) = \text{in} \wedge \exists y_i \in \{y_1, \dots, y_k\} : \text{Lab}(y_i) = \text{undec}$ .

In the case i), clearly  $z$  is assigned the label *out* by  $F_{GR}$  in  $\mathcal{M}_1$  and this also holds in  $\mathcal{M}_2$  since either  $z$  is attacked directly by an argument labelled *in* (if  $\text{Lab}(y_1) = \text{in}$ ) or, if this is not the case, necessarily  $\exists y_i \in \{y_2, \dots, y_k\} : \text{Lab}(y_i) = \text{in}$  and then  $\text{Lab}(q_1) = \text{out}$ ,  $\text{Lab}(p_1) = \text{in}$ ,  $\text{Lab}(z) = \text{out}$ .

In the case ii), clearly  $z$  is assigned the label *in* by  $F_{GR}$  in  $\mathcal{M}_1$  and this also holds in  $\mathcal{M}_2$ : given  $\forall y_i \in \{y_1, \dots, y_k\} : \text{Lab}(y_i) = \text{out}$  it follows  $\text{Lab}(q_1) = \text{in}$ ,  $\text{Lab}(p_1) = \text{out}$  and then both attackers ( $y_1$  and  $p_1$ ) of  $z$  are labelled *out* and  $z$  is labelled *in*.

In the case iii), clearly  $z$  is assigned the label *undec* by  $F_{GR}$  in  $\mathcal{M}_1$ . As to  $\mathcal{M}_2$ , first note that  $y_1$  is either labelled *undec* or *out* (in the latter case necessarily  $\exists y_i \in \{y_2, \dots, y_k\} : \text{Lab}(y_i) = \text{undec}$ ). Moreover,  $q_1$  is either labelled *in* or *undec* and consequently  $p_1$  is labelled *out* or *undec* (both are necessarily *undec* if  $\text{Lab}(y_1) = \text{out}$ ). Summing up,  $z$  is either attacked by two arguments labelled *undec* or by one labelled *undec* and one labelled *out* and hence is labelled *undec* by  $F_{GR}$ , as required.

A similar reasoning applies to the case where an argument attacks more than two other arguments, using the replacement sketched in Fig. 27.

### 9.2.2. Flattening attacks to attacks

In recent years several extensions of Dung's framework encompassing attacks to attacks have been considered, like the *EAF* (Extended Argumentation Framework) formalism [35], mainly designed for the purpose of preference modelling, and the more general (as, differently from EAF, they allow unlimited recursion of attacks on attacks) *AFRA* (Argumentation Framework with Recursive Attacks) [4,5] and *HAF* (Higher Level Argumentation Framework) [29].

For the sake of keeping the example compact, we focus here on the *EAF* formalism whose definition (taken from [35]) is given below.

**Definition 38.** An Extended Argumentation Framework (EAF) is a tuple  $(\text{Args}, \mathcal{R}, \mathcal{D})$  such that  $\text{Args}$  is a set of arguments and:

- $\mathcal{R} \subseteq \text{Args} \times \text{Args}$
- $\mathcal{D} \subseteq \text{Args} \times \mathcal{R}$
- if  $(X, (Y, Z)), (X', (Z, Y)) \in \mathcal{D}$  then  $(X, X'), (X', X) \in \mathcal{R}$ .

As typical in any kind of extension of Dung's framework, there is an interest in defining a translation procedure from the extended formalism to the basic one. This is useful for several purposes, including the opportunity to reuse or adapt, in the extended context, the large corpus of theoretical results available in Dung's framework, in particular as far as computational complexity is concerned.

In the case of attacks to attacks, as to our knowledge, two main translation procedures have been proposed in the literature. The first procedure (considered with some slight variants in [36,29,18]) involves replacing an attacked attack with

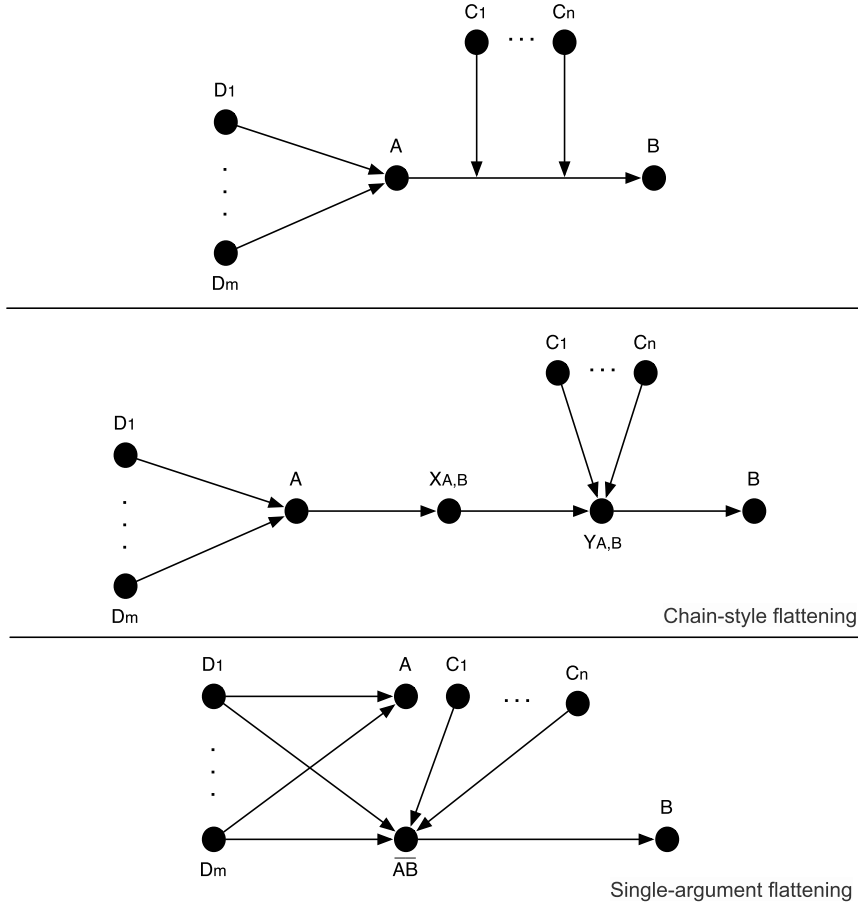


Fig. 28. The two translation procedures of attacks to attacks.

an attack chain consisting in two additional arguments, then every attack towards the replaced attack becomes an attack towards the second additional argument. The second procedure (considered in [4,5]) involves replacing an attack with a single new argument, with a proper rearrangement of the incoming and outgoing attacks involving it. We present in the following the formal definition of these procedures tailored to the case of EAF.

**Definition 39.** Let  $\Gamma = (Args, \mathcal{R}, \mathcal{D})$  be an EAF and let us define  $\mathcal{D}^{\rightarrow}(\Gamma) = \{(A, B) | \mathcal{D} \cap (Args \times \{(A, B)\}) \neq \emptyset\}$  i.e. the set of attacks receiving at least an attack according to the relation  $\mathcal{D}$ .

- The *chain-style flattening* of  $\Gamma$  is the argumentation framework  $AF_c^{\Gamma} = (Args_c, att_c)$  where  $Args_c = Args \cup \{X_{A,B}, Y_{A,B} | (A, B) \in \mathcal{D}^{\rightarrow}(\Gamma)\}$  and  $att_c = \mathcal{R} \cup \{(A, X_{A,B}), (X_{A,B}, Y_{A,B}), (Y_{A,B}, B) | (A, B) \in \mathcal{D}^{\rightarrow}(\Gamma)\} \cup \{(C, Y_{A,B}) | (C, (A, B)) \in \mathcal{D}\}$
- The *single-argument flattening* of  $\Gamma$  is the argumentation framework  $AF_{sa}^{\Gamma} = (Args_{sa}, att_{sa})$  where  $Args_{sa} = Args \cup \{\overline{AB} | (A, B) \in \mathcal{D}^{\rightarrow}(\Gamma)\}$  and  $att_{sa} = (\mathcal{R} \setminus \{(A, B) | (A, B) \in \mathcal{D}^{\rightarrow}(\Gamma)\}) \cup \{(\overline{AB}, B) | (A, B) \in \mathcal{D}^{\rightarrow}(\Gamma)\} \cup \{(D, \overline{AB}) | (D, A) \in \mathcal{R}\} \cup \{(C, \overline{AB}) | (C, (A, B)) \in \mathcal{D}\}$ .

In words, in chain-style flattening two arguments  $X_{A,B}$  and  $Y_{A,B}$  are added in replacement of every attacked attack  $(A, B)$  (with  $A, X_{A,B}, Y_{A,B}, B$  forming an attack chain) and the arguments attacking  $(A, B)$  according to the relation  $\mathcal{D}$  of  $\Gamma$  attack  $Y_{A,B}$  in  $AF_c$  (while the attacks between arguments in  $\mathcal{R}$  remain the same). In single-argument flattening every attacked attack  $(A, B)$  is replaced by a single argument  $\overline{AB}$  which attacks  $B$  (instead of  $A$ ) and is attacked by all attackers of  $A$  in  $\mathcal{R}$  and by all attackers of  $(A, B)$  in  $\mathcal{D}$ .

The two translation procedures are illustrated in Fig. 28.

Of course one may wonder whether the operational differences in the two flattening procedures give rise to any actual difference in the final outcome (i.e. in the justification status of the arguments originally included in  $\Gamma$ ) or, indeed, the two flattened frameworks treat the arguments originally included in  $\Gamma$  in the same way, showing that the two procedures, different as they are, basically capture the same intuition.

**Table 3**Contextually **CO**-legitimate replacement of  $\mathcal{M}_c$  with  $\mathcal{M}_{sa}$ .

$\{D_1, \dots, D_m\}$	$\{C_1, \dots, C_n\}$	$Lab(A)$	$Lab(X_{A,B})$	$Lab(Y_{A,B})$	$Lab\bar{A}\bar{B}$
$\exists in$	$\exists in$	out	in	out	out
$\exists in$	$\forall out$	out	in	out	out
$\exists in$	$\nexists in \wedge \exists undec$	out	in	out	out
$\forall out$	$\exists in$	in	out	out	out
$\forall out$	$\forall out$	in	out	in	in
$\forall out$	$\nexists in \wedge \exists undec$	in	out	undec	undec
$\nexists in \wedge \exists undec$	$\exists in$	undec	undec	out	out
$\nexists in \wedge \exists undec$	$\forall out$	undec	undec	undec	undec
$\nexists in \wedge \exists undec$	$\nexists in \wedge \exists undec$	undec	undec	undec	undec

To answer this question first observe that, given an EAF  $\Gamma = (Args, \mathcal{R}, \mathcal{D})$ , both the argumentation frameworks  $AF_c^\Gamma$  and  $AF_{sa}^\Gamma$  include all the original arguments  $Args$  and that they locally differ in correspondence of the additional arguments used to represent the elements of  $\mathcal{D} \rightarrow (\Gamma)$ . So  $AF_{sa}^\Gamma$  can be obtained from  $AF_c^\Gamma$  (and vice versa) through the replacements of a (possibly quite articulated) multipole  $\mathcal{M}_c$  with another multipole  $\mathcal{M}_{sa}$ , both referring to the same set of invariant arguments  $Args$ . Thus, answering the question amounts to analyzing the safeness of this replacement, based in turn on the equivalence between these multipoles.

First, observe that the multipoles  $\mathcal{M}_c$  and  $\mathcal{M}_{sa}$  consist of the union of  $|\mathcal{D} \rightarrow (\Gamma)|$  disjoint and non-interacting “submultipoles” each having the form illustrated in Fig. 28. In virtue of Proposition 13, we can then consider a sequence of (similar) replacements leading from  $AF_{sa}^\Gamma$  to  $AF_c^\Gamma$  and, to ensure that the whole sequence of replacements is safe (as far as the arguments  $Args$  are concerned), it is sufficient to show that each single step is safe, i.e. to analyze equivalence between the two multipoles representing the translation of a single attack to attack.

To this purpose, referring again to Fig. 28, we can identify the multipoles  $\mathcal{M}_c = (AF^c, R_{INP}^c, R_{OUTP}^c)$  with  $AF^c = (\{X_{A,B}, Y_{A,B}\}, \{X_{A,B}, Y_{A,B}\})$ ,  $R_{INP}^c = \{(A, X_{A,B})\} \cup \{(C_1, Y_{A,B}), \dots, (C_n, Y_{A,B})\}$ ,  $R_{OUTP}^c = \{(Y_{A,B}, B)\}$ , and  $\mathcal{M}_{sa} = (AF^{sa}, R_{INP}^{sa}, R_{OUTP}^{sa})$  with  $AF^{sa} = (\{\bar{A}\bar{B}\}, \emptyset)$ ,  $R_{INP}^{sa} = \{(D_1, \bar{A}\bar{B}), \dots, (D_m, \bar{A}\bar{B})\} \cup \{(C_1, \bar{A}\bar{B}), \dots, (C_n, \bar{A}\bar{B})\}$ , and  $R_{OUTP}^{sa} = \{(\bar{A}\bar{B}, B)\}$ .

It is immediate to observe that the replacement of  $\mathcal{M}_c$  with  $\mathcal{M}_{sa}$  (or vice versa) is in general not legitimate: for instance,  $\mathcal{M}_c$  and  $\mathcal{M}_{sa}$  are in general not equivalent if one considers a labelling  $Lab$  such that  $Lab(D_1) = in$  and  $Lab(A) = in$ . However, this labelling is clearly illegal in a context where  $D_1$  attacks  $A$ . More generally, the labels of arguments  $D_1, \dots, D_m$  completely determine the label of  $A$ , thus one may check whether the replacement is contextually legitimate. So, for any semantics  $\mathbf{S}$ , we are interested in showing that  $\mathcal{M}_c$  and  $\mathcal{M}_{sa}$  are  $\mathbf{S}$ -equivalent with respect to  $\mathcal{L}_{\mathcal{R}}^{\mathbf{S}} \triangleq \{F_{\mathbf{S}}(AF \downarrow_{Args}, \mathcal{M}_c^{outp}, \mathbf{L}_c, R_{OUTP}^c) \mid \mathbf{L}_c \in \mathcal{L}_{\mathcal{M}_c^{outp}}\} \cup \{F_{\mathbf{S}}(AF \downarrow_{Args}, \mathcal{M}_{sa}^{outp}, \mathbf{L}_{sa}, R_{OUTP}^{sa}) \mid \mathbf{L}_{sa} \in \mathcal{L}_{\mathcal{M}_{sa}^{outp}}\}$ .

Observe that since the output relation of both  $\mathcal{M}_c$  and  $\mathcal{M}_{sa}$  consists of a single attack (arising from  $Y_{A,B}$  and  $\bar{A}\bar{B}$  respectively) the two multipoles are equivalent if and only if the labels assigned to  $Y_{A,B}$  and  $\bar{A}\bar{B}$  are the same for any labelling in  $\mathcal{L}_{\mathcal{R}}^{\mathbf{S}}$ . Moreover, we focus on **CO**-equivalence of multipoles without loss of generality: in fact, it turns out (and it is easy to see) that, in any case, for both  $AF^c$  and  $AF^{sa}$ , the local function of complete semantics prescribes exactly one complete labelling, which implies that this is the only labelling also for all the other complete-compatible semantics considered in this paper.

A remark is now in order concerning the labellings of the arguments outside the multipoles, i.e.  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}}$  (note that  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}} \supseteq \mathcal{L}_{\mathcal{R}}^{\mathbf{S}}$  for any complete-compatible semantics  $\mathbf{S}$  considered in this paper). As we are interested in proving an equivalence result whatever the remaining part of the framework is (in addition to the arguments depicted in Fig. 28), the set of labellings  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}}$  cannot be precisely characterized as it also depends on the (unspecified) remaining part of the framework. As a consequence, we prove a slightly stronger result, considering any labelling in the set  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}}$  consisting of the labellings compatible with the attacks from the arguments  $D_i$  to the argument  $A$ . Since  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}} \supseteq \mathcal{L}_{\mathcal{R}}^{\mathbf{CO}}$  this implies the desired equivalence result.

Now, the examination of labellings in  $\mathcal{L}_{\mathcal{R}}^{\mathbf{CO}}$  can be carried out considering nine cases, i.e., all possible combination, for the sets  $\{D_1, \dots, D_m\}$  and  $\{C_1, \dots, C_n\}$ , of three basic cases: i) there is an argument labelled  $in$ ; ii) all arguments are labelled  $out$ ; iii) otherwise (note in particular that  $Lab(A)$  is determined by the labelling of the set  $\{D_1, \dots, D_m\}$  in both multipoles). As all cases are rather simple, for the sake of compactness we synthesize the analysis in Table 3 rather than providing trivial and verbose explanations: by inspection of the last two columns it appears that  $Lab(Y_{A,B}) = Lab(\bar{A}\bar{B})$  in all cases, as desired.

We have thus proved that the replacement of the considered multipoles is contextually **S**-legitimate for any complete-compatible semantics  $\mathbf{S}$  considered in this paper. Then, the replacement is safe for any such semantics  $\mathbf{S}$  which is strongly transparent with respect to these multipoles. Given that the multipoles are acyclic, from the results recalled in Table 2 it follows that the replacement is guaranteed to be safe for all semantics considered in this paper, but **SST**, for which the answer is negative in general and the question is open for this specific case.

The lesson learned is twofold: first, we have given a substantial formal confirmation of the intuition that the two translation procedures are equivalent as far as the “external effects” are concerned for a comprehensive set of semantics, second we have seen however that even “simple” and basically correct intuitions require a careful semantics-specific scrutiny which may point out specific exceptions or critical issues (like for semi-stable semantics in our case).

## 10. Related works

As mentioned in Section 1, the work presented in this paper has connections with three main (and non-disjoint) topics in the area of computational argumentation namely:

- local evaluation in argumentation semantics;
- argumentation dynamics;
- equivalence and interchangeability between argumentation frameworks.

We discuss the relationships with the relevant literature orderly in the following subsections.

### 10.1. Local evaluation in argumentation semantics

As to our knowledge the first analyses of semantics' properties exploitable for the purpose of local evaluation in the literature are provided by the work on SCC-recursiveness [10] and the notion of directionality introduced in [8].

Starting from the latter, in a nutshell a semantics is directional when it is guaranteed that, as far as extensions are concerned, a part of the framework which does not receive attacks from the rest of the framework is unaffected by the rest of the framework itself. Letting  $U$  be a set of arguments not receiving attacks from arguments not in  $U$ , this means that the same results (i.e. the same set of local extensions) are obtained either by computing the global extensions and then intersecting them with  $U$ , or by directly computing the extensions of the restricted framework consisting of the arguments in  $U$  and of the attacks among them.

Directionality allows for local computation when the results one is interested in can be obtained by focusing on an unattacked set, but has no embedded notion of progressive construction: it simply prescribes a relation of inclusion between the local extensions and the global ones. As such it is poorly related with the properties of semantics decomposability and transparency. To give some examples, stable semantics, which is not directional, is fully decomposable (and hence strongly transparent) while semi-stable semantics (which is non directional too) lacks any form of decomposability and transparency. Admissible and complete semantics are directional, fully decomposable and strongly transparent, while ideal semantics (which is directional too) lacks any form of decomposability and satisfies only a very weak form of transparency. To complete the picture, recall that grounded and preferred semantics (which feature intermediate properties) are directional too.

The notion of SCC-recursiveness has closer relationships with the present work, as already evidenced by the fact that we considered partition selectors based on the notion of SCC. Basically, the SCC-recursive scheme provides a general method to build the global extensions prescribed by a semantics by proceeding progressively following the (partial) order among SCCs induced by the attack relation (recall that the graph obtained by considering each SCC as a single node is acyclic). The SCC-recursive scheme applies to each SCC a semantics-specific *base function* and then prescribes how to “propagate the effects” of the choices made in the previous SCCs to the subsequent ones before applying in turn the base function to them. As such, SCC-recursiveness directly implies the property of semantics decomposability with respect to the selector  $\mathcal{F}_{SCC}$ .

Five of the semantics we have considered in this paper are SCC-recursive (namely admissible, complete, stable, grounded, and preferred semantics), and indeed we have proved that all of them feature stronger decomposability properties than the one implied by SCC-recursiveness. Moreover, the notion of local function introduced in this paper can be seen as a generalization of the notion of base function in the SCC-recursive scheme.

Drawing a more detailed analysis of the relationships of SCC-recursiveness with decomposability and transparency properties is an interesting line of future work. As a first note in this direction, we can observe that the two semantics lacking SCC-recursiveness considered in this paper (namely semi-stable and ideal semantics) lack also any decomposability property.

In [40] the problem of combining local evaluations is addressed in a multi-agent scenario context where each agent owns a part of the framework and may locally adopt a different semantics. This gives rise to the notion of multi-sorted argumentation framework where a global argumentation framework is regarded as composed of a set of interacting *cells*, each associated with a (possibly) different semantics. In this context, the investigation in [40] follows a sort of top-down approach: given a (global) set of arguments  $S$ , it addresses the problem of checking whether  $S$  is an extension of the multi-sorted framework, according to local evaluations carried out for each cell. Basically, the definition of local evaluation at the cell level, directly reuses notions taken from the SCC-recursive scheme, as explicitly stated in [40]: the acceptance functions used at the cell level (Definition 5 in [40]) correspond to the base functions of the SCC-recursive scheme, while the notions of subframework and qualified arguments of a subframework (Definitions 7 and 8 in [40]) also have a direct correspondence with key technical elements of the SCC-recursive scheme (respectively with  $AF \downarrow_{UP_{AF}(S,E)}$  and  $U_{AF}(S,E)$  in Definition 20 of [10]). Thus, in a sense, the work of [40] reuses some of the main notions of the SCC-recursive scheme by applying them into two important directions of generalization: considering arbitrary (rather than SCC-based) partitions of the framework and allowing heterogeneous local evaluations. However, the direct reuse of notions specifically conceived in the context of the SCC-recursive scheme limits the possibility to fully encompass situations of mutual interaction and cyclic dependence between cells, which are impossible in the case of SCCs but are possible with arbitrary partitions. The present work addresses the study of homogeneous local evaluations for arbitrary partitions of an argumentation framework by introducing novel notions to capture the more complex interactions between subframeworks arising in this context.



Extending the results presented in this paper to the case of heterogeneous local evaluations is an important direction of future work.

It has also to be mentioned that some results concerning the use of the same semantics (or of semantics with common properties) in all cells are provided in [40] (in particular the notion of *Uniform Case Extension Equivalence* in Definition 10 of [40] roughly corresponds to our notion of semantics decomposability). These results are not directly comparable with ours, due to the different modeling of the interactions between subframeworks mentioned previously. For instance, in Example 5 of [40] a counterexample is given disproving (a sort of) top-down decomposability of grounded semantics in multi-sorted frameworks, while in our context grounded semantics is actually top-down decomposable.

In the notion of conditional acceptance function introduced in [19], basically the acceptance function, corresponding to a given semantics, accepts as input not only an argumentation framework but also an (externally imposed) condition, which corresponds to the set of possible labellings of the framework. In other words, the acceptance function is constrained to produce a set of labellings which is a subset of the given condition. This expresses some form of external influence on argument evaluation, and in this sense could be related to our notion of argumentation framework with input. However, it is based on a rather different intuition, since it expresses a constraint on the labels of all arguments, independently of any attack relation coming from outside, while in our approach external influences manifest themselves through attack relations involving a well-identified set of arguments in the conditioned framework. In [19] the generic notion of conditional acceptance function is instantiated only for complete semantics, while its application to other semantics is, as to our knowledge, still to be developed.

Abstract dialectical frameworks (ADFs) [20] generalize Dung's framework by detaching the meaning of attack from the binary relation between arguments, so that each element of this relation is just a link representing a dependency. The meaning of the dependencies for each argument  $s$  is then expressed by an acceptance condition  $C_s$  which associates each subset of the set of parents of  $s$  with either in or out, namely gives a binary decision on the acceptance of  $s$  given the set of its parents which are accepted. Hence, in ADFs argument evaluation is, by definition, based on a strictly local criterion and any global evaluation arises bottom-up from the combination of the local ones.

While the present work is strictly focused on Dung's framework and the relevant semantics based on the attack relation, it appears that the basic ideas underlying our analysis have significant commonalities with the process of bottom-up evaluation in ADFs. Generalizing the results we have obtained to the context of ADFs is therefore a very important direction of future work.

## 10.2. Argumentation dynamics

Broadly speaking, in the context of abstract argumentation, dynamics concerns the evolution of a given framework to which one or more modifications (i.e. additions and/or deletions of arguments and/or attacks) are applied. These modifications can be exogenous and neutral, namely determined by some external event, or endogenous and goal-oriented, namely deliberately induced by an agent to reach some goal, like the acceptance of a desired argument. In the former case, the main interest is in determining the effect of the external modifications, in the latter, in identifying the minimal set of modifications sufficient to reach the goal. In both cases, one is typically interested in reusing as far as possible the results of previous computations carried out in the original framework so as to limit the amount of new computation required by the modification. Hence some of the pre-existing computation results have to be combined with the results of some partial computations in the new framework. Clearly the results presented in this paper are specifically related to this facet of argumentation dynamics and we focus on the relevant literature. A detailed analysis of the broader implications of our work on argumentation dynamics is beyond the scope of the present paper.

In [33] to save computation in a dynamic context the *division-based* method is proposed. Essentially, after a modification, the considered framework is divided into two parts, one unaffected and one affected. Briefly, the affected part consists of those arguments which are reachable (through a directed path of attacks) starting from any argument or attack involved in the modification. The identification of the unaffected part relies on the directionality property, which is required for the application of the method. To formalize the influence of the unaffected part over the affected part, the notion of conditioned argumentation framework is introduced, namely an argumentation framework receiving some attacks from arguments included in another argumentation framework. The paper then deals with incremental computation for some semantics satisfying the directionality property, namely complete, grounded, preferred and ideal<sup>14</sup> semantics. After the modification, one needs to recompute the extensions only for the affected part (modeled as a conditioned argumentation framework w.r.t. the unaffected part).

Some basic notions underlying the division-based method are related to our work. In particular, the notion of conditioned argumentation framework in [33] is similar to the notions of conditioning relation and of argumentation framework with input in our Definitions 10 and 11. Moreover, the incremental computation in a conditioned framework is analogous to the application of the local function introduced in Definition 11.

There is however an important difference due to the fact that the division-based method is essentially based on the directionality principle and, in particular, requires that there are no paths from the affected part to the unaffected part. As

<sup>14</sup> Actually the claim concerning ideal semantics turns out to be flawed, as recently pointed out in [11].

a consequence, the division-based method covers the cases where a framework is partitioned into two subframeworks such that one has an output, without having an input, and the other has an input (from the former) without having an output. The results concerning incremental computation of the four semantics considered in [33] correspond to a restricted form of semantics decomposability under these restrictive assumptions: both the unaffected and the affected part consist of a set of SCCs such that the SCCs included in the unaffected part precede those included in the affected part according to the partial order induced by the attack relation. Given this observation, the notion of decomposability in this context basically corresponds to a mild generalization of decomposability w.r.t.  $\mathcal{F}_{\text{SCC}}(AF)$ , i.e. of the weakest notion of decomposability considered in this paper, and is definitely weaker than decomposability w.r.t.  $\mathcal{F}_{\text{USCC}}(AF)$ . Our work is definitely more general as it concerns arbitrary partitions and does not rely on the directionality property. In particular, we prove full decomposability of stable semantics, which is not directional.

The work on splitting argumentation frameworks [17] focuses on modifications involving only additions of arguments or attacks (called expansions) and, apart of this restriction, shares the main basic assumptions with [33]. Considering a subclass of expansions called weak expansions, a splitting divides an argumentation framework into two subframeworks, such that only one of them receives attacks from the other: the two subframeworks correspond to the unaffected and affected parts of [33]. To model the effect of the unaffected subframework on the affected one, in [17] a modification of the affected subframework is introduced, which involves the addition of self-attacks and bears some similarity with our notion of standard argumentation framework for an argumentation framework with input. Then, the *splitting theorem* of [17] provides a decomposability result for stable, admissible, preferred, complete and grounded semantics, which, due to the restriction on the partitions considered, as in the case of [33], are weaker than the ones considered in this paper.

The restrictions that one of the two parts cannot receive attacks from the other one is lifted in [15] where an arbitrary partition of a framework into two parts is called *quasi splitting* and, using a technical arsenal rather different than ours, the decomposability property of stable semantics is proved. We achieved the same result for stable semantics in the context of a more general analysis, covering six additional literature semantics.

On the performance side, there are some empirical evidences that both the division-based method [32] and the splitting approach [16] may significantly reduce the computation time required for some standard problems in abstract argumentation w.r.t. to algorithms adopting a “monolithic” approach. Investigating the advantages provided by our more general approach in this respect is an important direction of future work.

### 10.3. Equivalence and interchangeability between argumentation frameworks

Various notions of equivalence for argumentation frameworks have been considered in the literature. The most basic ones focus either on structural correspondences (like the notion of syntactical equivalence, i.e. equality of arguments and attacks, used in [38] or the notion of isomorphism used in [8]) or on equality of extensions (w.r.t. a given semantics), which is called equivalence *tout court* in [38] and is analogous to the notion of equivalence between logic programs [34]. These notions are poorly or not at all related with modularity and interchangeability issues, that may arise in various contexts and in particular in presence of some form of argumentation dynamics.

To address this limitation, the notion of strong equivalence between argumentation framework (again, analogous to the one of strong equivalence between logic programs [34]) is introduced and investigated in [38]: two frameworks  $F$  and  $G$  are strongly equivalent w.r.t. a given semantics if for any argumentation framework  $H$ , the frameworks  $F \cup H$  and  $G \cup H$  have the same extensions. Basically,  $F$  and  $G$  must preserve the same outcomes in front of any operation of expansion. Since this requirement is, in fact, very strong, weaker notions of equivalence have subsequently been considered in the literature by restricting the set of expansions of the original frameworks encompassed. In particular four subclasses of expansions (called normal, weak, strong, and local<sup>15</sup>) are considered in [12] giving rise to four correspondent definitions of expansion equivalence all weaker than strong equivalence. A different notion of equivalence, introduced in [13], refers to the problem of minimal change: given a framework and a set of arguments  $E$  whose (credulous) acceptance has to be enforced, one is interested in identifying the minimal number of modifications that ensure the desired enforcement result. Two frameworks are minimal change equivalent, if for any set  $E$  the minimal number of modifications required to enforce  $E$  is the same in both frameworks. The relationships between all the above mentioned notions of equivalence have been analyzed in detail in [14].

The approach presented in this paper is complementary to the ones reviewed above: while these refer to several forms of invariance over the whole framework w.r.t. an operation of expansion, our work concerns invariance only in the unmodified part of the framework w.r.t. an operation of replacement. This involves a notion of equivalence in terms of Input/Output behavior and the study of the property of semantics transparency, which have no counterpart in the works cited above. As already mentioned, they can be related with the notion of strong equivalence in logic programming, while our approach is closer in spirit to the notion of modular equivalence between logic programs [37] and, more generally, with the study of modularity in this context [31]. A detailed analysis of the possible interplay between our results and the area of modular logic programming is beyond the scope of the present paper and is left for future work.

<sup>15</sup> This terminology, taken out of its context, may be a bit misleading: normal expansions are not the most general case of expansions, and the terms weak and strong here refer to the additional arguments, so strong expansions are not a subset of weak expansions.

The issue of substitutions within an argumentation framework is explicitly addressed by the notion of *fibring* [28] which indeed covers the more general case of combining together networks of different nature (e.g. embedding a neural network or a Bayesian network into an argumentation framework), including the special case of combination of networks of the same nature, called *self-fibring*. Due to the potential heterogeneity of the networks involved, however, fibring concerns the substitution of a *single node* of a network with an entire other network (neither of them having a notion of “interface” with the rest of the framework) and hence addresses a different kind of replacement than the one considered in this paper, which involves the two argumentation multipoles, i.e. two partial networks with well-defined interface. Moreover, the study presented in [28] covers generalized argumentation frameworks, featuring a richer set of relations (e.g. support, attacks to attacks, attacks arising from attacks, collective and disjoint attacks) than Dung’s framework, and investigates how this conceptual and technical arsenal can be used to properly transform the incoming and outgoing links involving just one node into links involving the nodes of the network replacing that node. Thus, the analysis in [28] goes deeply into these complex structural manipulations, which are mostly semantics independent, and does not concern the study of specific semantics properties. Our work, as already mentioned, concerns a different kind of replacement and lies in the context of traditional Dung frameworks, where we provide a systematic assessment of interchangeability-related properties for a comprehensive set of literature semantics. Extending and relating our results to generalized frameworks in the spirit of [28] is a further interesting direction of future work.

In [43] the notion of argumentation pattern is introduced in order to capture “general reusable solutions to commonly occurring problems in the design of argumentation frameworks”. Hence an argumentation pattern is understood as a reusable and modular component, in a spirit which has some analogy with the idea of argumentation multipole introduced in this paper. It has however to be observed that the notion of argumentation pattern lies at a higher level of abstraction than the one of argumentation multipole: the definition of argumentation pattern given in [43] involves a set of arguments and, basically, a set of possible labellings of these arguments. No notion of attack is explicitly involved, since an argumentation pattern captures a set of evaluation outcomes which together represent a “typical situation” seen from outside, independently of the (in fact, not necessarily univocal) underlying structure giving rise to this situation. Indeed, in [43] methods to translate (or flatten) a pattern into an argumentation framework and vice versa to extract a pattern from an argumentation framework (where arguments to be included in the pattern have been preliminarily identified) are devised. Our work, lying at different level, provides suitable technical foundations for further developments of the study of argumentation patterns. Indeed, our analysis concerning the equivalence of alternative representations of attacks to attacks in Section 9.2.2 strengthens the analysis of patterns for so called higher-order attacks in Section 3.2 of [43]. Moreover in [43] the issue of pattern combination is mentioned as a matter of future research, which may certainly benefit from the systematic set of results provided in this paper, applicable to the underlying flattened representation.

## 11. Conclusions

This paper contributes to the emerging research direction on modularity-based properties and techniques in abstract argumentation, by introducing a novel comprehensive formal corpus to describe the Input/Output behavior of argumentation frameworks along with the relevant semantics properties, and by providing a systematic assessment of seven well-known argumentation semantics in this context. Due to their foundational nature, we believe these results may play an enabling role in the development of a variety of more specific investigation lines, ranging from the sound combination of heterogeneous semantics to the definition of reusable argumentation patterns. As to future work, in addition to the many issues already included in the discussion of Section 10, we mention three further interesting lines. First, the extension of the analysis carried out in the paper to other literature semantics like the ones mainly based on the notion of conflict-freeness (e.g. stage [41], CF2 [7,10], stage2 [30] semantics) or those featuring a parametric definition (e.g. resolution-based semantics [6]). Second, the study of *argumentation synthesis* problems, namely, given a desired Input/Output behavior generating an argumentation framework which produces it, possibly under some constraints concerning its structure and/or the semantics to be adopted. Third, a systematic definition of modularity-related variations of traditional computational problems in abstract argumentation, e.g. checking whether two multipoles are equivalent according to a given semantics, and the analysis of their complexity properties.

## Acknowledgements

The authors thank the anonymous reviewers for their helpful comments.

## Appendix A. Proofs

### A.1. Proofs of Section 3

**Proposition 1.** Given a semantics  $\mathbf{S}$  and an argumentation framework  $AF$ ,  $F_{\mathbf{S}}(AF, \emptyset, \emptyset, \emptyset) = \mathbf{L}_{\mathbf{S}}(AF)$ .

**Proof.** The result is immediate by considering that the standard argumentation framework w.r.t.  $(AF, \emptyset, \emptyset, \emptyset)$  is  $AF$ .  $\square$

**Proposition 2.** Let  $\mathbf{S}$  be a complete-compatible semantics and let  $AF' = (Ar \cup \mathcal{J}', att \cup R_{\mathcal{J}'})$  be the standard argumentation framework w.r.t. an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Then for any  $Lab \in \mathbf{L}_{\mathbf{S}}(AF')$  it holds that  $Lab \downarrow_{\mathcal{J}'} = \{(A', in) \mid A \in out(L_{\mathcal{J}})\} \cup L_{\mathcal{J}}$  and  $Lab \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ .

**Proof.** Since  $\mathbf{S}$  is complete-compatible, on the basis of Definition 14 it is immediate to see that for any labelling  $Lab \in \mathbf{L}_{\mathbf{S}}(AF')$   $Lab \downarrow_{\mathcal{J}'} = \{(A', in) \mid A \in out(L_{\mathcal{J}})\} \cup L_{\mathcal{J}}$ , thus in particular  $Lab \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ .  $\square$

**Proposition 3.** Given a complete-compatible semantics  $\mathbf{S}$ , a set of arguments  $\mathcal{J}$  and a labelling  $L_{\mathcal{J}} \in \mathcal{L}_{\mathcal{J}}$ , it holds that  $F_{\mathbf{S}}(AF_{\emptyset}, \mathcal{J}, L_{\mathcal{J}}, \emptyset) = \{\emptyset\}$ .

**Proof.** According to Definition 12, the standard argumentation framework w.r.t.  $(AF_{\emptyset}, \mathcal{J}, L_{\mathcal{J}}, \emptyset)$  is  $AF' = (\mathcal{J}', att')$ , where  $\mathcal{J}' = \mathcal{J} \cup \{A' \mid A \in out(L_{\mathcal{J}})\}$  and  $att' = \{(A', A) \mid A \in out(L_{\mathcal{J}})\} \cup \{(A, A) \mid A \in undec(L_{\mathcal{J}})\}$ . According to Definition 13,  $F_{\mathbf{S}}(AF_{\emptyset}, \mathcal{J}, L_{\mathcal{J}}, \emptyset) = \{Lab \downarrow_{\emptyset} \mid Lab \in \mathbf{L}_{\mathbf{S}}(AF')\}$ . Since  $\mathbf{S}$  is complete-compatible,  $|\mathbf{L}_{\mathbf{S}}(AF')| = 1$ , thus  $F_{\mathbf{S}}(AF_{\emptyset}, \mathcal{J}, L_{\mathcal{J}}, \emptyset) = \{\emptyset\}$ .  $\square$

**Proposition 4.** GR, CO, ST, PR, SST, ID are all complete-compatible semantics.

**Proof.** First, it is easy too see that the three conditions of the first part of Definition 14 are satisfied by any complete labelling (see Definition 6). The desired conclusion then follows from the fact that, according to Definition 8, all labellings prescribed by GR, ST, PR, SST, ID are complete (note in particular that only the first two conditions have to be satisfied by ST, since there are no stable labellings in case a self-attacking argument is attacked by itself only). As to the second part of the definition, it is immediate to see that the labelling  $\{(A', in) \mid A \in out(L_{\mathcal{J}})\} \cup L_{\mathcal{J}}$  is the unique complete labelling, thus it is also grounded, preferred, semi-stable and ideal. Moreover, such labelling is also stable (note that in this case there are no self-attacking arguments in the framework).  $\square$

**Proposition 5.** Given a complete-compatible semantics  $\mathbf{S}$ , if  $\mathbf{S}$  is fully decomposable then there is a unique local function satisfying the conditions of Definition 15, coinciding with the canonical local function  $F_{\mathbf{S}}$ .

**Proof.** For each argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , consider the standard argumentation framework  $AF'$  w.r.t. it, and the partition  $\{Ar, \mathcal{J}'\}$  where  $\mathcal{J}' = \mathcal{J} \cup \{A' \mid A \in out(L_{\mathcal{J}})\}$ . Since  $\mathbf{S}$  is decomposable, according to Definition 15 it must be the case that  $\mathbf{L}_{\mathbf{S}}(AF') = \{L_{Ar} \cup L_{\mathcal{J}'} \mid L_{Ar} \in F(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}), L_{\mathcal{J}'} \in F(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)\}$ . By Proposition 2, for any  $Lab \in \mathbf{L}_{\mathbf{S}}(AF')$  it must be the case that  $Lab \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ , thus from the above condition  $\mathbf{L}_{\mathbf{S}}(AF') = \{L_{Ar} \cup L_{\mathcal{J}'} \mid L_{Ar} \in F(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}), L_{\mathcal{J}'} \in F(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)\}$ . Moreover,  $F(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$  is nonempty (in particular there is a unique labelling  $L_{\mathcal{J}'} \in F(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$ ), since Definition 15 applied to the partition  $\{\mathcal{J}'\}$  of  $AF' \downarrow_{\mathcal{J}'}$  yields  $\mathbf{L}_{\mathbf{S}}(AF' \downarrow_{\mathcal{J}'}) = F(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$  and  $|\mathbf{L}_{\mathbf{S}}(AF' \downarrow_{\mathcal{J}'})| = 1$  since  $\mathbf{S}$  is complete-compatible. This entails that  $F(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) = \{Lab \downarrow_{Ar} \mid Lab \in \mathbf{L}_{\mathbf{S}}(AF')\}$ , which according to Definition 13 coincides with the canonical local function  $F_{\mathbf{S}}$ .  $\square$

## A.2. Proofs of Section 4

**Theorem 1.** Admissible semantics AD is fully decomposable, with  $F_{\mathbf{AD}}$  satisfying the conditions of Definition 15.

**Proof.** We have to prove that, for any argumentation framework  $AF = (Ar, att)$  and any partition  $\{P_1, \dots, P_n\}$  of  $Ar$ ,  $\mathbf{L}_{\mathbf{AD}}(AF) = \{L_{P_1} \cup \dots \cup L_{P_n} \mid L_{P_i} \in F_{\mathbf{AD}}(AF \downarrow_{P_i}, P_i^{in}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{in}}, P_i^R)\}$ . As to  $\mathbf{L}_{\mathbf{AD}}(AF) \subseteq \{L_{P_1} \cup \dots \cup L_{P_n} \mid L_{P_i} \in F_{\mathbf{AD}}(AF \downarrow_{P_i}, P_i^{in}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{in}}, P_i^R)\}$ , given  $\mathbf{L} \in \mathbf{L}_{\mathbf{AD}}(AF)$ , let us consider a generic  $P = P_i$  and let  $L_P = \mathbf{L} \downarrow_{P_i}$  and  $L_{Ar \setminus P} = \mathbf{L} \downarrow_{(\bigcup_{j=1 \dots n, j \neq i} P_j)}$ : we have to prove that  $L_P \in F_{\mathbf{AD}}(AF \downarrow_P, P^{in}, L_{Ar \setminus P} \downarrow_{P^{in}}, P^R)$ , i.e. that the conditions of Definition 23 are satisfied for  $L_P$ . Given a generic argument  $A \in P$ :

- if  $L_P(A) = in$  then  $\mathbf{L}(A) = in$  and since  $\mathbf{L}$  is an admissible labelling it must be the case that  $\forall B \in Ar : (B, A) \in att, \mathbf{L}(B) = out$ . As a consequence,  $\forall B \in P : (B, A) \in att, \mathbf{L}(B) = out$  with  $\mathbf{L}(B) = L_P(B)$ , and  $\forall B \in P^{in} : (B, A) \in P^R$ , since  $P^R \subseteq att$  and  $L_{Ar \setminus P} \downarrow_{P^{in}}(B) = \mathbf{L}(B)$  it must be the case that  $L_{Ar \setminus P} \downarrow_{P^{in}}(B) = out$ ;
- if  $L_P(A) = out$  then  $\mathbf{L}(A) = out$  and since  $\mathbf{L}$  is an admissible labelling  $\exists B \in Ar : (B, A) \in att \wedge \mathbf{L}(B) = in$ . There are two cases to consider. If  $B \in P$ , then  $L_P(B) = \mathbf{L}(B) = in$  and the condition is verified. If  $B \notin P$ , then since  $(B, A) \in att$  it holds that  $B \in P^{in}$ , and  $L_{Ar \setminus P} \downarrow_{P^{in}}(B) = \mathbf{L}(B) = in$ ;

Turning to  $\mathbf{L}_{\mathbf{AD}}(AF) \supseteq \{L_{P_1} \cup \dots \cup L_{P_n} \mid L_{P_i} \in F_{\mathbf{AD}}(AF \downarrow_{P_i}, P_i^{in}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{in}}, P_i^R)\}$ , let  $L_{P_1}, \dots, L_{P_n}$  be  $n$  labellings of  $P_1, \dots, P_n$  such that  $\forall i L_{P_i} \in F_{\mathbf{AD}}(AF \downarrow_{P_i}, P_i^{in}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{in}}, P_i^R)$ . Letting  $\mathbf{L} = L_{P_1} \cup \dots \cup L_{P_n}$ , we prove that  $\mathbf{L} \in \mathbf{L}_{\mathbf{AD}}(AF)$ . We consider a generic argument  $A \in Ar$ , denoting as  $P$  the set  $P_i$  such that  $A \in P_i$ , as  $L_{Ar \setminus P}$  the labelling  $(\bigcup_{j=1 \dots n, j \neq i} L_{P_j})$ , and we prove that the conditions of Definition 5 are satisfied:

- if  $\mathbf{L}(A) = \text{in}$  then  $L_P(A) = \text{in}$ . On the basis of Definition 23, given  $B \in \text{Ar} : (B, A) \in \text{att}$ , if  $B \in P$  then  $\mathbf{L}(B) = L_P(B) = \text{out}$ , if  $B \notin P$  then  $B \in P^{\text{inp}}$ , thus  $\mathbf{L}(B) = L_{\text{Ar} \setminus P \downarrow P^{\text{inp}}}(B) = \text{out}$ ;
- if  $\mathbf{L}(A) = \text{out}$  then  $L_P(A) = \text{out}$ . On the basis of Definition 23,  $\exists B \in P : (B, A) \in \text{att} \wedge L_P(B) = \text{in}$  or  $\exists B \in P^{\text{inp}} : (B, A) \in P^R \wedge L_{\text{Ar} \setminus P \downarrow P^{\text{inp}}}(B) = \text{in}$ . Since  $P^R \subseteq \text{att}$  and  $\mathbf{L} = L_P \cup L_{\text{Ar} \setminus P}$ , it holds that  $\exists B : (B, A) \in \text{att}$  with  $\mathbf{L}(B) = \text{in}$ ;  $\square$

**Theorem 2.** The canonical local function of admissible semantics is  $F_{\text{AD}}$ , as defined in Definition 23.

**Proof.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where  $AF = (\text{Ar}, \text{att})$ , let  $AF' = (\text{Ar} \cup \mathcal{J}', \text{att} \cup R'_{\mathcal{J}'})$  be the standard argumentation framework w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . We have to show that  $\{\mathbf{L}' \downarrow_{\text{Ar}} \mid \mathbf{L}' \in \mathbf{L}_{\text{AD}}(AF')\} = F_{\text{AD}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ .

As to  $\{\mathbf{L}' \downarrow_{\text{Ar}} \mid \mathbf{L}' \in \mathbf{L}_{\text{AD}}(AF')\} \subseteq F_{\text{AD}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , given an admissible labelling  $\mathbf{L}'$  of  $AF'$ , we prove that  $\mathbf{L}' \downarrow_{\text{Ar}} \in F_{\text{AD}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , i.e. satisfies the conditions of Definition 23. Given an argument  $A \in \text{Ar}$  such that  $\mathbf{L}' \downarrow_{\text{Ar}}(A) = \text{in}$ , obviously  $\mathbf{L}'(A) = \text{in}$ , thus by the definition of admissible labelling all the attackers of  $A$  in  $AF'$  are labelled  $\text{out}$ , entailing the condition  $(\forall B \in \text{Ar} : (B, A) \in \text{att}, \text{Lab}(B) = \text{out}) \wedge (\forall B \in \mathcal{J}' : (B, A) \in R_{\mathcal{J}'}, L_{\mathcal{J}'}(B) = \text{out})$ . Similarly, given an argument  $A \in \text{Ar}$  such that  $\mathbf{L}' \downarrow_{\text{Ar}}(A) = \text{out}$ , since  $\mathbf{L}'$  is an admissible labelling of  $AF'$  there is an attacker of  $A$  labelled  $\text{in}$ , entailing the condition  $(\exists B \in \text{Ar} : (B, A) \in \text{att} \wedge \text{Lab}(B) = \text{in}) \vee (\exists B \in \mathcal{J}' : (B, A) \in R_{\mathcal{J}'} \wedge L_{\mathcal{J}'}(B) = \text{in})$ .

As to  $\{\mathbf{L}' \downarrow_{\text{Ar}} \mid \mathbf{L}' \in \mathbf{L}_{\text{AD}}(AF')\} \supseteq F_{\text{AD}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , consider a labelling  $\mathbf{L} \in F_{\text{AD}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : letting  $\mathbf{L}^*$  be the grounded labelling of  $AF' \downarrow_{\mathcal{J}'}$ , we prove that  $\mathbf{L}^* \cup \mathbf{L}$  is an admissible labelling of  $AF'$ . Since  $\mathbf{L}^*$  is an admissible labelling of  $AF' \downarrow_{\mathcal{J}'}$ , it is easy to see that  $\mathbf{L}^* \in F_{\text{AD}}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$ . Moreover, since  $\mathbf{GR}$  is complete-compatible  $\mathbf{L}^* \downarrow_{\mathcal{J}'} = L_{\mathcal{J}'}$ , thus  $\mathbf{L} \in F_{\text{AD}}(AF' \downarrow_{\text{Ar}}, \mathcal{J}, \mathbf{L}^* \downarrow_{\mathcal{J}'}, R_{\mathcal{J}})$  (where it has been taken into account that  $AF = AF' \downarrow_{\text{Ar}}$ ). Then the conclusion that  $\mathbf{L}^* \cup \mathbf{L}$  is an admissible labelling of  $AF'$  holds by Theorem 1.  $\square$

**Theorem 3.** Complete semantics  $\mathbf{CO}$  is fully decomposable and  $F_{\text{CO}}$  is its canonical local function.

**Proof.** We have to prove that, for any argumentation framework  $AF = (\text{Ar}, \text{att})$  and any partition  $\{P_1, \dots, P_n\}$  of  $\text{Ar}$ ,  $\mathbf{L}_{\text{CO}}(AF) = \{L_{P_1} \cup \dots \cup L_{P_n} \mid L_{P_i} \in F_{\text{CO}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}}, P_i^R)\}$ .

First, given  $\mathbf{L} \in \mathbf{L}_{\text{CO}}(AF)$ , let us consider a generic  $P = P_i$  and let  $L_P = \mathbf{L} \downarrow_{P_i}$  and  $L_{\text{Ar} \setminus P} = \mathbf{L} \downarrow_{(\bigcup_{j=1 \dots n, j \neq i} P_j)}$ : we have to prove that  $L_P \in F_{\text{CO}}(AF \downarrow_P, P^{\text{inp}}, L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}, P^R)$ . Since by Theorem 1  $L_P \in F_{\text{AD}}(AF \downarrow_P, P^{\text{inp}}, L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}, P^R)$ , we have only to prove that the third condition of Definition 24 is satisfied for  $L_P$ . Given a generic argument  $A \in P$ , if  $L_P(A) = \text{undec}$  then  $\mathbf{L}(A) = \text{undec}$ , and since  $\mathbf{L}$  is a complete labelling there is no  $B$  such that  $(B, A) \in \text{att}$  with  $\mathbf{L}(B) = \text{in}$ , and there is an argument  $C$  such that  $(C, A) \in \text{att}$  and  $\mathbf{L}(C) = \text{undec}$ . The first condition entails that  $(\forall B \in \text{Ar} : (B, A) \in \text{att}, L_P(B) \neq \text{in}) \wedge (\forall B \in P^{\text{inp}} : (B, A) \in P^R, L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}(B) \neq \text{in})$ . The second condition in turn entails that  $(\exists B \in \text{Ar} : (B, A) \in \text{att} \wedge L_P(B) = \text{undec}) \vee (\exists B \in P^{\text{inp}} : (B, A) \in P^R \wedge L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}(B) = \text{undec})$ .

As for the other direction of the proof, let  $L_{P_1}, \dots, L_{P_n}$  be  $n$  labellings of  $P_1, \dots, P_n$  such that  $\forall i L_{P_i} \in F_{\text{CO}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}}, P_i^R)$ . Letting  $\mathbf{L} = L_{P_1} \cup \dots \cup L_{P_n}$ , we prove that  $\mathbf{L} \in \mathbf{L}_{\text{CO}}(AF)$ . Since by Theorem 1  $\mathbf{L}$  is admissible, we have only to prove the third condition of Definition 6. We consider a generic argument  $A \in \text{Ar}$ , denoting as  $P$  the set  $P_i$  such that  $A \in P_i$ , as  $L_{\text{Ar} \setminus P}$  the labelling  $(\bigcup_{j=1 \dots n, j \neq i} L_{P_j})$ . If  $\mathbf{L}(A) = \text{undec}$  then  $L_P(A) = \text{undec}$ . For every  $B \in \text{Ar} : (B, A) \in \text{att}$ , according to Definition 24 if  $B \in P$  then  $\mathbf{L}(B) = L_P(B) \neq \text{in}$ , if  $B \notin P$  then  $B \in P^{\text{inp}}$ , thus  $\mathbf{L}(B) = L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}(B) \neq \text{in}$ . Moreover,  $\exists B \in P : (B, A) \in \text{att} \wedge L_P(B) = \text{undec}$  or  $\exists B \in P^{\text{inp}} : (B, A) \in P^R \wedge L_{\text{Ar} \setminus P} \downarrow_{P^{\text{inp}}}(B) = \text{undec}$ . Since  $P^R \subseteq \text{att}$  and  $\mathbf{L} = L_P \cup L_{\text{Ar} \setminus P}$ , this entails that  $\exists B \in \text{Ar} : (B, A) \in \text{att} \wedge \mathbf{L}(B) = \text{undec}$ .

Finally, since the proof adopts  $F_{\text{CO}}$  as the local function and  $\mathbf{CO}$  is complete-compatible, by Proposition 5 it holds that  $F_{\text{CO}}$  is the canonical local function of complete semantics.  $\square$

**Theorem 4.** Stable semantics  $\mathbf{ST}$  is fully decomposable and  $F_{\text{ST}}$  is its canonical local function.

**Proof.** First, consider a labelling  $\mathbf{L} \in \mathbf{L}_{\text{ST}}(AF)$ . By definition,  $\mathbf{L}$  is a complete labelling. Since complete semantics is fully decomposable and thus in particular top-down decomposable, for any partition  $\{P_1, \dots, P_n\}$  of  $\text{Ar}$  it holds that  $\forall i L_{P_i} \in F_{\text{CO}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, L_{\text{Ar} \setminus P_i} \downarrow_{P_i^{\text{inp}}}, P_i^R)$ . From the absence of  $\text{undec}$ -labelled arguments according to  $\mathbf{L}$ , it also holds that  $L_{P_i} \in F_{\text{ST}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, L_{\text{Ar} \setminus P_i} \downarrow_{P_i^{\text{inp}}}, P_i^R)$ .

As to the other direction of the proof, let  $L_{P_1}, \dots, L_{P_n}$  be  $n$  labellings of  $P_1, \dots, P_n$  such that  $\forall i L_{P_i} \in F_{\text{ST}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}}, P_i^R)$ . Let  $\mathbf{L} = L_{P_1} \cup \dots \cup L_{P_n}$ . By definition of  $F_{\text{ST}}$ , there are no  $\text{undec}$ -labelled arguments according to  $\mathbf{L}$ . Moreover, since  $F_{\text{ST}}$  always returns a set of complete labellings, bottom-up decomposability of complete semantics entails that  $\mathbf{L} \in \mathbf{L}_{\text{CO}}(AF)$ , thus  $\mathbf{L}$  is a stable labelling.

Finally, similarly to the proof of Theorem 3, by Proposition 5 it holds that  $F_{\text{ST}}$  is the canonical local function of stable semantics.  $\square$

**Proposition 6.** The canonical local function of grounded and preferred semantics are defined as



- $F_{GR}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L}^*\}$ , where  $\mathbf{L}^*$  is the minimal (w.r.t.  $\sqsubseteq$ ) labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$
- $F_{PR}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L} \mid \mathbf{L} \text{ is a maximal (w.r.t. } \sqsubseteq) \text{ labelling in } F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})\}$ .

**Proof.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where  $AF = (Ar, att)$ , let  $AF' = (Ar \cup \mathcal{J}', att \cup R'_{\mathcal{J}})$  be the standard argumentation w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ .

As to  $F_{GR}$ , let  $\mathbf{L}'$  be the grounded labelling of  $AF'$ : we prove that  $\{\mathbf{L}' \downarrow_{Ar}\}$  coincides with  $\{\mathbf{L} \mid \mathbf{L} \text{ is minimal in } F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})\}$ . Notice that we do not assume that the definition of  $F_{GR}$  is well founded, i.e. that there is a unique minimal labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : this is obtained as a by-product of the proof.

First we prove that  $\mathbf{L}' \downarrow_{Ar}$  is a minimal labelling of  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Since  $\mathbf{L}'$  is by definition a complete labelling and complete semantics is top-down decomposable (see Theorem 3), it holds that  $\mathbf{L}' \downarrow_{\mathcal{J}'} \in F_{CO}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$  and  $\mathbf{L}' \downarrow_{Ar} \in F_{CO}(AF, \mathcal{J}, \mathbf{L}' \downarrow_{\mathcal{J}}, R_{\mathcal{J}})$ . Since by Proposition 2 it holds that  $\mathbf{L}' \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ , the latter condition can be rewritten as  $\mathbf{L}' \downarrow_{Ar} \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , entailing in particular that  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  is nonempty. Suppose now by contradiction that  $\mathbf{L}' \downarrow_{Ar}$  is not a minimal labelling returned by  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : then there is a labelling  $\mathbf{L}^* \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}^* \sqsubseteq \mathbf{L}' \downarrow_{Ar}$  and  $\mathbf{L}^* \neq \mathbf{L}' \downarrow_{Ar}$ . Note that  $\mathbf{L}' \downarrow_{\mathcal{J}'}$  and  $\mathbf{L}^*$  are compatible (since  $\mathbf{L}' \downarrow_{\mathcal{J}'}$  coincides in  $\mathcal{J}'$  with  $L_{\mathcal{J}'}$ ), thus from bottom-up decomposability of complete semantics  $\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}^*$  is a complete labelling of  $AF'$ . Since  $(\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}^*) \sqsubseteq \mathbf{L}'$  and  $(\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}^*) \neq \mathbf{L}'$ , this contradicts the fact that  $\mathbf{L}'$  is by definition the minimal (w.r.t.  $\sqsubseteq$ ) complete labelling of  $AF'$ .

Let us now consider a labelling  $\mathbf{L}$  of  $AF$  which is minimal among those of  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : we prove that  $\mathbf{L} = \mathbf{L}' \downarrow_{Ar}$ , thus also showing that there is a unique minimal labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Since  $\mathbf{L}'$  is by definition a complete labelling and complete semantics is decomposable, it must be the case that  $\mathbf{L}' \downarrow_{\mathcal{J}'} \in F_{CO}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$ , thus by bottom-up decomposability of complete semantics  $\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}$  is a complete labelling of  $AF'$  (recall that  $\mathbf{L}'$  coincides in  $\mathcal{J}$  with  $L_{\mathcal{J}}$ ). By definition of grounded labelling,  $\mathbf{L}' \sqsubseteq (\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L})$ . Assume by contradiction that  $(\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}) \neq \mathbf{L}'$ : since the grounded labelling is minimal (w.r.t.  $\sqsubseteq$ ) among all complete labellings, it must be the case that  $\mathbf{L}' \downarrow_{Ar} \sqsubseteq \mathbf{L}$  and  $\mathbf{L}' \downarrow_{Ar} \neq \mathbf{L}$ . However, by top-down decomposability of complete semantics  $\mathbf{L}' \downarrow_{Ar} \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , contradicting the minimality of  $\mathbf{L}$ . This completes the proof as far as the canonical local function of grounded semantics is concerned.

As to  $F_{PR}$ , we have to show that  $\{\mathbf{L}' \downarrow_{Ar} \mid \mathbf{L}' \in \mathbf{LPR}(AF')\} = \{\mathbf{L} \mid \mathbf{L} \text{ is maximal in } F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})\}$ . The proof is similar to that for  $F_{GR}$ .

First, given a preferred labelling  $\mathbf{L}'$  of  $AF'$ , we prove that  $\mathbf{L}' \downarrow_{Ar}$  is a maximal labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Since  $\mathbf{L}'$  is by definition a complete labelling, by top-down decomposability of complete semantics  $\mathbf{L}' \downarrow_{\mathcal{J}'} \in F_{CO}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$  and  $\mathbf{L}' \downarrow_{Ar} \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where we exploit the fact that, by Proposition 2,  $\mathbf{L}' \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ . Assume by contradiction that  $\mathbf{L}' \downarrow_{Ar}$  is not a maximal labelling in  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : this entails that there is a labelling  $\mathbf{L}^* \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}' \downarrow_{Ar} \sqsubseteq \mathbf{L}^*$  and  $\mathbf{L}' \downarrow_{Ar} \neq \mathbf{L}^*$ . Note that  $\mathbf{L}^*$  and  $\mathbf{L}' \downarrow_{\mathcal{J}'}$  are compatible since  $\mathbf{L}'$  coincides in  $\mathcal{J}$  with  $L_{\mathcal{J}}$ , thus from bottom-up decomposability of complete semantics  $\mathbf{L}^* \cup \mathbf{L}' \downarrow_{\mathcal{J}'}$  is a complete labelling of  $AF'$ . However,  $\mathbf{L}' \sqsubseteq (\mathbf{L}^* \cup \mathbf{L}' \downarrow_{\mathcal{J}'})$  and  $\mathbf{L}' \neq (\mathbf{L}^* \cup \mathbf{L}' \downarrow_{\mathcal{J}'})$ , contradicting the fact that  $\mathbf{L}'$  is by definition a maximal (w.r.t.  $\sqsubseteq$ ) complete labelling.

Let us now consider a labelling  $\mathbf{L}$  of  $AF$  which is maximal among those of  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ : letting  $\mathbf{L}^*$  be the grounded labelling of  $AF'$ , we prove that  $\mathbf{L}^* \downarrow_{\mathcal{J}'} \cup \mathbf{L}$  is a preferred labelling of  $AF'$ . By bottom-up decomposability of complete semantics  $\mathbf{L}^* \downarrow_{\mathcal{J}'} \cup \mathbf{L}$  is a complete labelling of  $AF'$ . Assume by contradiction that it is not maximal: then there is a preferred labelling  $\mathbf{L}'$  of  $AF'$  such that  $(\mathbf{L}^* \downarrow_{\mathcal{J}'} \cup \mathbf{L}) \sqsubseteq \mathbf{L}'$  and  $(\mathbf{L}^* \downarrow_{\mathcal{J}'} \cup \mathbf{L}) \neq \mathbf{L}'$ . Since, by Proposition 2,  $\mathbf{L}' \downarrow_{\mathcal{J}'} = \mathbf{L}^* \downarrow_{\mathcal{J}'}$ , this entails that  $\mathbf{L} \sqsubseteq \mathbf{L}' \downarrow_{Ar}$  and  $\mathbf{L} \neq \mathbf{L}' \downarrow_{Ar}$ . However, top-down decomposability of complete semantics also entails that  $\mathbf{L}' \downarrow_{Ar} \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , contradicting the maximality of  $\mathbf{L}$  among the labellings of  $F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ .  $\square$

**Lemma 1.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where  $AF = (Ar, att)$ , let  $\mathbf{L}$  be a labelling such that  $\mathbf{L} \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $\mathbf{L} \notin F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Then there is an argument  $A \in Ar$  such that  $\mathbf{L}(A) = \text{undec}$  and a labelling  $\mathbf{L}^A \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}^A(A) \in \{\text{in}, \text{out}\}$  and  $\forall B \in Ar: B \neq A, \mathbf{L}^A(B) = \mathbf{L}(B)$ .

**Proof.** Since  $\mathbf{L} \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $\mathbf{L} \notin F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , according to Definitions 23 and 24 there must be at least an undec-labelled argument  $A$  such that either  $(\forall B \in Ar: (B, A) \in att, \mathbf{L}(B) = \text{out}) \wedge (\forall B \in \mathcal{J}: (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}(B) = \text{out})$  or  $(\exists B \in Ar: (B, A) \in att \wedge \mathbf{L}(B) = \text{in}) \vee (\exists B \in \mathcal{J}: (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}(B) = \text{in})$ . Then the labelling  $\mathbf{L}^A$  is constructed such that  $\forall B \neq A, \mathbf{L}^A(B) = \mathbf{L}(B)$ , and  $\mathbf{L}^A(A) = \text{in}$  in the first case and  $\mathbf{L}^A(A) = \text{out}$  in the second case. It is easy to see that  $\mathbf{L}^A \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  by checking the conditions of Definition 23. In particular,  $A$  satisfies the conditions of in-labelled arguments (in the first case) or the conditions of out-labelled arguments (in the second case) by construction, all of the arguments besides  $A$  labelled in by  $\mathbf{L}^A$  are labelled in by  $\mathbf{L}$  and thus have their attackers labelled out by  $\mathbf{L}$  (and  $L_{\mathcal{J}}$ ) and thus by  $\mathbf{L}^A$  (and  $L_{\mathcal{J}}$ ), all of the arguments besides  $A$  labelled out by  $\mathbf{L}^A$  are labelled out by  $\mathbf{L}$  and thus have an attacker which is labelled in by  $\mathbf{L}$  (or  $L_{\mathcal{J}}$ ) and thus by  $\mathbf{L}^A$  (or  $L_{\mathcal{J}}$ ).  $\square$

**Lemma 2.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , for every labelling  $\mathbf{L}_1 \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  there exists a labelling  $\mathbf{L}_2 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ .

**Proof.** Given  $\mathbf{L}_1 \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , if  $\mathbf{L}_1 \in F_{CO}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  then the conclusion trivially follows, otherwise according to Lemma 1 there is an argument  $A \in \text{Args}$  and a labelling  $\mathbf{L}_1^A \in F_{AD}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_1^A$ ,  $\mathbf{L}_1^A(A) = \text{undec}$  and  $\mathbf{L}_1^A(A) \in \{\text{in}, \text{out}\}$ . Since  $AF$  is finite and it is thus impossible to indefinitely turn undec-labelled



arguments to in or out, iterating this step must yield a labelling  $\mathbf{L}_2$  such that  $\mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . The relation  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$  holds by construction.  $\square$

**Proposition 7.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , let  $L_{\mathcal{J}}^1, L_{\mathcal{J}}^2 \in \mathcal{L}_{\mathcal{J}}$  be two labellings of  $\mathcal{J}$  such that  $L_{\mathcal{J}}^1 \sqsubseteq L_{\mathcal{J}}^2$ . Then it holds that

1.  $\forall \mathbf{L}_1 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}}), \exists \mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ ; and
2.  $\forall \mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}}), \exists \mathbf{L}_1 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ .

**Proof.** As to the first point, since  $\mathbf{L}_1 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$  and all arguments in  $\mathcal{J}$  that are labelled in (respectively out) by  $L_{\mathcal{J}}^1$  are also labelled in (respectively out) by  $L_{\mathcal{J}}^2$ , it is easy to see that  $\mathbf{L}_1 \in \mathbf{Fad}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$ . Then the conclusion follows from Lemma 2.

As to the second point, given  $\mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$  consider the set of labellings  $\Lambda_{\mathbf{L}_2} = \{\mathbf{L} \in \mathbf{Fad}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}}) \mid \mathbf{L} \sqsubseteq \mathbf{L}_2\}$ . Note that  $\Lambda_{\mathbf{L}_2}$  is nonempty, as it includes the labelling that assigns undec to all arguments, and since  $AF$  is finite there is at least a maximal (w.r.t.  $\sqsubseteq$ ) labelling among those of  $\Lambda_{\mathbf{L}_2}$ . Let  $\mathbf{L}_1$  be a maximal labelling in  $\Lambda_{\mathbf{L}_2}$ : we prove the claim by showing that  $\mathbf{L}_1 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$ . Suppose by contradiction that this is not the case: then, by Lemma 1 there is an argument  $A \in \text{Args}$  and a labelling  $\mathbf{L}_1^A \in \mathbf{Fad}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$  such that  $\mathbf{L}_1^A(A) = \text{undec}$ ,  $\mathbf{L}_1^A(A) \in \{\text{in}, \text{out}\}$  and  $\forall B \in \text{Ar} : B \neq A, \mathbf{L}_1^A(B) = \mathbf{L}_1(B)$ . There are two cases to consider.

- If  $\mathbf{L}_1^A(A) = \text{in}$ , then it must be the case that  $\mathbf{L}_2(A) \neq \text{in}$ , otherwise from  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$  we would also have  $\mathbf{L}_1^A \sqsubseteq \mathbf{L}_2$  and  $\mathbf{L}_1$  would not be a maximal element of  $\Lambda_{\mathbf{L}_2}$  (observe that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_1^A$  and  $\mathbf{L}_1^A \not\sqsubseteq \mathbf{L}_1$ ). Since  $\mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$ , this in turns entails that  $\exists B \in \text{Ar} : (B, A) \in \text{att} \wedge \mathbf{L}_2(B) \neq \text{out}$ , or  $\exists B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}^2(B) \neq \text{out}$  (otherwise  $\mathbf{L}_2(A)$  would be in according to Definition 24). However, since  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$  and  $L_{\mathcal{J}}^1 \sqsubseteq L_{\mathcal{J}}^2$ , the same condition holds for  $\mathbf{L}_1$  and  $L_{\mathcal{J}}^1$ , i.e.  $\exists B \in \text{Ar} : (B, A) \in \text{att} \wedge \mathbf{L}_1(B) \neq \text{out}$ , or  $\exists B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}} \wedge L_{\mathcal{J}}^1(B) \neq \text{out}$ . This contradicts the fact that  $\mathbf{L}_1^A \in \mathbf{Fad}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$ , since  $A$  does not satisfy the conditions of Definition 23 required by in-labelled arguments.
- The other case to consider is  $\mathbf{L}_1^A(A) = \text{out}$ , which similarly to the first case entails  $\mathbf{L}_2(A) \neq \text{out}$ , otherwise  $\mathbf{L}_1^A \sqsubseteq \mathbf{L}_2$  and  $\mathbf{L}_1$  would not be a maximal element of  $\Lambda_{\mathbf{L}_2}$ . Since  $\mathbf{L}_2 \in \mathbf{Fco}(AF, \mathcal{J}, L_{\mathcal{J}}^2, R_{\mathcal{J}})$ , this in turns entails that  $\forall B \in \text{Ar} : (B, A) \in \text{att}, \mathbf{L}_2(B) \neq \text{in}$  and  $\forall B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}^2(B) \neq \text{in}$ . However, since  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$  and  $L_{\mathcal{J}}^1 \sqsubseteq L_{\mathcal{J}}^2$ , the same condition holds for  $\mathbf{L}_1$  and  $L_{\mathcal{J}}^1$ , i.e.  $\forall B \in \text{Ar} : (B, A) \in \text{att}, \mathbf{L}_1(B) \neq \text{in}$  and  $\forall B \in \mathcal{J} : (B, A) \in R_{\mathcal{J}}, L_{\mathcal{J}}^1(B) \neq \text{in}$ , contradicting the fact that  $\mathbf{L}_1^A \in \mathbf{Fad}(AF, \mathcal{J}, L_{\mathcal{J}}^1, R_{\mathcal{J}})$ .  $\square$

**Theorem 5.** Given an argumentation framework  $AF = (Ar, \text{att})$ , let  $\mathbf{L}$  be the grounded labelling of  $AF$ . For any set  $P \subseteq Ar$ ,  $\mathbf{L} \downarrow_P \in \mathbf{FGR}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L} \downarrow_{P^{\text{inp}}}, P^R)$ .

**Proof.** Let  $Q = Ar \setminus P$ . Since  $\mathbf{L}$  is the grounded labelling of  $AF$  which by definition is also complete, letting  $\mathbf{L}_Q = \mathbf{L} \downarrow_Q$  and  $\mathbf{L}_P = \mathbf{L} \downarrow_P$  we have by Theorem 3 that  $\mathbf{L}_Q \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P \downarrow_{Q^{\text{inp}}}, Q^R)$  and  $\mathbf{L}_P \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q \downarrow_{P^{\text{inp}}}, P^R)$ . Suppose by contradiction that  $\mathbf{L}_P$  is not the minimal element w.r.t.  $\sqsubseteq$  in  $\mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q \downarrow_{P^{\text{inp}}}, P^R)$ , and let  $\mathbf{L}_P^2 \sqsubseteq \mathbf{L}_P$  (with  $\mathbf{L}_P \not\sqsubseteq \mathbf{L}_P^2$ ) be the minimal element. Since in particular  $\mathbf{L}_P^2 \downarrow_{Q^{\text{inp}}} \sqsubseteq \mathbf{L}_P \downarrow_{Q^{\text{inp}}}$ , by Proposition 7 (second point) there is a labelling  $\mathbf{L}_Q^2 \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^2 \downarrow_{Q^{\text{inp}}}, Q^R)$  such that  $\mathbf{L}_Q^2 \sqsubseteq \mathbf{L}_Q$ . Since in particular  $\mathbf{L}_Q^2 \downarrow_{P^{\text{inp}}} \sqsubseteq \mathbf{L}_Q \downarrow_{P^{\text{inp}}}$ , again by Proposition 7 there is a labelling  $\mathbf{L}_P^3 \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^2 \downarrow_{P^{\text{inp}}}, P^R)$  such that  $\mathbf{L}_P^3 \sqsubseteq \mathbf{L}_P^2$ . Iterating the same reasoning, for any  $i > 2$  we get  $\exists \mathbf{L}_P^i \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^{i-1} \downarrow_{P^{\text{inp}}}, P^R)$  such that  $\mathbf{L}_P^i \sqsubseteq \mathbf{L}_P^{i-1} \sqsubseteq \dots \sqsubseteq \mathbf{L}_P^2 \sqsubseteq \mathbf{L}_P$ , and  $\exists \mathbf{L}_Q^i \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^i \downarrow_{Q^{\text{inp}}}, Q^R)$  such that  $\mathbf{L}_Q^i \sqsubseteq \mathbf{L}_Q^{i-1} \sqsubseteq \dots \sqsubseteq \mathbf{L}_Q^2 \sqsubseteq \mathbf{L}_Q$ . Since the argumentation framework is finite, there must be an  $i^*$  such that  $\forall j > i^* \mathbf{L}_P^j = \mathbf{L}_P^{i^*}$  and  $\mathbf{L}_Q^j = \mathbf{L}_Q^{i^*}$ . This in turn yields  $\mathbf{L}_P^{i^*} \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^{i^*} \downarrow_{P^{\text{inp}}}, P^R)$  and  $\mathbf{L}_Q^{i^*} \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^{i^*} \downarrow_{Q^{\text{inp}}}, Q^R)$ , which by Theorem 3 entails that  $(\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*})$  is a complete labelling of  $AF$ . However, since  $\mathbf{L}_P^{i^*} \sqsubseteq \mathbf{L}_P$  and  $\mathbf{L}_Q^{i^*} \sqsubseteq \mathbf{L}_Q$  it holds that  $(\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*}) \sqsubseteq \mathbf{L}$ , and since  $\mathbf{L}_P \not\sqsubseteq \mathbf{L}_P^2$  it must also be the case that  $\mathbf{L} \not\sqsubseteq (\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*})$ , contradicting the fact that  $\mathbf{L}$  is the grounded labelling of  $AF$ .  $\square$

**Theorem 6.** Given an argumentation framework  $AF = (Ar, \text{att})$ , let  $\mathbf{L}$  be a preferred labelling of  $AF$ . For any set  $P \subseteq Ar$ ,  $\mathbf{L} \downarrow_P \in \mathbf{FPR}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L} \downarrow_{P^{\text{inp}}}, P^R)$ .

**Proof.** Let  $Q = Ar \setminus P$ . Since  $\mathbf{L}$  is a preferred labelling of  $AF$  and thus complete, letting  $\mathbf{L}_Q = \mathbf{L} \downarrow_Q$  and  $\mathbf{L}_P = \mathbf{L} \downarrow_P$  we have by Theorem 3 that  $\mathbf{L}_Q \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P \downarrow_{Q^{\text{inp}}}, Q^R)$  and  $\mathbf{L}_P \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q \downarrow_{P^{\text{inp}}}, P^R)$ . Suppose by contradiction that  $\mathbf{L}_P$  is not a maximal element w.r.t.  $\sqsubseteq$  in  $\mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q \downarrow_{P^{\text{inp}}}, P^R)$ , and let  $\mathbf{L}_P^2$  be a maximal element such that  $\mathbf{L}_P \sqsubseteq \mathbf{L}_P^2$  (with  $\mathbf{L}_P^2 \not\sqsubseteq \mathbf{L}_P$ ). Since  $\mathbf{L}_P \downarrow_{Q^{\text{inp}}} \sqsubseteq \mathbf{L}_P^2 \downarrow_{Q^{\text{inp}}}$ , by Proposition 7 (first point) there is a labelling  $\mathbf{L}_Q^2 \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^2 \downarrow_{Q^{\text{inp}}}, Q^R)$  such that  $\mathbf{L}_Q \sqsubseteq \mathbf{L}_Q^2$ . Since in particular  $\mathbf{L}_Q \downarrow_{P^{\text{inp}}} \sqsubseteq \mathbf{L}_Q^2 \downarrow_{P^{\text{inp}}}$ , again by Proposition 7 there is a labelling  $\mathbf{L}_P^3 \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^2 \downarrow_{P^{\text{inp}}}, P^R)$  such that  $\mathbf{L}_P^2 \sqsubseteq \mathbf{L}_P^3$ . Iterating the same reasoning, for any  $i > 2$  we get  $\exists \mathbf{L}_P^i \in \mathbf{Fco}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^{i-1} \downarrow_{P^{\text{inp}}}, P^R)$  such that  $\mathbf{L}_P \sqsubseteq \mathbf{L}_P^2 \sqsubseteq \dots \sqsubseteq \mathbf{L}_P^{i-1} \sqsubseteq \mathbf{L}_P^i$ , and  $\exists \mathbf{L}_Q^i \in \mathbf{Fco}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^i \downarrow_{Q^{\text{inp}}}, Q^R)$  such that  $\mathbf{L}_Q \sqsubseteq \mathbf{L}_Q^2 \sqsubseteq \dots \sqsubseteq \mathbf{L}_Q^{i-1} \sqsubseteq \mathbf{L}_Q^i$ . Since the argumentation framework is finite, there must be an  $i^*$  such that  $\forall j > i^* \mathbf{L}_P^j =$

$\mathbf{L}_P^{i^*}$  and  $\mathbf{L}_Q^j = \mathbf{L}_Q^{i^*}$ . This in turn yields  $\mathbf{L}_P^{i^*} \in F_{\mathbf{CO}}(AF \downarrow_P, P^{\text{inp}}, \mathbf{L}_Q^{i^*} \downarrow_{P^{\text{inp}}}, P^R)$  and  $\mathbf{L}_Q^{i^*} \in F_{\mathbf{CO}}(AF \downarrow_Q, Q^{\text{inp}}, \mathbf{L}_P^{i^*} \downarrow_{Q^{\text{inp}}}, Q^R)$ , which by [Theorem 3](#) entails that  $(\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*})$  is a complete labelling of  $AF$ . However, since  $\mathbf{L}_P \subseteq \mathbf{L}_P^{i^*}$  and  $\mathbf{L}_Q \subseteq \mathbf{L}_Q^{i^*}$  we have that  $\mathbf{L} \subseteq (\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*})$ , and since  $\mathbf{L}_P^2 \not\subseteq \mathbf{L}_P$  it must also be the case that  $(\mathbf{L}_P^{i^*} \cup \mathbf{L}_Q^{i^*}) \not\subseteq \mathbf{L}$ , contradicting the fact that  $\mathbf{L}$  is a preferred labelling of  $AF$ .  $\square$

**Lemma 3.** *Let  $\mathbf{S}$  be a complete-compatible semantics which is top-down decomposable, with the canonical local function  $F_{\mathbf{S}}$ . Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , consider a labelling  $\mathbf{L} \in F_{\mathbf{S}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and let  $P \subseteq Ar$  be an arbitrary set of arguments of  $AF$ . Then, letting  $P^{\text{F-inp}} \triangleq P^{\text{inp}} \cup \{A \in \mathcal{J} \mid \exists B \in P, (A, B) \in R_{\mathcal{J}}\}$  and  $P_F^R \triangleq P^R \cup (R_{\mathcal{J}} \cap (\mathcal{J} \times P))$ , it holds that  $\mathbf{L} \downarrow_P \in F_{\mathbf{S}}(AF \downarrow_P, P^{\text{F-inp}}, (\mathbf{L} \cup L_{\mathcal{J}}) \downarrow_{P^{\text{F-inp}}}, P_F^R)$ .*

**Proof.** Let  $AF' = (Ar \cup \mathcal{J}', att \cup R'_{\mathcal{J}})$  be the standard argumentation w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Since  $\mathbf{L} \in F_{\mathbf{S}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , according to the definition of canonical local function (see [Definition 13](#)) it holds that  $\exists \mathbf{L}' \in F_{\mathbf{S}}(AF')$  such that  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ . By top-down decomposability of  $\mathbf{S}$  applied to  $AF'$  and  $P$ , it must be the case that  $\mathbf{L}' \downarrow_P \in F_{\mathbf{S}}(AF' \downarrow_P, (P^{\text{inp}})', \mathbf{L}' \downarrow_{(P^{\text{inp}})', (P^R)'})$ , where we use  $(P^{\text{inp}})'$  and  $(P^R)'$  to denote  $P^{\text{inp}}$  and  $P^R$  in the context of  $AF'$ , respectively. Then the conclusion follows by observing that  $\mathbf{L}' \downarrow_P = \mathbf{L} \downarrow_P$ ,  $AF' \downarrow_P = AF \downarrow_P$ ,  $(P^{\text{inp}})' = P^{\text{F-inp}}$ ,  $\mathbf{L}' \downarrow_{(P^{\text{inp}})' } = L_{\mathcal{J}} \downarrow_{(P^{\text{F-inp}} \cap \mathcal{J})} \cup \mathbf{L} \downarrow_{P^{\text{inp}}}$ , and  $(P^R)' = P_F^R$ .  $\square$

**Theorem 7.** *Grounded and preferred semantics are decomposable w.r.t.  $\mathcal{F}_{\text{USCC}}$ .*

**Proof.** Since grounded and preferred semantics are top-down decomposable, we have only to prove bottom-up decomposability w.r.t.  $\mathcal{F}_{\text{USCC}}$ .

Let us first consider grounded semantics. Given an argumentation framework  $AF = (Ar, att)$ , consider a partition  $\{P_1, \dots, P_n\} \in \mathcal{F}_{\text{USCC}}(AF)$ , and let  $L_{P_1}, \dots, L_{P_n}$  be  $n$  labellings of  $P_1, \dots, P_n$  such that  $\forall i \ L_{P_i} \in F_{\mathbf{GR}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}}, P_i^R)$ . Letting  $\mathbf{L} = L_{P_1} \cup \dots \cup L_{P_n}$ , we prove that  $\mathbf{L}$  is the grounded labelling of  $AF$ . Since  $F_{\mathbf{GR}}$  returns a labelling of  $F_{\mathbf{CO}}$  and complete semantics is bottom-up decomposable, we know that  $\mathbf{L}$  is a complete labelling of  $AF$ : assume by contradiction that it is not the grounded labelling, thus letting  $\mathbf{L}'$  be the grounded labelling of  $AF$  it must be the case that  $\mathbf{L}' \subseteq \mathbf{L}$  and  $\mathbf{L}' \neq \mathbf{L}$ . Taking into account that the graph obtained by considering strongly-connected components as single nodes is acyclic, there must be a strongly connected component  $S \in \text{SCCS}_{AF}$  such that  $\mathbf{L}' \downarrow_S \subseteq \mathbf{L} \downarrow_S$  and  $\mathbf{L}' \downarrow_S \neq \mathbf{L} \downarrow_S$ , while  $\mathbf{L}' \downarrow_{S^{\text{inp}}} = \mathbf{L} \downarrow_{S^{\text{inp}}}$  (otherwise, considering a strongly connected component  $S_1 \in \text{SCCS}_{AF}$  such that  $\mathbf{L}' \downarrow_{S_1} \neq \mathbf{L} \downarrow_{S_1}$ , from  $\mathbf{L}' \downarrow_{S_1^{\text{inp}}} \neq \mathbf{L} \downarrow_{S_1^{\text{inp}}}$  we have  $\mathbf{L}' \downarrow_{S_2} \neq \mathbf{L} \downarrow_{S_2}$  where  $S_2$  is a strongly connected component such that  $S_1^{\text{inp}} \cap S_2 \neq \emptyset$ , and iterating this step we either obtain an infinite number of strongly connected components, which is impossible since  $AF$  is finite, or end up with a cycle of strongly connected components, which is impossible as well). By top-down decomposability of grounded semantics we have  $\mathbf{L}' \downarrow_S \in F_{\mathbf{GR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L}' \downarrow_{S^{\text{inp}}}, S^R) = F_{\mathbf{GR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$ . Now, according to the definition of  $\mathcal{F}_{\text{USCC}}$ , there is an element  $P_k$  of the partition such that  $S \subseteq P_k$ , and by [Lemma 3](#) applied to  $L_{P_k} \in F_{\mathbf{GR}}(AF \downarrow_{P_k}, P_k^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_k^{\text{inp}}}, P_k^R)$  it must be the case that  $\mathbf{L} \downarrow_S \in F_{\mathbf{GR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$ . However, this contradicts the two conditions  $\mathbf{L}' \downarrow_S \in F_{\mathbf{GR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$  and  $\mathbf{L}' \downarrow_S \neq \mathbf{L} \downarrow_S$ , since  $F_{\mathbf{GR}}$  always returns a unique labelling.

The proof for preferred semantics is similar. Given an argumentation framework  $AF = (Ar, att)$ , consider a partition  $\{P_1, \dots, P_n\} \in \mathcal{F}_{\text{USCC}}(AF)$ , and let  $L_{P_1}, \dots, L_{P_n}$  be  $n$  labellings of  $P_1, \dots, P_n$  such that  $\forall i \ L_{P_i} \in F_{\mathbf{PR}}(AF \downarrow_{P_i}, P_i^{\text{inp}}, (\bigcup_{j=1 \dots n, j \neq i} L_{P_j}) \downarrow_{P_i^{\text{inp}}}, P_i^R)$ . Letting  $\mathbf{L} = L_{P_1} \cup \dots \cup L_{P_n}$ , we prove that  $\mathbf{L} \in \mathbf{LPR}(AF)$ . By definition of  $F_{\mathbf{PR}}$  and bottom-up decomposability of complete semantics, we know that  $\mathbf{L}$  is a complete labelling of  $AF$ : assume by contradiction that it is not preferred, thus  $\exists \mathbf{L}' \in \mathbf{LPR}(AF)$  such that  $\mathbf{L} \subseteq \mathbf{L}'$  and  $\mathbf{L} \neq \mathbf{L}'$ . Taking into account that the graph obtained by considering strongly-connected components as single nodes is acyclic, there must be a strongly connected component  $S \in \text{SCCS}_{AF}$  such that  $\mathbf{L} \downarrow_S \subseteq \mathbf{L}' \downarrow_S$  and  $\mathbf{L} \downarrow_S \neq \mathbf{L}' \downarrow_S$ , while  $\mathbf{L} \downarrow_{S^{\text{inp}}} = \mathbf{L}' \downarrow_{S^{\text{inp}}}$ . By top-down decomposability of preferred semantics we have  $\mathbf{L}' \downarrow_S \in F_{\mathbf{PR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L}' \downarrow_{S^{\text{inp}}}, S^R) = F_{\mathbf{PR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$ . According to the definition of  $\mathcal{F}_{\text{USCC}}$ , there is a  $k$  such that  $S \subseteq P_k$ , and by [Lemma 3](#) it holds that  $\mathbf{L} \downarrow_S \in F_{\mathbf{PR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$ . However, from  $\mathbf{L}' \downarrow_S \in F_{\mathbf{PR}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R) \subseteq F_{\mathbf{CO}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$ ,  $\mathbf{L} \downarrow_S \subseteq \mathbf{L}' \downarrow_S$  and  $\mathbf{L} \downarrow_S \neq \mathbf{L}' \downarrow_S$  we get a contradiction, since  $\mathbf{L} \downarrow_S$  should be a maximal labelling in  $F_{\mathbf{CO}}(AF \downarrow_S, S^{\text{inp}}, \mathbf{L} \downarrow_{S^{\text{inp}}}, S^R)$  by definition of  $F_{\mathbf{PR}}$ .  $\square$

**Proposition 8.** *The canonical local function of ideal semantics is defined as  $F_{\mathbf{ID}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L}^*\}$ , where  $\mathbf{L}^*$  is the maximal (w.r.t.  $\subseteq$ ) labelling in  $F_{\mathbf{CO}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that for each  $\mathbf{L}_P \in F_{\mathbf{PR}}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  it holds that  $\mathbf{L}^* \subseteq \mathbf{L}_P$ .*

**Proof.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , let  $AF' = (Ar \cup \mathcal{J}', att \cup R'_{\mathcal{J}})$  be the standard argumentation framework w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Taking into account [Proposition 2](#), it is immediate to see that

$$F_{\mathbf{CO}}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset) = F_{\mathbf{PR}}(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset) = \{ \{ (A', \text{in}) \mid A \in \text{out}(L_{\mathcal{J}}) \} \cup L_{\mathcal{J}} \} \quad (\text{A.1})$$

As a preliminary result, we prove that

$$\begin{aligned} & \{ \mathbf{L}' \downarrow_{Ar} \mid \mathbf{L}' \in \mathbf{Lco}(AF') \wedge \forall \mathbf{L}'_p \in \mathbf{Lpr}(AF') \mathbf{L}' \sqsubseteq \mathbf{L}'_p \} \\ &= \{ \mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mid \forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L} \sqsubseteq \mathbf{L}_p \} \end{aligned} \quad (\text{A.2})$$

- As to the first direction of this proof, consider a labelling  $\mathbf{L}' \in \mathbf{Lco}(AF')$  such that  $\forall \mathbf{L}'_p \in \mathbf{Lpr}(AF') \mathbf{L}' \sqsubseteq \mathbf{L}'_p$ . Since complete semantics is top-down decomposable and, by Proposition 2,  $\mathbf{L}' \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ ,  $\mathbf{L}' \downarrow_{Ar} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . Moreover, Proposition 2 and condition (A.1) entail that  $\mathbf{L}' \downarrow_{\mathcal{J}'} \in Fpr(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$ , thus by bottom-up decomposability of preferred semantics w.r.t.  $\mathcal{F}_{USCC}$  it must be the case that  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) (\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}_p) \in \mathbf{Lpr}(AF')$ . As a consequence, by the hypothesis on  $\mathbf{L}'$  it holds that  $\mathbf{L}' \sqsubseteq (\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}_p)$ , entailing in particular  $\mathbf{L}' \downarrow_{Ar} \sqsubseteq \mathbf{L}_p$ .
- As to the other direction, consider a labelling  $\mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L} \sqsubseteq \mathbf{L}_p$ . By bottom-up decomposability of complete semantics and condition (A.1) there is a labelling  $\mathbf{L}' \in \mathbf{Lco}(AF')$  such that  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ . Moreover,  $\forall \mathbf{L}'_p \in \mathbf{Lpr}(AF')$  Proposition 2 entails that  $\mathbf{L}'_p \downarrow_{\mathcal{J}'} = \mathbf{L}' \downarrow_{\mathcal{J}'}$ , and by top-down decomposability of preferred semantics  $\mathbf{L}'_p \downarrow_{Ar} \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  thus  $\mathbf{L} \sqsubseteq \mathbf{L}'_p \downarrow_{Ar}$ . As a consequence,  $(\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}' \downarrow_{Ar}) \sqsubseteq (\mathbf{L}'_p \downarrow_{\mathcal{J}'} \cup \mathbf{L}'_p \downarrow_{Ar})$ , i.e.  $\mathbf{L}' \sqsubseteq \mathbf{L}'_p$ .

Let us now prove the desired result. Let  $\mathbf{L}'_{ID}$  be the ideal labelling of  $AF'$ : we prove that  $\{\mathbf{L}'_{ID} \downarrow_{Ar}\}$  coincides with  $\{\mathbf{L} \mid \mathbf{L} \text{ is maximal in } Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \text{ such that } \forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L} \sqsubseteq \mathbf{L}_p\}$ . Notice that we do not assume that the definition of  $F_{ID}$  is well founded, i.e. that there is a unique maximal labelling in the latter set: this is obtained as a by-product of the proof.

First consider the ideal labelling  $\mathbf{L}'_{ID}$ . According to (A.2),  $\mathbf{L}'_{ID} \downarrow_{Ar} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L}'_{ID} \downarrow_{Ar} \sqsubseteq \mathbf{L}_p$ . Assume by contradiction that  $\mathbf{L}'_{ID} \downarrow_{Ar}$  is not maximal, i.e. there is a labelling  $\mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L} \sqsubseteq \mathbf{L}_p$ ,  $\mathbf{L}'_{ID} \downarrow_{Ar} \sqsubseteq \mathbf{L}$  and  $\mathbf{L}'_{ID} \downarrow_{Ar} \neq \mathbf{L}$ . According to (A.2), there is a labelling  $\mathbf{L}' \in \mathbf{Lco}(AF')$  such that  $\forall \mathbf{L}'_p \in \mathbf{Lpr}(AF') \mathbf{L}' \sqsubseteq \mathbf{L}'_p$  and  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ . However, we can write  $\mathbf{L}'_{ID}$  as  $\mathbf{L}'_{ID} \downarrow_{\mathcal{J}'} \cup \mathbf{L}'_{ID} \downarrow_{Ar}$ , and since  $\mathbf{L}'_{ID} \downarrow_{\mathcal{J}'} = \mathbf{L}' \downarrow_{\mathcal{J}'}$  (Proposition 2) and  $\mathbf{L}'_{ID} \downarrow_{Ar} \sqsubseteq \mathbf{L}$ ,  $\mathbf{L}'_{ID} \sqsubseteq (\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L})$ , which is in turn equal to  $\mathbf{L}'$  since  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ , i.e.  $\mathbf{L}'_{ID} \sqsubseteq \mathbf{L}'$ . Taking into account that  $\mathbf{L}'_{ID} \downarrow_{Ar} \neq \mathbf{L}$  it is also the case that  $\mathbf{L}'_{ID} \neq \mathbf{L}'$ , contradicting the maximality of  $\mathbf{L}'_{ID}$ .

Let us turn to the proof of the reverse condition. Let  $\mathbf{L}$  be a maximal labelling among those in  $Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  that satisfy the condition  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L} \sqsubseteq \mathbf{L}_p$ . According to (A.2), there is a labelling  $\mathbf{L}' \in \mathbf{Lco}(AF')$  such that  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$  and  $\forall \mathbf{L}'_p \in \mathbf{Lpr}(AF') \mathbf{L}' \sqsubseteq \mathbf{L}'_p$ . Assume by contradiction that  $\mathbf{L}'$  is not maximal, i.e. there is a labelling  $\mathbf{L}'' \in \mathbf{Lco}(AF')$  such that  $\forall \mathbf{L}'_p \in \mathbf{Lpr}(AF') \mathbf{L}'' \sqsubseteq \mathbf{L}'_p$ ,  $\mathbf{L}' \sqsubseteq \mathbf{L}''$  and  $\mathbf{L}' \neq \mathbf{L}''$ . Note in particular that  $\mathbf{L}' \downarrow_{Ar} \sqsubseteq \mathbf{L}'' \downarrow_{Ar}$ , and since, by Proposition 2,  $\mathbf{L}' \downarrow_{\mathcal{J}'} = \mathbf{L}'' \downarrow_{\mathcal{J}'}$ , it must be the case that  $\mathbf{L}' \downarrow_{Ar} \neq \mathbf{L}'' \downarrow_{Ar}$ . According to (A.2),  $\mathbf{L}'' \downarrow_{Ar} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $\forall \mathbf{L}_p \in Fpr(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mathbf{L}'' \downarrow_{Ar} \sqsubseteq \mathbf{L}_p$ . However, since  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ , it holds that  $\mathbf{L} \sqsubseteq \mathbf{L}'' \downarrow_{Ar}$  and  $\mathbf{L} \neq \mathbf{L}'' \downarrow_{Ar}$ , contradicting the maximality of  $\mathbf{L}$  among the labellings of  $Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ .  $\square$

**Proposition 9.** *The canonical local function of semi-stable semantics is defined as  $F_{SST}(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \triangleq \{\mathbf{L} \mid \mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \text{ such that } \text{undec}(\mathbf{L}) \text{ is minimal w.r.t. set inclusion}\}$ .*

**Proof.** Given an argumentation framework with input  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , let  $AF' = (Ar \cup \mathcal{J}', att \cup R'_{\mathcal{J}})$  be the standard argumentation w.r.t.  $(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . It is immediate to see that

$$Fco(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset) = \{ \{ (A', in) \mid A \in out(L_{\mathcal{J}}) \} \cup L_{\mathcal{J}} \} \quad (\text{A.3})$$

We have to show that  $\{\mathbf{L}' \downarrow_{Ar} \mid \mathbf{L}' \in \mathbf{L}_{SST}(AF')\} = \{\mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}}) \mid \text{undec}(\mathbf{L}) \text{ is minimal}\}$ .

First, given a semi-stable labelling  $\mathbf{L}'$  of  $AF'$ , by top-down decomposability of complete semantics  $\mathbf{L}' \downarrow_{Ar} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , where we exploit the fact that, by Proposition 2,  $\mathbf{L}' \downarrow_{\mathcal{J}} = L_{\mathcal{J}}$ . Assume by contradiction that  $\text{undec}(\mathbf{L}' \downarrow_{Ar})$  is not minimal, i.e. there is a labelling  $\mathbf{L}_m \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\text{undec}(\mathbf{L}_m) \subsetneq \text{undec}(\mathbf{L}' \downarrow_{Ar})$ . By Proposition 2 and condition (A.3)  $\mathbf{L}' \downarrow_{\mathcal{J}'} \in Fco(AF' \downarrow_{\mathcal{J}'}, \emptyset, \emptyset, \emptyset)$ , thus, letting  $\mathbf{L}'' = (\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}_m)$ , by bottom-up decomposability of complete semantics  $\mathbf{L}'' \in \mathbf{Lco}(AF')$ . However, since  $\text{undec}(\mathbf{L}_m) \subsetneq \text{undec}(\mathbf{L}' \downarrow_{Ar})$  it must also be the case that  $\text{undec}(\mathbf{L}'') \subsetneq \text{undec}(\mathbf{L}' \downarrow_{\mathcal{J}'} \cup \mathbf{L}' \downarrow_{Ar})$ , i.e.  $\text{undec}(\mathbf{L}'') \subsetneq \text{undec}(\mathbf{L}')$ . This contradicts the fact that  $\mathbf{L}'$  is a semi-stable labelling of  $AF'$ , and thus  $\text{undec}(\mathbf{L}')$  should be minimal among the complete labellings of  $AF$ .

Let us now consider a labelling  $\mathbf{L} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  such that  $\text{undec}(\mathbf{L})$  is minimal. By bottom-up decomposability of complete semantics and condition (A.3), there is a labelling  $\mathbf{L}' \in \mathbf{Lco}(AF')$  such that  $\mathbf{L}' \downarrow_{Ar} = \mathbf{L}$ . Assume by contradiction that  $\mathbf{L}' \notin \mathbf{L}_{SST}(AF')$ , i.e. there is a labelling  $\mathbf{L}'' \in \mathbf{Lco}(AF')$  such that  $\text{undec}(\mathbf{L}'') \subsetneq \text{undec}(\mathbf{L}')$ . Since, by Proposition 2,  $\mathbf{L}'' \downarrow_{\mathcal{J}'} = \mathbf{L}' \downarrow_{\mathcal{J}'}$ , it must be the case that  $\text{undec}(\mathbf{L}'' \downarrow_{Ar}) \subsetneq \text{undec}(\mathbf{L}' \downarrow_{Ar}) = \text{undec}(\mathbf{L})$ . Moreover, by top-down decomposability of complete semantics and Proposition 2 it holds that  $\mathbf{L}'' \downarrow_{Ar} \in Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ . These conditions, however, contradict the hypothesis that  $\text{undec}(\mathbf{L})$  is minimal among the labellings in  $Fco(AF, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ .  $\square$

### A.3. Proofs of Section 5

**Lemma 4.** *Given a set of arguments  $\mathcal{J}$ , two labellings  $L^1_{\mathcal{J}}, L^2_{\mathcal{J}} \in \mathcal{L}_{\mathcal{J}}$ , a set of arguments  $Args$  such that  $\mathcal{J} \cap Args = \emptyset$  and a relation  $R_{INP} \subseteq \mathcal{J} \times Args$ , if  $L^1_{\mathcal{J}} \sqsubseteq L^2_{\mathcal{J}}$  then  $\text{eff}_{Args}(\mathcal{J}, L^1_{\mathcal{J}}, R_{INP}) \sqsubseteq \text{eff}_{Args}(\mathcal{J}, L^2_{\mathcal{J}}, R_{INP})$ .*

**Proof.** We have to prove that if an argument  $A$  is labelled in by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^1_{\mathcal{J}}, R_{\text{INP}})$  then it is labelled in by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^2_{\mathcal{J}}, R_{\text{INP}})$ , and if it is labelled out by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^1_{\mathcal{J}}, R_{\text{INP}})$  then it is labelled out by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^2_{\mathcal{J}}, R_{\text{INP}})$ . In the first case, either  $A$  has not attackers in  $\mathcal{J}$  or all of its attackers in  $\mathcal{J}$  are labelled out by  $L^1_{\mathcal{J}}$ . Since  $L^1_{\mathcal{J}} \subseteq L^2_{\mathcal{J}}$ , all these attackers are labelled out also by  $L^2_{\mathcal{J}}$ , thus  $A$  is labelled in by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^2_{\mathcal{J}}, R_{\text{INP}})$ . In the second case,  $A$  has an attacker  $B \in \mathcal{J}$  which is labelled in by  $L^1_{\mathcal{J}}$ . Again, since  $L^1_{\mathcal{J}} \subseteq L^2_{\mathcal{J}}$  the attacker  $B$  is labelled in also by  $L^2_{\mathcal{J}}$ , thus  $A$  is labelled out by  $\text{eff}_{\text{Args}}(\mathcal{J}, L^2_{\mathcal{J}}, R_{\text{INP}})$ .  $\square$

**Proposition 10.** Every semantics  $\mathbf{S} \in \{\mathbf{AD}, \mathbf{CO}, \mathbf{ST}, \mathbf{GR}, \mathbf{PR}, \mathbf{ID}, \mathbf{SST}\}$  is effect-dictated.

**Proof.** By inspection of Definition 23 and Definition 24, it is easy to see that the proposition holds for  $\mathbf{S} \in \{\mathbf{AD}, \mathbf{CO}\}$ . As to the other semantics, according to Definition 25, Proposition 6 and Proposition 9 the definition of  $F_{\mathbf{ST}}, F_{\mathbf{GR}}, F_{\mathbf{PR}}, F_{\mathbf{SST}}$  select those labellings of  $F_{\mathbf{CO}}$  that satisfy a requirement which does not depend on the input to the local function (i.e. absence of undec-labelled arguments, minimality w.r.t.  $\sqsubseteq$ , maximality w.r.t.  $\sqsubseteq$  and minimality of undec arguments, respectively): thus the conclusion is entailed by the result for  $F_{\mathbf{CO}}$ . Finally, according to Proposition 8  $F_{\mathbf{ID}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  is completely determined by  $F_{\mathbf{CO}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$  and  $F_{\mathbf{PR}}(\text{AF}, \mathcal{J}, L_{\mathcal{J}}, R_{\mathcal{J}})$ , thus the result follows from the above ones for  $\mathbf{CO}$  and  $\mathbf{PR}$ .  $\square$

#### A.4. Proofs of Section 6

**Proposition 11.** Consider a semantics  $\mathbf{S}$  such that  $\mathbf{L}_{\mathbf{S}}(\text{AF}_{\emptyset}) = \{\emptyset\}$ . Given a set of arguments  $E$  and a labelling  $\mathbf{L}_E \in \mathcal{L}_E$ , it holds that  $\mathbf{S}\text{-eff}_E(\mathcal{M}_{\emptyset}, \mathbf{L}_E) = \{(A, \text{in}) \mid A \in E\}$ .

**Proof.** According to Definition 29,  $\mathbf{S}\text{-eff}_E(\mathcal{M}_{\emptyset}, \mathbf{L}_E) = \{\text{eff}_E(\emptyset, \emptyset, \emptyset) \mid \mathbf{L} \in F_{\mathbf{S}}(\text{AF}_{\emptyset}, \emptyset, \emptyset, \emptyset)\}$ . By Proposition 1,  $F_{\mathbf{S}}(\text{AF}_{\emptyset}, \emptyset, \emptyset, \emptyset) = \mathbf{L}_{\mathbf{S}}(\text{AF}_{\emptyset})$ , which by the hypothesis is nonempty. Thus,  $\mathbf{S}\text{-eff}_E(\mathcal{M}_{\emptyset}, \mathbf{L}_E) = \{\text{eff}_E(\emptyset, \emptyset, \emptyset)\} = \{(A, \text{in}) \mid A \in E\}$ .  $\square$

**Proposition 12.** Consider a semantics  $\mathbf{S}$  such that  $\mathbf{L}_{\mathbf{S}}(\text{AF}_{\emptyset}) = \{\emptyset\}$  and a replacement  $\mathcal{R} = (\text{AF}, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E = \emptyset$ . Letting  $\text{AF}_2 = T(\mathcal{R})$ , the following conditions are equivalent:

- $\mathcal{R}$  is  $\mathbf{S}$ -legitimate
- $\mathcal{R}$  is contextually  $\mathbf{S}$ -legitimate
- $|\mathbf{L}_{\mathbf{S}}(\text{AF})| > 0 \wedge |\mathbf{L}_{\mathbf{S}}(\text{AF}_2)| > 0$ , or  $\mathbf{L}_{\mathbf{S}}(\text{AF}) = \mathbf{L}_{\mathbf{S}}(\text{AF}_2) = \emptyset$
- $\mathcal{R}$  is  $\mathbf{S}$ -safe.

**Proof.** Since  $E = \emptyset$ , according to Definition 31  $\mathcal{M}_1 = (\text{AF}, \emptyset, \emptyset)$  and  $\mathcal{M}_2 = (\text{AF}_2, \emptyset, \emptyset)$ . Moreover, according to Definition 32 it turns out that  $\mathcal{L}_{\mathcal{R}}^{\mathbf{S}} = \{F_{\mathbf{S}}(\text{AF} \downarrow_{\emptyset}, \emptyset, \emptyset, \emptyset)\}$ , which by Proposition 1 is equal to  $\mathbf{L}_{\mathbf{S}}(\text{AF}_{\emptyset})$ , in turn equal to  $\{\emptyset\}$  by the hypothesis. Since it also holds that  $\mathcal{L}_E = \{\emptyset\}$ ,  $\mathcal{R}$  is  $\mathbf{S}$ -legitimate if and only if it is contextually  $\mathbf{S}$ -legitimate. In particular,  $\mathcal{R}$  is legitimate if and only if  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{S}$ -equivalent, which according to Definition 30 holds if and only if  $\mathbf{S}\text{-eff}_{\emptyset}(\mathcal{M}_1, \emptyset) = \mathbf{S}\text{-eff}_{\emptyset}(\mathcal{M}_2, \emptyset)$ . Following Definition 29,  $\mathbf{S}\text{-eff}_{\emptyset}(\mathcal{M}_1, \emptyset) = \{\text{eff}_{\emptyset}(\emptyset, \mathbf{L} \downarrow_{\emptyset}, \emptyset) \mid \mathbf{L} \in F_{\mathbf{S}}(\text{AF}, \emptyset, \emptyset, \emptyset)\}$ , which by Proposition 1 is equal to  $\{\text{eff}_{\emptyset}(\emptyset, \mathbf{L} \downarrow_{\emptyset}, \emptyset) \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\}$ , i.e.  $\{\emptyset\}$  if  $|\mathbf{L}_{\mathbf{S}}(\text{AF})| > 0$ ,  $\emptyset$  if  $\mathbf{L}_{\mathbf{S}}(\text{AF}) = \emptyset$ . By the same reasoning,  $\mathbf{S}\text{-eff}_{\emptyset}(\mathcal{M}_2, \emptyset) = \{\emptyset\}$  if  $|\mathbf{L}_{\mathbf{S}}(\text{AF}_2)| > 0$ ,  $\emptyset$  if  $\mathbf{L}_{\mathbf{S}}(\text{AF}_2) = \emptyset$ . As a consequence,  $\mathcal{R}$  is  $\mathbf{S}$ -legitimate if and only if  $|\mathbf{L}_{\mathbf{S}}(\text{AF})| > 0 \wedge |\mathbf{L}_{\mathbf{S}}(\text{AF}_2)| > 0$ , or  $\mathbf{L}_{\mathbf{S}}(\text{AF}) = \mathbf{L}_{\mathbf{S}}(\text{AF}_2) = \emptyset$ . Let us finally consider the last condition. According to Definition 32,  $\mathcal{R}$  is  $\mathbf{S}$ -safe if and only if  $\{\mathbf{L} \downarrow_{\emptyset} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{\emptyset} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_2)\}$ . It is easy to see that the first term of this equality is equal to  $\{\emptyset\}$  if  $|\mathbf{L}_{\mathbf{S}}(\text{AF})| > 0$ , it is equal to  $\emptyset$  if  $\mathbf{L}_{\mathbf{S}}(\text{AF}) = \emptyset$ . Analogously, the second term is equal to  $\{\emptyset\}$  if  $|\mathbf{L}_{\mathbf{S}}(\text{AF}_2)| > 0$ , it is equal to  $\emptyset$  if  $\mathbf{L}_{\mathbf{S}}(\text{AF}_2) = \emptyset$ . Thus  $\mathcal{R}$  is  $\mathbf{S}$ -safe if and only if the third condition holds.  $\square$

**Proposition 13.** Let  $\text{AF} = (\text{Ar}, \text{att})$  be an argumentation framework. Consider a sequence of replacements  $(R_1, R_2, \dots, R_n)$  where  $R_i = (\text{AF}_i, \mathcal{M}_{i,1}, \mathcal{M}_{i,2})$ ,  $E_i$  is the invariant set of  $R_i$ ,  $\text{AF}_1 = \text{AF}$  and, for any  $1 < i \leq n$ ,  $\text{AF}_i = T(\text{AF}_{i-1}, \mathcal{M}_{i-1,1}, \mathcal{M}_{i-1,2})$ . Let  $\text{AF}_*$  be the result of the sequence of replacements, i.e.  $\text{AF}_* = T(\text{AF}_n, \mathcal{M}_{n,1}, \mathcal{M}_{n,2})$ . If all replacements  $R_i$  are  $\mathbf{S}$ -safe, then letting  $E = E_1 \cap \dots \cap E_n$  it holds that  $\{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_E \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_*)\}$ . Moreover, any argument  $A \in E$  is skeptically/credulously justified according to  $\mathbf{S}$  in  $\text{AF}$  if and only if it is skeptically/credulously justified according to  $\mathbf{S}$  in  $\text{AF}_*$ .

**Proof.** In order to prove the thesis, we show by induction on  $i$  that for  $i \in \{1, \dots, n\}$   $\{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_{i+1})\}$ , where  $E^i = E_1 \cap \dots \cap E_i$ .

For  $i = 1$ , since  $\text{AF}_2 = T(\text{AF}, \mathcal{M}_{1,1}, \mathcal{M}_{1,2})$ ,  $E_1 = E^1$  and  $(\text{AF}, \mathcal{M}_{1,1}, \mathcal{M}_{1,2})$  is  $\mathbf{S}$ -safe it holds that  $\{\mathbf{L} \downarrow_{E^1} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{E^1} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_2)\}$ .

For  $i > 1$ , we assume inductively that  $\{\mathbf{L} \downarrow_{E^{i-1}} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{E^{i-1}} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_i)\}$ , and we prove that  $\{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_{i+1})\}$ . First, since  $E^i = E^{i-1} \cap E_i$  it holds that  $E^i \subseteq E^{i-1}$ , thus the inductive hypothesis yields  $\{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF})\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_i)\}$ . Since  $R_i$  is  $\mathbf{S}$ -safe, it holds that  $\{\mathbf{L} \downarrow_{E_i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_i)\} = \{\mathbf{L} \downarrow_{E_i} \mid \mathbf{L} \in \mathbf{L}_{\mathbf{S}}(\text{AF}_{i+1})\}$ . Taking into account

that  $E^i \subseteq E_i$ , the latter condition entails that  $\{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_S(AF_i)\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_S(AF_{i+1})\}$ . Summing up,  $\{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_S(AF)\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_S(AF_i)\} = \{\mathbf{L} \downarrow_{E^i} \mid \mathbf{L} \in \mathbf{L}_S(AF_{i+1})\}$ .

As to the last point, an argument  $A \in E$  is skeptically justified according to  $\mathbf{S}$  in  $AF$  if and only if  $\forall \mathbf{L} \in \mathbf{L}_S(AF), \mathbf{L} \downarrow_E(A) = \text{in}$ , and by the previous point this holds if and only if  $\forall \mathbf{L} \in \mathbf{L}_S(AF_*) , \mathbf{L} \downarrow_E(A) = \text{in}$ , i.e. if and only if  $A$  is skeptically justified in  $AF_*$ . The proof for credulous justification is analogous.  $\square$

#### A.5. Proofs of Section 7

The proof of [Theorem 8](#) requires two preliminary lemmas, proving that any semantics satisfies a property corresponding to top-down and bottom-up decomposability in the degenerate case of two subframeworks including an empty one.

**Lemma 9.** *Let  $\mathbf{S}$  be a semantics such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ ,  $AF = (Ar, att)$  be an argumentation framework, and  $E \subseteq Ar$  be a subset of its arguments. Let  $D = Ar \setminus E$  and  $\mathcal{M} = (AF \downarrow_D, R_{INP}, R_{OUTP})$ , where  $R_{INP} = att \cap (E \times D)$  and  $R_{OUTP} = att \cap (D \times E)$ . Given a labelling  $\mathbf{L} \in \mathbf{L}_S(AF)$ , let  $\mathbf{L}^E = \mathbf{L} \downarrow_E$  and  $\mathbf{L}^D = \mathbf{L} \downarrow_D$ . If  $D = \emptyset$ , then  $\mathbf{L}^E \in F_S(AF \downarrow_E, \mathcal{M}^{\text{outp}}, \mathbf{L}^D \downarrow_{\mathcal{M}^{\text{outp}}, R_{OUTP}})$  and  $\mathbf{L}^D \in F_S(AF \downarrow_D, \mathcal{M}^{\text{inp}}, \mathbf{L}^E \downarrow_{\mathcal{M}^{\text{inp}}, R_{INP}})$ .*

**Proof.** In the specific case that  $D = \emptyset$ , it obviously holds that  $E = Ar$ ,  $AF \downarrow_D = AF_\emptyset$ ,  $R_{INP} = \emptyset$ ,  $R_{OUTP} = \emptyset$ ,  $\mathcal{M}^{\text{inp}} = \emptyset$  and  $\mathcal{M}^{\text{outp}} = \emptyset$ . It is then easy to see that  $\mathbf{L}^D = \emptyset$  and  $\mathbf{L}^E = \mathbf{L} \in \mathbf{L}_S(AF)$ . By [Proposition 1](#),  $\mathbf{L}_S(AF) = F_S(AF, \emptyset, \emptyset, \emptyset)$ , which is equal to  $F_S(AF \downarrow_E, \mathcal{M}^{\text{outp}}, \mathbf{L}^D \downarrow_{\mathcal{M}^{\text{outp}}, R_{OUTP}})$ . Moreover,  $\mathbf{L}^D = \emptyset \in \mathbf{L}_S(AF_\emptyset)$  by the hypothesis, and again by [Proposition 1](#) it holds that  $\mathbf{L}_S(AF_\emptyset) = F_S(AF_\emptyset, \emptyset, \emptyset, \emptyset)$ , where the latter is equal to  $F_S(AF \downarrow_D, \mathcal{M}^{\text{inp}}, \mathbf{L}^E \downarrow_{\mathcal{M}^{\text{inp}}, R_{INP}})$ .  $\square$

**Lemma 10.** *Let  $\mathbf{S}$  be a semantics such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ ,  $AF = (Ar, att)$  be an argumentation framework, and  $E \subseteq Ar$  be a subset of its arguments. Let  $D = Ar \setminus E$  and  $\mathcal{M} = (AF \downarrow_D, R_{INP}, R_{OUTP})$ , where  $R_{INP} = att \cap (E \times D)$  and  $R_{OUTP} = att \cap (D \times E)$ . Given two labellings  $\mathbf{L}^E$  and  $\mathbf{L}^D$  such that  $\mathbf{L}^E \in F_S(AF \downarrow_E, \mathcal{M}^{\text{outp}}, \mathbf{L}^D \downarrow_{\mathcal{M}^{\text{outp}}, R_{OUTP}})$  and  $\mathbf{L}^D \in F_S(AF \downarrow_D, \mathcal{M}^{\text{inp}}, \mathbf{L}^E \downarrow_{\mathcal{M}^{\text{inp}}, R_{INP}})$ , if  $D = \emptyset$  then  $(\mathbf{L}^E \cup \mathbf{L}^D) \in \mathbf{L}_S(AF)$ .*

**Proof.** In the specific case that  $D = \emptyset$ , it obviously holds that  $E = Ar$ ,  $AF \downarrow_D = AF_\emptyset$ ,  $R_{INP} = \emptyset$ ,  $R_{OUTP} = \emptyset$ ,  $\mathcal{M}^{\text{inp}} = \emptyset$  and  $\mathcal{M}^{\text{outp}} = \emptyset$ . Then the hypothesis yields  $\mathbf{L}^D \in F_S(AF_\emptyset, \emptyset, \emptyset, \emptyset)$ , which is equal to  $\mathbf{L}_S(AF_\emptyset)$  by [Proposition 1](#). Since by the hypothesis  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ , it must be the case that  $\mathbf{L}^D = \emptyset$ . Moreover,  $\mathbf{L}^E \in F_S(AF, \emptyset, \emptyset, \emptyset)$ , which is equal to  $\mathbf{L}_S(AF)$  by [Proposition 1](#). Thus,  $(\mathbf{L}^E \cup \mathbf{L}^D) = \mathbf{L}^E \in \mathbf{L}_S(AF)$ .  $\square$

**Theorem 8.** *Consider an effect-dictated semantics  $\mathbf{S}$  such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ . If  $\mathbf{S}$  is decomposable w.r.t. a partition selector  $\mathcal{F}$  then  $\mathbf{S}$  is strongly transparent w.r.t.  $\mathcal{F}$ .*

**Proof.** Consider a  $\mathcal{F}$ -preserving and contextually  $\mathbf{S}$ -legitimate replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , where  $\mathcal{M}_1 = (AF_1 \downarrow_{D_1}, R_{INP}^1, R_{OUTP}^1)$  and  $\mathcal{M}_2 = ((D_2, R_{D_2}), R_{INP}^2, R_{OUTP}^2)$ , and let  $AF_2 = T(AF_1, \mathcal{M}_1, \mathcal{M}_2)$ : we have to prove that the replacement is  $\mathbf{S}$ -safe, i.e. that  $\{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{L}_S(AF_1)\} = \{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{L}_S(AF_2)\}$ . Since  $(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  is  $\mathcal{F}$ -preserving,  $\{E, D_1\} \setminus \emptyset \in \mathcal{F}(AF_1)$  and  $\{E, D_2\} \setminus \emptyset \in \mathcal{F}(AF_2)$ .

First, in the particular case where  $E = \emptyset$  the replacement is  $\mathbf{S}$ -safe by [Proposition 12](#). Thus, in the remainder of the proof we assume  $E \neq \emptyset$ .

Let us first prove that  $\{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{L}_S(AF_1)\} \subseteq \{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{L}_S(AF_2)\}$ , i.e. that given an arbitrary labelling  $\mathbf{L}_1 \in \mathbf{L}_S(AF_1)$  there is a labelling  $\mathbf{L}_2 \in \mathbf{L}_S(AF_2)$  such that  $\mathbf{L}_1 \downarrow_E = \mathbf{L}_2 \downarrow_E$ . Let  $\mathbf{L}_1^E = \mathbf{L}_1 \downarrow_E$  and  $\mathbf{L}_1^{D_1} = \mathbf{L}_1 \downarrow_{D_1}$ . If  $D_1 \neq \emptyset$ , since  $\mathbf{S}$  is decomposable w.r.t.  $\mathcal{F}$  and  $\{E, D_1\} \in \mathcal{F}(AF_1)$ , by top-down decomposability it holds that  $\mathbf{L}_1^E \in F_S(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}, R_{OUTP}^1})$  and  $\mathbf{L}_1^{D_1} \in F_S(AF_1 \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}, R_{INP}^1})$ . In the case where  $D_1 = \emptyset$ , the same conditions are entailed by [Lemma 9](#). Note that  $\mathbf{L}_1^E \in \mathcal{L}_S^{\mathcal{R}}$ . Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{S}$ -equivalent under  $\mathcal{L}_S^{\mathcal{R}}$ , there is a labelling  $\mathbf{L}_2^{D_2} \in F_S(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}, R_{INP}^2})$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{OUTP}^2}) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}, R_{OUTP}^1})$ . Taking into account that  $\mathbf{S}$  is effect-dictated, this condition entails that  $F_S(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}, R_{OUTP}^1}) = F_S(AF_1 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{OUTP}^2})$ , and since  $AF_1 \downarrow_E = AF_2 \downarrow_E$  it holds that  $\mathbf{L}_1^E \in F_S(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{OUTP}^2})$ . Now, we have two cases to consider. In case  $D_2 \neq \emptyset$ ,  $\{E, D_2\} \in \mathcal{F}(AF_2)$  and bottom-up decomposability w.r.t.  $\mathcal{F}$  entails that  $(\mathbf{L}_1^E \cup \mathbf{L}_2^{D_2}) \in \mathbf{L}_S(AF_2)$ . Otherwise, the same condition holds by [Lemma 10](#). In both cases, the conclusion follows by letting  $\mathbf{L}_2 = (\mathbf{L}_1^E \cup \mathbf{L}_2^{D_2})$ .

Taking into account that  $AF_1 = T(AF_2, \mathcal{M}_2, \mathcal{M}_1)$  and that also  $(AF_2, \mathcal{M}_2, \mathcal{M}_1)$  is a contextually  $\mathbf{S}$ -legitimate replacement, by a symmetric reasoning it can be proved that  $\{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{L}_S(AF_2)\} \subseteq \{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{L}_S(AF_1)\}$ .

Thus,  $\{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{L}_S(AF_1)\} \subseteq \{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{L}_S(AF_2)\}$  and  $\{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{L}_S(AF_2)\} \subseteq \{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{L}_S(AF_1)\}$  entail the desired conclusion.  $\square$

**Theorem 9.** *Let  $\mathbf{S}$  be an effect-dictated single-status semantics such that  $\mathbf{L}_S(AF_\emptyset) = \{\emptyset\}$ . Suppose that  $\mathbf{S}$  is top-down decomposable w.r.t. a partition selector  $\mathcal{F}$  and satisfies the following property: for any argumentation framework  $AF$  and any partition  $\{E, D\} \in \mathcal{F}(AF)$ ,*



letting  $\mathbf{L}$  be the labelling prescribed by  $\mathbf{S}$  for  $AF$ , if  $\mathbf{L}^E \in \mathfrak{L}_E$  and  $\mathbf{L}^D \in \mathfrak{L}_D$  are two labellings such that  $\mathbf{L}^E \in F_S(AF \downarrow_E, E^{\text{inp}}, \mathbf{L}^D \downarrow_{E^{\text{inp}}}, E^R)$  and  $\mathbf{L}^D \in F_S(AF \downarrow_D, D^{\text{inp}}, \mathbf{L}^E \downarrow_{D^{\text{inp}}}, D^R)$ , then  $\mathbf{L} \sqsubseteq \mathbf{L}^E \cup \mathbf{L}^D$ . Then  $\mathbf{S}$  is strongly transparent w.r.t.  $\mathcal{F}$ .

**Proof.** Consider a  $\mathcal{F}$ -preserving and contextually  $\mathbf{S}$ -legitimate replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , where  $\mathcal{M}_1 = (AF_1 \downarrow_{D_1}, R_{1\text{INP}}^1, R_{1\text{OUTP}}^1)$  and  $\mathcal{M}_2 = ((D_2, R_{D_2}), R_{2\text{INP}}^2, R_{2\text{OUTP}}^2)$ , and let  $AF_2 = T(AF_1, \mathcal{M}_1, \mathcal{M}_2)$ . We have to prove that the replacement is  $\mathbf{S}$ -safe, i.e. letting  $\mathbf{L}_1$  and  $\mathbf{L}_2$  be the labellings prescribed for  $AF_1$  and  $AF_2$ , respectively, that  $\mathbf{L}_1 \downarrow_E = \mathbf{L}_2 \downarrow_E$ .

As in the proof of Theorem 8, in the case that  $E = \emptyset$  the replacement is  $\mathbf{S}$ -safe by Proposition 12, thus in the remainder of the proof we assume  $E \neq \emptyset$ .

Let  $\mathbf{L}_1^E = \mathbf{L}_1 \downarrow_E$  and  $\mathbf{L}_1^{D_1} = \mathbf{L}_1 \downarrow_{D_1}$ . If  $D_1 \neq \emptyset$ ,  $\{E, D_1\} \in \mathcal{F}(AF_1)$  and since  $\mathbf{S}$  is top-down decomposable w.r.t.  $\mathcal{F}$  it holds that  $\mathbf{L}_1^E \in F_S(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1)$  and  $\mathbf{L}_1^{D_1} \in F_S(AF_1 \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{1\text{INP}}^1)$ . Otherwise, if  $D_1 = \emptyset$  then the same conditions hold by Lemma 9. Note that  $\mathbf{L}_1^E \in \mathfrak{L}_E^{\mathbf{S}}$ . Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{S}$ -equivalent under  $\mathfrak{L}_E^{\mathbf{S}}$ , there is a labelling  $\mathbf{L}_2^{D_2} \in F_S(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{2\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1)$ , entailing that  $F_S(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1) = F_S(AF_1 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2)$ . Taking into account that  $AF_1 \downarrow_E = AF_2 \downarrow_E$ , it holds that  $\mathbf{L}_1^E \in F_S(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2)$ . We have two cases to consider. If  $D_2 \neq \emptyset$ , then  $\{E, D_2\} \in \mathcal{F}(AF_2)$  and the hypothesis condition applied to  $\mathbf{L}_1^E$  and  $\mathbf{L}_2^{D_2}$  in  $AF_2$  yields  $\mathbf{L}_2 \sqsubseteq (\mathbf{L}_1^E \cup \mathbf{L}_2^{D_2})$ , thus in particular  $\mathbf{L}_2 \downarrow_E \sqsubseteq \mathbf{L}_1 \downarrow_E$ . In the other case,  $D_2 = \emptyset$  and by Lemma 10 it holds that  $(\mathbf{L}_1^E \cup \mathbf{L}_2^{D_2}) \in \mathfrak{L}_S(AF_2)$ . Since  $\mathbf{S}$  is single-status, it must be the case that  $(\mathbf{L}_1^E \cup \mathbf{L}_2^{D_2}) = \mathbf{L}_2$ , thus in particular  $\mathbf{L}_1 \downarrow_E = \mathbf{L}_2 \downarrow_E$ , which is a special case of the condition  $\mathbf{L}_2 \downarrow_E \sqsubseteq \mathbf{L}_1 \downarrow_E$ .

Taking into account that  $AF_1 = T(AF_2, \mathcal{M}_2, \mathcal{M}_1)$  and that also  $(AF_2, \mathcal{M}_2, \mathcal{M}_1)$  is a contextually  $\mathbf{S}$ -legitimate replacement, by a symmetric reasoning with  $\mathbf{L}_2$  it can be proved that there exists a labelling  $\mathbf{L}_1^{D_1}$  of  $D_1$  such that  $\mathbf{L}_1 \sqsubseteq (\mathbf{L}_2 \downarrow_E \cup \mathbf{L}_1^{D_1})$  and in particular  $\mathbf{L}_1 \downarrow_E \sqsubseteq \mathbf{L}_2 \downarrow_E$ .

Now, from  $\mathbf{L}_2 \downarrow_E \sqsubseteq \mathbf{L}_1 \downarrow_E$  and  $\mathbf{L}_1 \downarrow_E \sqsubseteq \mathbf{L}_2 \downarrow_E$  it turns out that  $\mathbf{L}_1 \downarrow_E = \mathbf{L}_2 \downarrow_E$ .  $\square$

## A.6. Proofs of Section 8

**Theorem 10.** Admissible semantics  $\mathbf{AD}$ , complete semantics  $\mathbf{CO}$  and stable semantics  $\mathbf{ST}$  are strongly transparent.

**Proof.** Since decomposability and strong transparency of a semantics are equivalent to decomposability w.r.t.  $\mathcal{F}_{\text{ALL}}$  and strong transparency w.r.t.  $\mathcal{F}_{\text{ALL}}$ , respectively, an immediate consequence of Theorem 8 is that an effect-dictated decomposable semantics is strongly transparent. Then the conclusion follows from Proposition 10 and Theorem 1 (for  $\mathbf{AD}$ ), Theorem 3 (for  $\mathbf{CO}$ ) and Theorem 4 (for  $\mathbf{ST}$ ).  $\square$

**Theorem 11.** Grounded semantics  $\mathbf{GR}$  is strongly transparent.

**Proof.** We prove that grounded semantics satisfies the hypotheses of Theorem 9 that ensure strong transparency. First, grounded semantics is single-status by definition. Moreover, we know by Proposition 10 that  $\mathbf{GR}$  is effect-dictated and by Theorem 5 that it is top-down decomposable. Finally, consider an argumentation framework  $AF$ , an arbitrary partition  $\{E, D\} \in \mathcal{F}_{\text{ALL}}(AF)$  and two labellings  $\mathbf{L}^E \in \mathfrak{L}_E$  and  $\mathbf{L}^D \in \mathfrak{L}_D$  such that  $\mathbf{L}^E \in F_{\mathbf{GR}}(AF \downarrow_E, E^{\text{inp}}, \mathbf{L}^D \downarrow_{E^{\text{inp}}}, E^R)$  and  $\mathbf{L}^D \in F_{\mathbf{GR}}(AF \downarrow_D, D^{\text{inp}}, \mathbf{L}^E \downarrow_{D^{\text{inp}}}, D^R)$ ; taking into account that  $F_{\mathbf{GR}}$  returns a subset of  $F_{\mathbf{CO}}$  (see Proposition 6), by bottom-up decomposability of complete semantics (see Theorem 3)  $(\mathbf{L}^E \cup \mathbf{L}^D)$  is a complete labelling of  $AF$ . Thus, letting  $\mathbf{L}$  be the grounded labelling of  $AF$ , by definition it must be the case that  $\mathbf{L} \sqsubseteq \mathbf{L}^E \cup \mathbf{L}^D$ .  $\square$

**Theorem 12.** For any contextually  $\mathbf{PR}$ -legitimate replacement  $\mathcal{R} = (AF, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , any argument  $A \in E$  is credulously justified according to  $\mathbf{PR}$  in  $AF$  if and only if it is credulously justified according to  $\mathbf{PR}$  in  $T(AF, \mathcal{M}_1, \mathcal{M}_2)$ .

**Proof.** First, if  $E = \emptyset$  then the claim is trivially verified, thus in the following we consider the case that  $E \neq \emptyset$ .

Assume that  $\mathcal{M}_1 = (AF \downarrow_{D_1}, R_{1\text{INP}}^1, R_{1\text{OUTP}}^1)$ ,  $AF_2 = T(AF, \mathcal{M}_1, \mathcal{M}_2)$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{D_2}, R_{2\text{INP}}^2, R_{2\text{OUTP}}^2)$ . Given an argument  $A \in E$  which is credulously justified in  $AF$ , there is a preferred labelling  $\mathbf{L}_1 \in \mathbf{LPR}(AF)$  such that  $\mathbf{L}_1(A) = \text{in}$ . Letting  $\mathbf{L}_1^E = \mathbf{L}_1 \downarrow_E$  and  $\mathbf{L}_1^{D_1} = \mathbf{L}_1 \downarrow_{D_1}$ , we have that  $\mathbf{L}_1^E \in F_{\mathbf{PR}}(AF \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1)$  and  $\mathbf{L}_1^{D_1} \in F_{\mathbf{PR}}(AF \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{1\text{INP}}^1)$ . In particular, if  $D_1 \neq \emptyset$  these conditions hold by top-down decomposability of preferred semantics (see Theorem 6), otherwise by Lemma 9. Note that, according to the first condition,  $\mathbf{L}_1^E \in \mathfrak{L}_E^{\mathbf{PR}}$ . Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{PR}$ -equivalent under  $\mathfrak{L}_E^{\mathbf{PR}}$ , there is a labelling  $\mathbf{L}_1^{D_2} \in F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{2\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1)$ , thus it is the case that  $F_{\mathbf{PR}}(AF \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{1\text{OUTP}}^1) = F_{\mathbf{PR}}(AF \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2)$ . Taking then into account that  $AF \downarrow_E = AF_2 \downarrow_E$ , it holds that  $\mathbf{L}_1^E \in F_{\mathbf{PR}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2)$ . Since  $F_{\mathbf{PR}}$  returns a subset of  $F_{\mathbf{CO}}$  (see Proposition 6), the previous conditions yield  $\mathbf{L}_1^E \in F_{\mathbf{CO}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{2\text{OUTP}}^2)$  and  $\mathbf{L}_1^{D_2} \in F_{\mathbf{CO}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{2\text{INP}}^2)$ , thus by bottom-up decomposability of complete semantics proved in Theorem 3 (if  $D_2 \neq \emptyset$ ) or by Lemma 10 (if  $D_2 = \emptyset$ ),  $(\mathbf{L}_1^E \cup \mathbf{L}_1^{D_2})$  is a complete labelling of  $AF_2$ . As a consequence, there is

a preferred labelling  $\mathbf{L}_2 \in \mathbf{LPR}(AF_2)$  such that  $(\mathbf{L}_1^E \cup \mathbf{L}_1^{D_2}) \sqsubseteq \mathbf{L}_2$ . Since  $A \in E$  and  $\mathbf{L}_1(A) = \text{in}$  it must also be the case that  $\mathbf{L}_2(A) = \text{in}$ , i.e.  $A$  is credulously justified under preferred semantics in  $AF_2$ .

Since the hypotheses are symmetric for  $AF$  and  $AF_2$ , the other direction of the proof is proved in the same way.  $\square$

**Theorem 13.** Any replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are homogeneously  $\mathbf{PR}$ -equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ , is  $\mathbf{PR}$ -safe.

**Proof.** First, if  $E = \emptyset$  then the replacement is  $\mathbf{PR}$ -safe by Proposition 12, thus in the following we consider the case that  $E \neq \emptyset$ .

Assume that  $\mathcal{M}_1 = (AF_1 \downarrow_{D_1}, R_{\text{INP}}^1, R_{\text{OUTP}}^1)$ ,  $AF_2 = T(AF_1, \mathcal{M}_1, \mathcal{M}_2)$  and  $\mathcal{M}_2 = (AF_2 \downarrow_{D_2}, R_{\text{INP}}^2, R_{\text{OUTP}}^2)$ . We have to prove that  $\{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{LPR}(AF_1)\} = \{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{LPR}(AF_2)\}$ . Similarly to previous proofs, we only prove that  $\{\mathbf{L}_1 \downarrow_E \mid \mathbf{L}_1 \in \mathbf{LPR}(AF_1)\} \subseteq \{\mathbf{L}_2 \downarrow_E \mid \mathbf{L}_2 \in \mathbf{LPR}(AF_2)\}$ , since the reverse inclusion relation can be proved in the same way by exploiting the fact that the hypotheses are symmetric for  $AF_1$  and  $AF_2$ .

Let  $\mathbf{L}_1 \in \mathbf{LPR}(AF_1)$  be an arbitrary preferred labelling of  $AF_1$ : we have to identify a labelling  $\mathbf{L}_2 \in \mathbf{LPR}(AF_2)$  such that  $\mathbf{L}_1 \downarrow_E = \mathbf{L}_2 \downarrow_E$ . Letting  $\mathbf{L}_1^E = \mathbf{L}_1 \downarrow_E$  and  $\mathbf{L}_1^{D_1} = \mathbf{L}_1 \downarrow_{D_1}$ , by top-down decomposability of preferred semantics proved in Theorem 6 (if  $D_1 \neq \emptyset$ ) or by Lemma 9 (if  $D_1 = \emptyset$ ),  $\mathbf{L}_1^E \in F_{\mathbf{PR}}(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$  and  $\mathbf{L}_1^{D_1} \in F_{\mathbf{PR}}(AF_1 \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$ . Note that, by the first condition,  $\mathbf{L}_1^E \in \mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ . Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{PR}$ -equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ , there is a labelling  $\mathbf{L}_1^{D_2} \in F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ , entailing that  $F_{\mathbf{PR}}(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) = F_{\mathbf{PR}}(AF_1 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ . Taking then into account that  $AF_1 \downarrow_E = AF_2 \downarrow_E$ , it holds that  $\mathbf{L}_1^E \in F_{\mathbf{PR}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ . Since  $F_{\mathbf{PR}}$  returns a subset of  $F_{\mathbf{CO}}$  (see Proposition 6), the previous conditions yield  $\mathbf{L}_1^E \in F_{\mathbf{CO}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$  and  $\mathbf{L}_1^{D_2} \in F_{\mathbf{CO}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ , thus by bottom-up decomposability of complete semantics proved in Theorem 3 (if  $D_2 \neq \emptyset$ ) or by Lemma 10 (if  $D_2 = \emptyset$ ),  $(\mathbf{L}_1^E \cup \mathbf{L}_1^{D_2})$  is a complete labelling of  $AF_2$ . We prove that  $\mathbf{L}_2 = (\mathbf{L}_1^E \cup \mathbf{L}_1^{D_2}) \in \mathbf{LPR}(AF_2)$ , which yields the desired conclusion. Assume by contradiction that  $\mathbf{L}_2$  is not a preferred labelling of  $AF_2$ : then there is a preferred labelling  $\mathbf{L}_2^* \in \mathbf{LPR}(AF_2)$  such that  $\mathbf{L}_2 \sqsubseteq \mathbf{L}_2^*$  and  $\mathbf{L}_2 \neq \mathbf{L}_2^*$ . Letting  $\mathbf{L}_2^{*E} = \mathbf{L}_2^* \downarrow_E$  and  $\mathbf{L}_2^{*D_2} = \mathbf{L}_2^* \downarrow_{D_2}$ , by top-down decomposability of preferred semantics (if  $D_2 \neq \emptyset$ ) or by Lemma 9 (if  $D_2 = \emptyset$ ), it holds that  $\mathbf{L}_2^{*E} \in F_{\mathbf{PR}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{*D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$  and  $\mathbf{L}_2^{*D_2} \in F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_2^{*E} \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ . Note that  $\mathbf{L}_2^{*E} \in \mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ . It must be the case that  $\mathbf{L}_1^E \sqsubseteq \mathbf{L}_2^{*E}$  and  $\mathbf{L}_1^E \neq \mathbf{L}_2^{*E}$ . The first condition is entailed by  $\mathbf{L}_2 \sqsubseteq \mathbf{L}_2^*$  and  $\mathbf{L}_2 \downarrow_E = \mathbf{L}_1^E$ . The second condition holds since otherwise  $\mathbf{L}_1^E = \mathbf{L}_2^{*E}$ , entailing both  $\mathbf{L}_1^{D_2}$  and  $\mathbf{L}_2^{*D_2}$  to belong to  $F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ : since  $\mathbf{L}_1^{D_2} \sqsubseteq \mathbf{L}_2^{*D_2}$  and  $F_{\mathbf{PR}}$  returns maximal (w.r.t.  $\sqsubseteq$ ) elements of  $F_{\mathbf{CO}}$ , we would have  $\mathbf{L}_1^{D_2} = \mathbf{L}_2^{*D_2}$  besides  $\mathbf{L}_1^E = \mathbf{L}_2^{*E}$ , violating the condition  $\mathbf{L}_2 \neq \mathbf{L}_2^*$ . Now, taking into account the hypothesis that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are homogeneously  $\mathbf{PR}$ -equivalent under  $\mathcal{L}_{\mathcal{R}}^{\mathbf{PR}}$ , from  $\mathbf{L}_1^{D_2} \in F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ ,  $\mathbf{L}_2^{*D_2} \in F_{\mathbf{PR}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_2^{*E} \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ ,  $\mathbf{L}_1^{D_2} \sqsubseteq \mathbf{L}_2^{*D_2}$ ,  $\mathbf{L}_1^E \sqsubseteq \mathbf{L}_2^{*E}$ ,  $\mathbf{L}_1^{D_1} \in F_{\mathbf{PR}}(AF_1 \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  and  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) = \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ , there is a labelling  $\mathbf{L}_1^{*D_1} \in F_{\mathbf{PR}}(AF_1 \downarrow_{D_1}, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_2^{*E} \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  such that  $\mathbf{L}_1^{D_1} \sqsubseteq \mathbf{L}_1^{*D_1}$  and  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{*D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) = \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{*D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ . Since  $\mathbf{L}_2^{*E} \in F_{\mathbf{PR}}(AF_2 \downarrow_E, \mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{*D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$  and  $AF_2 \downarrow_E = AF_1 \downarrow_E$ , the latter condition entails  $\mathbf{L}_2^{*E} \in F_{\mathbf{PR}}(AF_1 \downarrow_E, \mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{*D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ . Since  $F_{\mathbf{PR}}$  returns a subset of  $F_{\mathbf{CO}}$  and complete semantics is decomposable,  $\mathbf{L}_2^{*E} \cup \mathbf{L}_1^{*D_1}$  is a complete labelling of  $AF_1$  (again, in case  $D_1 = \emptyset$  this is a consequence of Lemma 10). However, from  $\mathbf{L}_1^E \sqsubseteq \mathbf{L}_2^{*E}$ ,  $\mathbf{L}_1^{D_1} \sqsubseteq \mathbf{L}_1^{*D_1}$  and  $\mathbf{L}_1^E \neq \mathbf{L}_2^{*E}$  it holds that  $\mathbf{L}_1 \sqsubseteq (\mathbf{L}_2^{*E} \cup \mathbf{L}_1^{*D_1})$  and  $\mathbf{L}_1 \neq (\mathbf{L}_2^{*E} \cup \mathbf{L}_1^{*D_1})$ , contradicting the fact that  $\mathbf{L}_1 \in \mathbf{LPR}(AF_1)$ , i.e. that it is a maximal (w.r.t.  $\sqsubseteq$ ) complete labelling of  $AF_1$ .  $\square$

**Lemma 5.** Consider two multipoles  $\mathcal{M}_1 = (AF_1, R_{\text{INP}}^1, R_{\text{OUTP}}^1)$  and  $\mathcal{M}_2 = (AF_2, R_{\text{INP}}^2, R_{\text{OUTP}}^2)$  w.r.t. a set  $E$  which are internally  $\mathbf{S}$ -homogeneous under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . If  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{S}$ -equivalent under  $\mathcal{L}'$ , then they are homogeneously  $\mathbf{S}$ -equivalent under  $\mathcal{L}'$ .

**Proof.** We have to prove the two conditions of Definition 34. We prove the first condition, since the other one can be obtained by a symmetric reasoning. Given  $\mathbf{L}_1^E, \mathbf{L}_2^E \in \mathcal{L}'$  such that  $\mathbf{L}_1^E \sqsubseteq \mathbf{L}_2^E$ , consider two labellings  $\mathbf{L}_1^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  and  $\mathbf{L}_2^{D_1} \in F_{\mathbf{S}}(AF_1, \mathcal{M}_1^{\text{inp}}, \mathbf{L}_2^E \downarrow_{\mathcal{M}_1^{\text{inp}}}, R_{\text{INP}}^1)$  such that  $\mathbf{L}_1^{D_1} \sqsubseteq \mathbf{L}_2^{D_1}$ . Consider also a labelling  $\mathbf{L}_2^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_1^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ . Since  $\mathbf{L}_1^{D_1} \sqsubseteq \mathbf{L}_2^{D_1}$ , by Lemma 4 it holds that  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_1^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1) \sqsubseteq \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ , thus  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) \sqsubseteq \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ . Moreover, since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are  $\mathbf{S}$ -equivalent under  $\mathcal{L}'$ , there is a labelling  $\mathbf{L}_2^{D_2} \in F_{\mathbf{S}}(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_2^E \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) = \text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}}, R_{\text{OUTP}}^1)$ , thus it holds that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2) \sqsubseteq \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}}, R_{\text{OUTP}}^2)$ . Summing up,  $\exists \mathbf{L}_1^{D_2} \in$



$F_S(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}_2^{\text{inp}}, R_{\text{INP}}^2}, \mathbf{L}_2^{D_2} \in F_S(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_2^{\text{inp}}, R_{\text{INP}}^2})$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_1^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{\text{OUTP}}^2}) \subseteq \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{\text{OUTP}}^2})$ , with  $\mathbf{L}_E^1, \mathbf{L}_E^2 \in \mathcal{L}'$  and  $\mathbf{L}_E^1 \sqsubseteq \mathbf{L}_E^2$ . The conclusion can then be derived from the fact that  $\mathcal{M}_2$  is internally **S**-homogeneous under  $\mathcal{L}'$ , since this entails that there is a labelling  $\mathbf{L}_2^{D_2} \in F_S(AF_2, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}_2^{\text{inp}}, R_{\text{INP}}^2})$  such that  $\text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{\text{OUTP}}^2}) = \text{eff}_E(\mathcal{M}_2^{\text{outp}}, \mathbf{L}_2^{D_2} \downarrow_{\mathcal{M}_2^{\text{outp}}, R_{\text{OUTP}}^2})$ , where the latter is equal to  $\text{eff}_E(\mathcal{M}_1^{\text{outp}}, \mathbf{L}_2^{D_1} \downarrow_{\mathcal{M}_1^{\text{outp}}, R_{\text{OUTP}}^1})$ , and  $\mathbf{L}_1^{D_2} \sqsubseteq \mathbf{L}_2^{D_2}$ .  $\square$

**Corollary 1.** Any replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$  and both  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are internally **PR**-homogeneous under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ , is **PR**-safe.

**Proof.** Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are internally **PR**-homogeneous under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$  and **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ , by Lemma 5 they are homogeneously **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ . The desired conclusion then follows from Theorem 13.  $\square$

**Theorem 14.** Preferred semantics **PR** is strongly transparent w.r.t.  $\mathcal{F}_{\text{USCC}}$ .

**Proof.** Immediate from Theorem 7 and Theorem 8.  $\square$

**Lemma 6.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . Then  $\forall \mathbf{L}_E \in \mathcal{L}', F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}}) = F_S(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$  for any  $\mathbf{S} \in \{\text{GR}, \text{PR}, \text{ID}, \text{SST}\}$ .

**Proof.** According to the relevant definitions,  $F_{\text{GR}}, F_{\text{PR}}, F_{\text{ID}}$  and  $F_{\text{SST}}$  return a subset of the labellings returned by  $F_{\text{CO}}$  and are always able to return at least a labelling. Then the conclusion follows by taking into account that  $\mathcal{M}$  is **CO**-univocal under  $\mathcal{L}'$ .  $\square$

**Lemma 7.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ , and let  $\mathbf{L}_E^1, \mathbf{L}_E^2$  be two labellings of  $\mathcal{L}'$  such that  $\mathbf{L}_E^1 \sqsubseteq \mathbf{L}_E^2$ . Then, for any two labellings  $\mathbf{L}_1, \mathbf{L}_2$  such that  $\mathbf{L}_1 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$  and  $\mathbf{L}_2 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$ , it holds that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ .

**Proof.** Taking into account Lemma 6,  $\mathbf{L}_1 \in F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$ , thus by Proposition 7 there must be a labelling  $\mathbf{L}'_2 \in F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$  such that  $\mathbf{L}_1 \sqsubseteq \mathbf{L}'_2$ . By Lemma 6 it also holds that  $\mathbf{L}_2 \in F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$ , and since  $|F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})| = 1$  it must be the case that  $\mathbf{L}'_2 = \mathbf{L}_2$ . Thus  $\mathbf{L}_1 \sqsubseteq \mathbf{L}'_2 = \mathbf{L}_2$ .  $\square$

**Lemma 8.** Let  $\mathcal{M}$  be an argumentation multipole  $(AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  which is **CO**-univocal under a set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ . Then  $\mathcal{M}$  is internally **PR**-homogeneous under  $\mathcal{L}'$ .

**Proof.** Let  $\mathbf{L}_E^1, \mathbf{L}_E^2$  be two labellings of  $\mathcal{L}'$  such that  $\mathbf{L}_E^1 \sqsubseteq \mathbf{L}_E^2$ , and let  $\mathbf{L}_1, \mathbf{L}_2$  be two labellings such that  $\mathbf{L}_1 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^1 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$ ,  $\mathbf{L}_2 \in F_{\text{PR}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E^2 \downarrow_{\mathcal{M}^{\text{inp}}, R_{\text{INP}}})$ . By Lemma 7,  $\mathbf{L}_1 \sqsubseteq \mathbf{L}_2$ . Then the condition required in Definition 35 is trivially verified with  $\mathbf{L}'_2 = \mathbf{L}_2$ .  $\square$

**Theorem 15.** Any contextually **CO**-legitimate replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **CO**-univocal under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$ , is **ID**-safe.

**Proof.** As in previous proofs, the claim in the case where  $E = \emptyset$  is a direct consequence of Proposition 12, thus in the following we assume  $E \neq \emptyset$ .

Let  $AF_2 = T(AF_1, \mathcal{M}_1, \mathcal{M}_2)$ , and let  $\mathbf{L}_1^*$  and  $\mathbf{L}_2^*$  be the ideal labellings of  $AF_1$  and  $AF_2$ , respectively. We prove that  $\mathbf{L}_1^* \downarrow_E \sqsubseteq \mathbf{L}_2^* \downarrow_E$  (the other direction of the proof can be obtained by a symmetric reasoning). Since by definition  $\mathbf{L}_1^*$  is a complete labelling of  $AF_1$  and **CO** is strongly transparent, there is a complete labelling  $\mathbf{L}_2$  of  $AF_2$  such that  $\mathbf{L}_2 \downarrow_E = \mathbf{L}_1^* \downarrow_E$ . We prove that for any preferred labelling  $\mathbf{L}_p^2$  of  $AF_2$  it holds that  $\mathbf{L}_2 \sqsubseteq \mathbf{L}_p^2$ : since  $\mathbf{L}_2^*$  is by definition the maximal (w.r.t.  $\sqsubseteq$ ) complete labelling satisfying this condition, it must be the case that  $\mathbf{L}_2 \sqsubseteq \mathbf{L}_2^*$ , thus in particular  $\mathbf{L}_2 \downarrow_E = \mathbf{L}_1^* \downarrow_E \sqsubseteq \mathbf{L}_2^* \downarrow_E$ .

First we prove that the hypotheses of Corollary 1 are satisfied for  $\mathcal{R}$ , thus  $\mathcal{R}$  is **PR**-safe. Since  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **CO**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$ , by Lemma 6 they are also **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$ , and since  $F_{\text{PR}}$  always returns a subset of  $F_{\text{CO}}$  it holds that  $\mathcal{L}_{\mathcal{R}}^{\text{PR}} \subseteq \mathcal{L}_{\mathcal{R}}^{\text{CO}}$ , therefore  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **PR**-equivalent under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ . Moreover, by Lemma 8  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are internally **PR**-homogeneous under  $\mathcal{L}_{\mathcal{R}}^{\text{PR}}$ .

Let us turn to the proof that  $\mathbf{L}_2 \sqsubseteq \mathbf{L}_p^2$ . Let  $\text{Args}_2$  be the set of arguments of  $AF_2$ , and  $D_2 = \text{Args}_2 \setminus E$ . First, it must be the case that  $\mathbf{L}_2 \downarrow_E \sqsubseteq \mathbf{L}_p^2 \downarrow_E$ : since  $\mathcal{R}$  is **PR**-safe, there is a preferred labelling  $\mathbf{L}_p^1$  of  $AF_1$  such that  $\mathbf{L}_p^2 \downarrow_E = \mathbf{L}_p^1 \downarrow_E$ , and by definition of ideal labelling  $\mathbf{L}_1^* \sqsubseteq \mathbf{L}_p^1$ , thus in particular  $\mathbf{L}_1^* \downarrow_E \sqsubseteq \mathbf{L}_p^1 \downarrow_E$ , and the conclusion follows by taking into account that  $\mathbf{L}_2 \downarrow_E = \mathbf{L}_1^* \downarrow_E$ . Second, by top-down decomposability of complete semantics (if  $D_2 \neq \emptyset$ ) or by Lemma 9 (if  $D_2 = \emptyset$ ),  $\mathbf{L}_2 \downarrow_{D_2} \sqsubseteq$

$F_{\text{CO}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_2 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$  and  $\mathbf{L}_P^2 \downarrow_{D_2} \in F_{\text{CO}}(AF_2 \downarrow_{D_2}, \mathcal{M}_2^{\text{inp}}, \mathbf{L}_P^2 \downarrow_{\mathcal{M}_2^{\text{inp}}}, R_{\text{INP}}^2)$ , thus by Lemma 7  $\mathbf{L}_2 \downarrow_E \subseteq \mathbf{L}_P^2 \downarrow_E$  also entails that  $\mathbf{L}_2 \downarrow_{D_2} \subseteq \mathbf{L}_P^2 \downarrow_{D_2}$ . Summing up, it holds that  $\mathbf{L}_2 \downarrow_E \subseteq \mathbf{L}_P^2 \downarrow_E$  and  $\mathbf{L}_2 \downarrow_{D_2} \subseteq \mathbf{L}_P^2 \downarrow_{D_2}$ , i.e.  $\mathbf{L}_2 \subseteq \mathbf{L}_P^2$ .  $\square$

**Proposition 14.** *An acyclic argumentation multipole  $\mathcal{M} = (AF, R_{\text{INP}}, R_{\text{OUTP}})$  w.r.t. a set  $E$  is **CO-univocal** under any set of labellings  $\mathcal{L}' \subseteq \mathcal{L}_E$ .*

**Proof.** Consider a labelling  $\mathbf{L}_E \in \mathcal{L}'$ . We reason by contradiction, assuming that there are two distinct labellings  $\mathbf{L}_1, \mathbf{L}_2 \in F_{\text{CO}}(AF, \mathcal{M}^{\text{inp}}, \mathbf{L}_E \downarrow_{\mathcal{M}^{\text{inp}}}, R_{\text{INP}})$ . Let  $\mathcal{L}_D$  be the set of arguments which are assigned different labels by  $\mathbf{L}_1$  and  $\mathbf{L}_2$ . Since  $\mathcal{M}$  is acyclic, there must be at least an argument  $A \in \mathcal{L}_D$  such that all arguments of  $AF$  that attack  $A$  (possibly none) are assigned the same labels from  $\mathbf{L}_1$  and  $\mathbf{L}_2$ . However, according to the conditions of Definition 24 the label of  $A$  is univocally determined by the labels of its attackers, both in the case  $A$  does not attack itself and in the case that  $A$  is self-attacking (in particular, if  $A$  is self-attacking then  $A$  is **out**-labelled if it has an **in**-labelled attacker, it is **undec** otherwise). This contradicts the fact that  $A \in \mathcal{L}_D$ .  $\square$

**Proposition 15.** *Any contextually **PR-legitimate** (**ID-legitimate**) replacement  $\mathcal{R} = (AF_1, \mathcal{M}_1, \mathcal{M}_2)$  with invariant set  $E$ , such that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are acyclic, is **PR-safe** (**ID-safe**).*

**Proof.** Considering preferred semantics, Proposition 14 and Lemma 8 entail that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are internally **PR-homogeneous** under any set  $\mathcal{L}' \subseteq \mathcal{L}_E$ , thus according to Corollary 1 the replacement  $\mathcal{R}$  is **PR-safe**. The result for ideal semantics is entailed by Theorem 15, taking into account that  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are **CO-univocal** under  $\mathcal{L}_{\mathcal{R}}^{\text{CO}}$  by Proposition 14, and they are **CO-equivalent** according to Lemma 6.  $\square$

## References

- [1] K. Atkinson, Introduction to special issue on modelling Popov v. Hayashi, *Artif. Intell. Law* 20 (1) (2012) 1–14.
- [2] P. Baroni, G. Boella, F. Cerutti, M. Giacomin, L.W.N. van der Torre, S. Villata, On input/output argumentation frameworks, in: *Proc. of the 4th Int. Conf. on Computational Models of Argument, COMMA 2012*, 2012, pp. 358–365.
- [3] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, *Knowl. Eng. Rev.* 26 (4) (2011) 365–410.
- [4] P. Baroni, F. Cerutti, M. Giacomin, G. Guida, Encompassing attacks to attacks in abstract argumentation frameworks, in: *Proc. of the 10th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2009*, 2009, pp. 83–94.
- [5] P. Baroni, F. Cerutti, M. Giacomin, G. Guida, AFRA: Argumentation framework with recursive attacks, *Int. J. Approx. Reason.* 52 (1) (2011) 19–37.
- [6] P. Baroni, P.E. Dunne, M. Giacomin, On the resolution-based family of abstract argumentation semantics and its grounded instance, *Artif. Intell.* 175 (3–4) (2011) 791–813.
- [7] P. Baroni, M. Giacomin, Solving semantic problems with odd-length cycles in argumentation, in: *Proc. of the 7th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2003*, 2003, pp. 440–451.
- [8] P. Baroni, M. Giacomin, On principle-based evaluation of extension-based argumentation semantics, *Artif. Intell.* 171 (10/15) (2007) 675–700 (Special issue on Argumentation in A.I.).
- [9] P. Baroni, M. Giacomin, Skepticism relations for comparing argumentation semantics, *Int. J. Approx. Reason.* 50 (6) (2009) 854–866.
- [10] P. Baroni, M. Giacomin, G. Guida, SCC-recursiveness: a general schema for argumentation semantics, *J. Artif. Intell.* 168 (1–2) (2005) 165–210.
- [11] P. Baroni, M. Giacomin, B. Liao, On topology-related properties of abstract argumentation semantics. A correction and extension to dynamics of argumentation systems: a division-based method, *Artif. Intell.* 212 (2014) 104–115.
- [12] R. Baumann, Normal and strong expansion equivalence for argumentation frameworks, *Artif. Intell.* 193 (2012) 18–44.
- [13] R. Baumann, What does it take to enforce an argument? Minimal change in abstract argumentation, in: *Proc. of the 20th European Conf. on Artificial Intelligence, ECAI 2012*, 2012, pp. 127–132.
- [14] R. Baumann, G. Brewka, Analyzing the equivalence zoo in abstract argumentation, in: *Proc. of the 14th Int. Workshop on Computational Logic in Multi-Agent Systems, CLIMA XIV*, 2013, pp. 18–33.
- [15] R. Baumann, G. Brewka, W. Dvorák, S. Woltran, Parameterized splitting: a simple modification-based approach, in: E. Erdem, J. Lee, Y. Lierler, D. Pearce (Eds.), *Correct Reasoning – Essays on Logic-Based AI in Honour of Vladimir Lifschitz*, in: *Lect. Notes Comput. Sci.*, vol. 7265, Springer, 2012, pp. 57–71.
- [16] R. Baumann, G. Brewka, R. Wong, Splitting argumentation frameworks: an empirical evaluation, in: *Theory and Applications of Formal Argumentation – First Int. Workshop, TFAA 2011, Revised Selected Papers*, in: *Lect. Notes Comput. Sci.*, vol. 7132, Springer, 2011, pp. 17–31.
- [17] R. Baumann, Splitting an argumentation framework, in: *Proc. of LPNMR 2011 11th Int. Conf. on Logic Programming and Nonmonotonic Reasoning*, 2011, pp. 40–53.
- [18] G. Boella, D.M. Gabbay, L. van der Torre, S. Villata, Meta-argumentation modelling i: methodology and techniques, *Stud. Log.* 93 (2–3) (2009) 297–355.
- [19] R. Booth, S. Kaci, T. Rienstra, L.W.N. van der Torre, Conditional acceptance functions, in: *Proc. of the 4th Int. Conf. on Computational Models of Argument, COMMA 2012*, 2012, pp. 470–477.
- [20] G. Brewka, S. Woltran, Abstract dialectical frameworks, in: *Proc. of the 12th Int. Conf. on Principles of Knowledge Representation and Reasoning, KR 2010*, 2010, pp. 102–111.
- [21] M.W.A. Caminada, On the issue of reinstatement in argumentation, in: *Proc. of the 10th European Conference on Logics in Artificial Intelligence, JELIA 2006*, 2006, pp. 111–123.
- [22] M.W.A. Caminada, Gabriella Pigozzi, On judgment aggregation in abstract argumentation, *Auton. Agents Multi-Agent Syst.* 22 (1) (2011) 64–102.
- [23] F. Cerutti, M. Giacomin, M. Vallati, M. Zanella, A SCC recursive meta-algorithm for computing preferred labellings in abstract argumentation, in: *Proc. of the 14th Int. Conf. on Principles of Knowledge Representation and Reasoning, KR 2014*, 2014, pp. 42–51.
- [24] S. Coste-Marquis, C. Devred, P. Marquis, Prudent semantics for argumentation frameworks, in: *Proc. of the 17th IEEE International Conf. on Tools with Artificial Intelligence, ICTAI 2005*, IEEE Computer Society, Hong Kong, China, 2005, pp. 568–572.
- [25] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and  $n$ -person games, *Artif. Intell.* 77 (1995) 321–357.
- [26] P.E. Dunne, A. Hunter, P. McBurney, S. Parsons, M. Wooldridge, Weighted argument systems: basic definitions, algorithms, and complexity results, *Artif. Intell.* 175 (2) (2011) 457–486.

- [27] W. Dvorák, S. Woltran, On the intertranslatability of argumentation semantics, *J. Artif. Intell. Res.* 41 (2011) 445–475.
- [28] D.M. Gabbay, Fibring argumentation frames, *Stud. Log.* 93 (2–3) (2009) 231–295.
- [29] D.M. Gabbay, Semantics for higher level attacks in extended argumentation frames part 1: overview, *Stud. Log.* 93 (2–3) (2009) 357–381.
- [30] S.A. Gaggl, W. Dvorák, Stage semantics and the SCC-recursive schema for argumentation semantics, *J. Log. Comput.* (2014), <http://dx.doi.org/10.1093/logcom/exu006>, in press.
- [31] T. Janhunen, E. Oikarinen, H. Tompits, S. Woltran, Modularity aspects of disjunctive stable models, *J. Artif. Intell. Res.* 35 (2009) 813–857.
- [32] B. Liao, H. Huang, Partial semantics of argumentation: basic properties and empirical results, *J. Log. Comput.* 23 (3) (2013) 541–562.
- [33] B. Liao, L. Jin, R.C. Koons, Dynamics of argumentation systems: a division-based method, *Artif. Intell.* 175 (2011) 1790–1814.
- [34] V. Lifschitz, D. Pearce, A. Valverde, Strongly equivalent logic programs, *ACM Trans. Comput. Log.* 2 (4) (2001) 526–541.
- [35] S. Modgil, Reasoning about preferences in argumentation frameworks, *Artif. Intell.* 173 (9–10) (2009) 901–934.
- [36] S. Modgil, T.J.M. Bench-Capon, Integrating object and meta-level value based argumentation, in: *Proc. of the 2nd Int. Conf. on Computational Models of Argument, COMMA 2008*, 2008, pp. 240–251.
- [37] E. Oikarinen, T. Janhunen, Modular equivalence for normal logic programs, in: *Proc. of the 17th European Conf. on Artificial Intelligence, ECAI 2006*, 2006, pp. 412–416.
- [38] E. Oikarinen, S. Woltran, Characterizing strong equivalence for argumentation frameworks, *Artif. Intell.* 175 (2011) 1985–2009.
- [39] H. Prakken, Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungian semantics, *Artif. Intell. Law* 20 (1) (2012) 57–82.
- [40] T. Rienstra, A. Perotti, S. Villata, D. Gabbay, L. van Der Torre, G. Boella, Multi-sorted argumentation frameworks, in: *Theory and Applications of Formal Argumentation – First Int. Workshop (TAFa 2011)*, Revised Selected Papers, in: *Lect. Notes Comput. Sci.*, vol. 7132, Springer, 2011, pp. 231–245.
- [41] B. Verheij, Two approaches to dialectical argumentation: admissible sets and argumentation stages, in: *Proc. of the Eighth Dutch Conf. on Artificial Intelligence, NAIC'96*, 1996, pp. 357–368.
- [42] B. Verheij, Two approaches to dialectical argumentation: admissible sets and argumentation stages, in: *Proc. of the Eighth Dutch Conf. on Artificial Intelligence, NAIC'96*, Utrecht, NL, 1996, pp. 357–368.
- [43] S. Villata, G. Boella, L. van Der Torre, Argumentation patterns, in: *Proc. of ARGMAS 2011 8th Int. Workshop on Argumentation in Multi-Agent Systems*, 2011, pp. 133–150.
- [44] A.Z. Wyner, T.J.M. Bench-Capon, Argument schemes for legal case-based reasoning, in: *Proc. of the 20th Annual Conf. on Legal Knowledge and Information Systems, JURIX 2007*, 2007, pp. 139–149.