Chapter 2
Semantics of Abstract Argument Systems

Pietro Baroni and Massimiliano Giacomin

1 Abstract Argument Systems

An *abstract argument system* or *argumentation framework*, as introduced in a seminal paper by Dung [13], is simply a pair $\langle \mathcal{A}, \mathcal{R} \rangle$ consisting of a set $\mathcal{A}$ whose elements are called *arguments* and of a binary relation $\mathcal{R}$ on $\mathcal{A}$ called *attack relation*. The set $\mathcal{A}$ may be finite or infinite in general, however, given the introductory purpose of this chapter, we will restrict the presentation to the case of finite sets of arguments. An argumentation framework has an obvious representation as a directed graph where nodes are arguments and edges are drawn from attacking to attacked arguments. A simple example of argumentation framework $AF_{2.1} = \langle \{a,b\}, \{(b,a)\} \rangle$ is shown in Figure 2.1.

While the word *argument* may recall several intuitive meanings, like the ones of "line of reasoning leading from some premise to a conclusion" or of "utterance in a dispute", abstract argument systems are not (even implicitly or indirectly) bound to any of them: an abstract argument is not assumed to have any specific structure but, roughly speaking, an argument is anything that may attack or be attacked by another argument. Accordingly, the argumentation framework depicted in Figure 2.1 is suitable to represent many different situations. For instance, in a context of reasoning about weather, argument *a* may be associated with the inferential step

$$\mathrm{b} \longrightarrow \mathrm{a}$$

Fig. 2.1 $AF_{2.1}$: a simple argumentation framework

Pietro Baroni
Dip. di Elettronica per l'Automazione, University of Brescia, Via Branze 38, 25123 Brescia, Italy
e-mail: baroni@ing.unibs.it

Massimiliano Giacomin
Dip. di Elettronica per l'Automazione, University of Brescia, Via Branze 38, 25123 Brescia, Italy
e-mail: giacomin@ing.unibs.it

I. Rahwan, G. R. Simari (eds.), *Argumentation in Artificial Intelligence,*
DOI 10.1007/978-0-387-98197-0_2,

"Tomorrow will rain because the national weather forecast says so", while b with "Tomorrow will not rain because the regional weather forecast says so". In a legal dispute a might be associated with a prosecutor's statement "The suspect is guilty because an eyewitness, Mr. Smith, says so" while b with the defense reply "Mr. Smith is notoriously alcohol-addicted and it is proved that he was completely drunk the night of the crime, therefore his testimony should not be taken into account". In a marriage arrangement setting (corresponding to the game-theoretic formulation of *stable marriage problem* [13]), a may be associated with "The marriage between Alice and John", while b with "The marriage between Barbara and John". Similarly, the attack relation has no specific meaning: again at a rough level, if an argument b attacks another argument a, this means that if b holds then a can not hold or, putting it in other words, that the evaluation of the status of b affects the evaluation of the status of a. In the weather example, two intrinsically conflicting conclusions are confronted and the attack relation corresponds to the fact that one conclusion is preferred to the other, e.g. because the regional forecast is considered more reliable than the national one (this kind of attack is often called *rebut* in the literature). In the legal dispute, the fact that Mr. Smith was drunk is not incompatible *per se* with the fact that the suspect is guilty, however it affects the reason why s/he should be believed to be guilty (this kind of attack has sometimes been called *undercut* in the literature, but the use of this term has not been uniform). In the stable marriage problem, the attack from b to a may simply be due to the fact that John prefers Barbara to Alice.

Abstracting away from the structure and meaning of arguments and attacks enables the study of properties which are independent of any specific aspect, and, as such, are relevant to any context that can be captured by the very terse formalization of abstract argument systems. As a counterpart of this generality, abstract argument systems feature a limited expressiveness and can hardly be adopted as a modeling formalism directly usable in application contexts. In fact the gap between a practical problem and its representation as an abstract argument system is patently too wide and requires to be filled by a less abstract formalism, dealing in particular with the construction of arguments and the conditions for an argument to attack another one. For instance, in a given formalism, attack may be identified at a purely syntactic level (e.g. the conclusion of an argument being the negation of the conclusion of another one). In another formalism, attack may not be a purely syntactic notion (e.g. the conclusions of two arguments being incompatible because they give rise to contradiction through some strict logical deduction). Only after the attack relation is defined in these "concrete" terms, it is appropriate to derive its abstract representation and exploit it for the analysis of general properties.

2 Abstract Argumentation Semantics

Given that arguments may attack each other, it is clearly the case that they can not stand all together and their status is subject to evaluation. In particular what we are usually interested in is the *justification state* of arguments: while this notion will be given a precise formal meaning[1] in the following, intuitively an argument is regarded as justified if it has some way to survive the attacks it receives, as not justified (or rejected) otherwise. In the following we will use the term *argument evaluation* to refer to a process aimed at determining the justification state of the arguments in an abstract argument system, i.e. on the basis of the attack relation only. In the case of Figure 2.1 the outcome of the evaluation process is almost undebatable: b should be justified (in particular it does not receive attacks) while a should not be justified (it has no way to survive the attack from b). The case of mutual attack of Figure 2.2 is less obvious: evident symmetry reasons suggest that the evaluation process should assign to b and a the same state, which should correspond neither to "full" justification nor to "full" rejection. An intuitive counterpart to Figure 2.2 is given by the case of two contradicting weather forecasts about rain tomorrow, coming from sources considered equally reliable.

It is important to remark that the evaluation process in abstract argument systems concerns the justification state of arguments, not of their conclusions (conclusions "do not exist" in the abstract setting). To clarify this distinction consider a new argument put forward by the prosecutor in the example of legal dispute: "The suspect is guilty because his fingerprints have been found on the crime scene". The new argument (let say c) has no attack relationships with the previous ones (in particular, the fact that the fingerprints have been found on the crime scene does not conflict with the fact that the witness was drunk that night) and, as a consequence, the corresponding abstract representation is the one given in Figure 2.3. Then a is still rejected, b is still justified and c (not receiving any attack) is justified too. Note that, at the underlying level, a and c have the same conclusion ("the suspect is guilty") and that this conclusion should now be intuitively regarded as justified. However, summarizing the justification state of conclusions is outside the scope of the abstract representation.

An *argumentation semantics* is the formal definition of a method (either declarative or procedural) ruling the argument evaluation process. Two main styles of argumentation semantics definition can be identified in the literature: *extension-based* and *labelling-based*.

In the extension-based approach a semantics definition specifies how to derive from an argumentation framework a set of *extensions*, where an extension E of an argumentation framework $\langle \mathcal{A}, \mathcal{R} \rangle$ is simply a subset of $\mathcal{A}$, intuitively representing a set of arguments which can "survive together" or are "collectively acceptable".

$b \rightleftarrows a$

Fig. 2.2 $AF_{2.2}$: a mutual attack

$c \quad b \longrightarrow a$

Fig. 2.3 $AF_{2.3}$: presence of an isolated argument

[1] Actually the term "justification" is used informally and with different meanings in the literature.

Putting things in more formal terms, given an extension-based semantics $\mathcal{S}$ and an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ we will denote the set of extensions prescribed by $\mathcal{S}$ for AF as $\mathcal{E}_{\mathcal{S}}(AF) \subseteq 2^{\mathcal{A}}$. The justification state of an argument $a \in \mathcal{A}$ according to an extension-based semantics $\mathcal{S}$ is then a derived concept, defined in terms of the membership of a to the elements of $\mathcal{E}_{\mathcal{S}}(AF)$.

In the labelling-based approach a semantics definition specifies how to derive from an argumentation framework a set of *labellings*, where a labelling L is the assignment to each argument in $\mathcal{A}$ of a label taken from a predefined set $\mathbb{L}$, which corresponds to the possible alternative states of an argument *in the context of a single labelling*. Putting things in formal terms, given a labelling-based semantics $\mathcal{S}$ with set of labels $\mathbb{L}$, a labelling of an argumentation framework $\langle \mathcal{A}, \mathcal{R} \rangle$ is a mapping $L : \mathcal{A} \rightarrow \mathbb{L}$. We denote the set of all possible labellings, i.e. of all possible mappings from $\mathcal{A}$ to $\mathbb{L}$, as $\mathfrak{L}(\mathcal{A}, \mathbb{L})$, and the set of labellings prescribed by $\mathcal{S}$ for AF as $\mathcal{L}_{\mathcal{S}}(AF) \subseteq \mathfrak{L}(\mathcal{A}, \mathbb{L})$. Again, the justification state of an argument a according to a labelling-based semantics $\mathcal{S}$ turns out to be a derived concept, defined in terms of the labels assigned to a in the various elements of $\mathcal{L}_{\mathcal{S}}(AF)$.

Let us exemplify the semantics definition styles with reference to the simple case of Figure 2.2. Let $\mathcal{S}^1_{ext}$ be a hypothetical extension-based semantics whose underlying principle consists in identifying as extensions the largest sets of non-conflicting arguments which reply to the attacks they receive. Intuitively, both the sets $\{a\}$ and $\{b\}$ satisfy this principle, yielding $\mathcal{E}_{\mathcal{S}^1_{ext}}(AF_{2.2}) = \{\{a\}, \{b\}\}$. Alternatively, let $\mathcal{S}^2_{ext}$ be a hypothetical extension-based semantics whose underlying principle consists in identifying as extensions the sets of arguments which are unattacked. Clearly, no argument features this property in $AF_{2.2}$, yielding as unique extension the empty set: $\mathcal{E}_{\mathcal{S}^2_{ext}}(AF_{2.2}) = \{\emptyset\}$. Let us turn to the labelling-based style and adopt the set of labels $\mathbb{L} = \{in, out, undecided\}$. Let $\mathcal{S}^1_{lab}$ be a hypothetical labelling-based semantics whose underlying principle consists in labelling *in* the elements of the largest sets of non-conflicting arguments which reply to the attacks they receive and labelling *out* the arguments attacked by arguments labelled *in*. Then two labellings are prescribed: $\mathcal{L}_{\mathcal{S}^1_{lab}}(AF_{2.2}) = \{\{(a, in), (b, out)\}, \{(a, out), (b, in)\}\}$. On the other hand, let $\mathcal{S}^2_{lab}$ be a hypothetical labelling-based semantics whose underlying principle consists in labelling *in* only unattacked arguments, labelling *out* the arguments attacked by arguments labelled *in*, and labelling *undecided* all the remaining ones. Clearly this would yield $\mathcal{L}_{\mathcal{S}^2_{lab}}(AF_{2.2}) = \{\{(a, undecided), (b, undecided)\}\}$.

As it is also evident from the above examples, for a given argumentation framework one or more extensions (labellings) may be prescribed by a given semantics. It has to be remarked that also the case where no extensions (labellings) are prescribed for an argumentation framework AF is in general possible, namely $\mathcal{E}_{\mathcal{S}}(AF) = \emptyset$ ($\mathcal{L}_{\mathcal{S}}(AF) = \emptyset$). This corresponds to the case where the semantics $\mathcal{S}$ is undefined in AF since no extensions (labellings) compliant with the definition of $\mathcal{S}$ exist. For extension-based semantics, note in particular that the case $\mathcal{E}_{\mathcal{S}}(AF) = \emptyset$ is very different from $\mathcal{E}_{\mathcal{S}}(AF) = \{\emptyset\}$. In the following we will denote as $\mathcal{D}_{\mathcal{S}}$ the set of argumentation frameworks where a semantics $\mathcal{S}$ is defined, namely, $\mathcal{D}_{\mathcal{S}} = \{AF \mid \mathcal{E}_{\mathcal{S}}(AF) \neq \emptyset\}$ or $\mathcal{D}_{\mathcal{S}} = \{AF \mid \mathcal{L}_{\mathcal{S}}(AF) \neq \emptyset\}$. A semantics $\mathcal{S}$ is called *universally defined* if any argumentation framework belongs to $\mathcal{D}_{\mathcal{S}}$. Further, an important terminological convention concerning the cardinality of the set of extensions (labellings) is worth introducing. If a semantics $\mathcal{S}$ always prescribes exactly one extension (labelling) for any argumentation framework belonging to $\mathcal{D}_{\mathcal{S}}$ then $\mathcal{S}$ is said to belong to the *unique-status* (or *single-status*) approach, otherwise it is said to belong to the *multiple-status* approach.

While the adoption of the labelling or extension-based style is a matter of subjective preference by the proponent(s) of a given semantics, a natural question concerns the expressiveness of the two styles. It is immediate to observe that any extension-based definition can be equivalently expressed in a simple labelling-based formulation, where a set of two labels is adopted (let say $\mathbb{L} = \{in, out\}$) corresponding to extension membership. On the other hand, an arbitrary labelling can not in general be formulated in terms of extensions. It is however a matter of fact that labellings considered in the literature typically include a label called *in* which naturally corresponds to extension membership, while other labels correspond to a partition of other arguments, easily derivable from extension membership and the attack relation. As a consequence, equivalent extension-based formulations of labelling-based semantics are typically available. Given this fact and the definitely prevailing custom of adopting the extension-based style in the literature, this chapter will focus on extension-based semantics.

3 Principles for extension-based semantics

While alternative semantics proposals differ from each other by the specific notion of extension they endorse, one may wonder whether there are reasonable general properties which are shared by all (or, at least, most) existing semantics and can be regarded as evaluation principles for new semantics. We discuss these principles in the following, distinguishing between properties of individual extensions and properties of the whole set of extensions. The reader may refer to [2] for a more extensive analysis of this issue.

The first basic requirement for any extension E corresponds to the idea that E is a set of arguments which "can stand together". Consequently, if an argument a attacks another argument b, then a and b can not be included together in an extension. This corresponds to the following *conflict-free* principle, which, as to our knowledge, is satisfied by all existing semantics in the literature.

Definition 2.1. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $S \subseteq \mathcal{A}$ is *conflict-free*, denoted as $cf(S)$, if and only if $\nexists a, b \in S$ such that $a\mathcal{R}b$. A semantics $\mathcal{S}$ satisfies the *conflict-free principle* if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}$, $\forall E \in \mathcal{E}_{\mathcal{S}}(AF)$ E is conflict-free.

A further requirement corresponds to the idea that an extension is a set of arguments which "can stand on its own", namely is able to withstand the attacks it receives from other arguments by "replying" with other attacks. Formally, this has a counterpart in the property of *admissibility*, which lies at the heart of all semantics discussed in [13] and is shared by many more recent proposals. The definition is based on the notions of acceptable argument and admissible set. In words, a is *acceptable* wrt. (or, equivalently, *defended* by) a set S if S “counterattacks” all attackers of a and a set S is *admissible* if it is conflict-free and defends all its elements.

Definition 2.2. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, an argument $a \in \mathcal{A}$ is *acceptable* wrt. a set $S \subseteq \mathcal{A}$ if and only if $\forall b \in \mathcal{A}\ b\mathcal{R}a \Rightarrow S\mathcal{R}b$.[2]

The function $\mathcal{F}_{AF} : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ which, given a set $S \subseteq \mathcal{A}$, returns the set of the acceptable arguments wrt. S, is called the *characteristic function* of AF.

Definition 2.3. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $S \subseteq \mathcal{A}$ is *admissible* if and only if $cf(S)$ and $\forall a \in S\ a$ is acceptable wrt. S. The set of all the admissible sets of AF is denoted as $\mathcal{AS}(AF)$.

Definition 2.4. A semantics $\mathcal{S}$ satisfies the *admissibility principle* if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}\ \mathcal{E}_{\mathcal{S}}(AF) \subseteq \mathcal{AS}(AF)$, namely $\forall E \in \mathcal{E}_{\mathcal{S}}(AF)$ it holds that:

$$a \in E \Rightarrow (\forall b \in \mathcal{A}, b\mathcal{R}a \Rightarrow E\mathcal{R}b). \tag{2.1}$$

The property of reinstatement is somewhat dual with respect to the notion of defense. Intuitively, if the attackers of an argument a are in turn attacked by an extension E one may assume that they have no effect on a: then a should be, in a sense, *reinstated*, therefore it should belong to E. This leads to the following *reinstatement principle* which turns out to be the converse of the implication (2.1).

Definition 2.5. A semantics $\mathcal{S}$ satisfies the *reinstatement principle* if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall E \in \mathcal{E}_{\mathcal{S}}(AF)$ it holds that:

$$(\forall b \in \mathcal{A}, b\mathcal{R}a \Rightarrow E\mathcal{R}b) \Rightarrow a \in E. \tag{2.2}$$

To exemplify these properties consider the argumentation framework $AF_{2.4}$ represented in Figure 2.4. It is easy to see that $\mathcal{AS}(AF_{2.4}) = \{\emptyset, \{a\}, \{b\}, \{a, c\}\}$, in particular c needs the “help” of a in order to be defended from the attack of b. The set $\{b\}$ obviously satisfies the reinstatement condition (2.2), and so does $\{a, c\}$, while the set $\{a\}$ does not, since it attacks all attackers of c, but does not include it.

Let us turn now to principles concerning sets of extensions. A first fundamental principle corresponds to the fact that the set of extensions only depends on the attack relation between arguments while it is totally independent of any property of arguments at the underlying language level. Formally, this principle corresponds to the fact that argumentation frameworks which are isomorphic have the “same” (modulo the isomorphism) extensions, as stated by the following definitions.

$$c \longleftarrow b \rightleftarrows a$$

Fig. 2.4 $AF_{2.4}$: not just a mutual attack

[2] With a little abuse of notation we define $S\mathcal{R}b \equiv \exists a \in S : a\mathcal{R}b$. Similarly, $b\mathcal{R}S \equiv \exists a \in S : b\mathcal{R}a$.

Definition 2.6. Two argumentation frameworks $AF_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$ and $AF_2 = \langle \mathcal{A}_2, \mathcal{R}_2 \rangle$ are isomorphic if and only if there is a bijective mapping $m : \mathcal{A}_1 \to \mathcal{A}_2$, such that $(a,b) \in \mathcal{R}_1$ if and only if $(m(a), m(b)) \in \mathcal{R}_2$. This is denoted as $AF_1 \doteq_m AF_2$.

Definition 2.7. A semantics $\mathcal{S}$ satisfies the *language independence principle* if and only if $\forall AF_1 \in \mathcal{D}_{\mathcal{S}}, \forall AF_2 \in \mathcal{D}_{\mathcal{S}}$ such that $AF_1 \doteq_m AF_2, \mathcal{E}_{\mathcal{S}}(AF_2) = \{M(E) \mid E \in \mathcal{E}_{\mathcal{S}}(AF_1)\}$, where $M(E) = \{b \mid \exists a \in E, b = m(a)\}$.

All argumentation semantics we are aware of adhere to this principle.

Another principle concerns possible inclusion relationships between extensions. Considering for instance the admissible sets of the example in Figure 2.4, the empty set is of course included in all other ones, moreover the set $\{a\}$ is included in $\{a,c\}$. In such a situation, the question arises whether an extension may be a strict subset of another one. The answer is typically negative, for reasons which are related in particular with the notion of skeptical justification, to be discussed later. Accordingly, the I-maximality principle can be introduced.

Definition 2.8. A set of extensions $\mathcal{E}$ is I-maximal if and only if $\forall E_1, E_2 \in \mathcal{E}$, if $E_1 \subseteq E_2$ then $E_1 = E_2$. A semantics $\mathcal{S}$ satisfies the *I-maximality principle* if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}\ \mathcal{E}_{\mathcal{S}}(AF)$ is I-maximal.

A further principle is related with the notion of attack which is intrinsically directional: if a attacks b this corresponds to the fact that a has the power to affect b, while not vice versa (unless, in turn, b attacks a). Generalizing this consideration, one may require the evaluation of an argument a to be only affected by its attackers, the attackers of its attackers and so on, i.e. by its ancestors in the attack relationship. In other words, an argument a may affect another argument b only if there is a directed path from a to b. For instance, in Figure 2.4 the evaluations of a and b affect each other and both affect c, while the evaluation of c should not affect those of a and b but rather depend on them. It is reasonable to require this notion of *directionality* to be reflected by the set of extensions prescribed by a semantics. This can be formalized by referring to sets of arguments not receiving attacks from outside.

Definition 2.9. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a non-empty set $S \subseteq \mathcal{A}$ is *unattacked* if and only if $\nexists a \in (\mathcal{A} \setminus S) : a\mathcal{R}S$. The set of unattacked sets of AF is denoted as $\mathcal{US}(AF)$.

We also need to introduce the notion of restriction of an argumentation framework to a subset S of its arguments.

Definition 2.10. Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation framework. The *restriction* of AF to $S \subseteq \mathcal{A}$ is the argumentation framework $AF{\downarrow_S} = \langle S, \mathcal{R} \cap (S \times S) \rangle$.

The directionality principle can then be defined by requiring an unattacked set to be unaffected by the remaining part of the argumentation framework as far as extensions are concerned.

Definition 2.11. A semantics $\mathcal{S}$ satisfies the *directionality principle* if and only if $\forall AF \in \mathcal{D}_{\mathcal{S}}, \forall S \in \mathcal{US}(AF), \mathcal{AE}_{\mathcal{S}}(AF,S) = \mathcal{E}_{\mathcal{S}}(AF{\downarrow_S})$, where $\mathcal{AE}_{\mathcal{S}}(AF,S) \triangleq \{(E \cap S) \mid E \in \mathcal{E}_{\mathcal{S}}(AF)\} \subseteq 2^S$.

In words, the intersection of any extension prescribed by $\mathcal{S}$ for AF with an unattacked set S is equal to one of the extensions prescribed by $\mathcal{S}$ for the restriction of AF to S, and vice versa. Referring to the example of Figure 2.4, $\mathcal{US}(AF_{2.4}) = \{\{a,b\},\{a,b,c\}\}$. Then, the restriction of $AF_{2.4}$ to its unattacked set $\{a,b\}$ coincides with $AF_{2.2}$ shown in Figure 2.2. A hypothetical semantics $\mathcal{S}_1$ such that $\mathcal{E}_{\mathcal{S}_1}(AF_{2.4}) = \{\{a,c\},\{b\}\}$ and $\mathcal{E}_{\mathcal{S}_1}(AF_{2.2}) = \{\{a\},\{b\}\}$ would satisfy the directionality principle in this case. On the other hand, a hypothetical semantics $\mathcal{S}_2$ such that $\mathcal{E}_{\mathcal{S}_2}(AF_{2.4}) = \{\{a,c\}\}$ and $\mathcal{E}_{\mathcal{S}_2}(AF_{2.2}) = \{\{a\},\{b\}\}$ would not.

4 The notion of justification state

At a first level, the justification state of an argument a can be conceived in terms of its extension membership. A basic classification encompasses only two possible states for an argument, namely justified or not justified. In this respect, two alternative types of justification, namely *skeptical* and *credulous* can be considered.

Definition 2.12. Given a semantics $\mathcal{S}$ and an argumentation framework $AF \in \mathcal{D}_{\mathcal{S}}$, an argument a is:

- *skeptically justified* if and only if $\forall E \in \mathcal{E}_{\mathcal{S}}(AF)\ a \in E$;
- *credulously justified* if and only if $\exists E \in \mathcal{E}_{\mathcal{S}}(AF)\ a \in E$.

Clearly the two notions coincide for unique-status approaches, while, in general, credulous justification includes skeptical justification. To refine this relatively rough classification, a consolidated tradition considers three justification states.

Definition 2.13. [18] Given a semantics $\mathcal{S}$ and an argumentation framework $AF \in \mathcal{D}_{\mathcal{S}}$, an argument a is:

- *justified* if and only if $\forall E \in \mathcal{E}_{\mathcal{S}}(AF), a \in E$ (this is clearly the "strongest" possible level of justification, corresponding to skeptical justification);
- *defensible* if and only if $\exists E_1, E_2 \in \mathcal{E}_{\mathcal{S}}(AF) : a \in E_1, a \notin E_2$ (this is a weaker level of justification, corresponding to arguments which are credulously but not skeptically justified);
- *overruled* if and only if $\forall E \in \mathcal{E}_{\mathcal{S}}(AF), a \notin E$ (in this case a can not be justified in any way and should be rejected).

Some remarks are worth about the above classification, which is largely (and sometimes implicitly) adopted in the literature: on one hand, while (two) different levels of justified arguments are encompassed, no distinctions are drawn among rejected arguments; on the other hand, the attack relation plays no role in the derivation of justification state from extensions. These points are addressed by a different classification of justification states introduced by Pollock [16] in the context of a unique-status approach. Again, three cases are possible for an argument a:

- a belongs to the (unique) extension E: then it is justified, or, using Pollock's terminology, *undefeated*;

- a does not belong to the (unique) extension E and is attacked by (some member of) E: then, using Pollock's terminology, it is *defeated outright*, corresponding to a strong form of rejection;
- a does not belong to the (unique) extension E but does not receive attacks from E: then it is *provisionally defeated*, corresponding to a weaker form of rejection.

Both these classifications of justification states are unsatisfactory in some respect. It is however possible to combine the intuitions underlying both of them, obtaining a systematic classification of seven possible justification states [6]. As a starting point, considering the relationship between an argument a and a specific extension E, three main situations[3], as in Pollock's classification, can be envisaged:

- a is *in* E, denoted as $in(a,E)$, if $a \in E$;
- a is *definitely out from* E, denoted as $do(a,E)$, if $a \notin E \wedge E\mathcal{R}a$;
- a is *provisionally out from* E, denoted as $po(a,E)$, if $a \notin E \wedge \neg E\mathcal{R}a$.

Then, taking into account the possible existence of multiple extensions, an argument can be in any of the above three states with respect to all, some or none of the extensions. This gives rise to 27 hypothetical combinations. It is however easy to see that some of them are impossible: for instance, if an argument is in a given state with respect to all extensions this clearly excludes that it is in another state with respect to any extension. Directly applying this kind of considerations, seven possible justification states emerge.

Definition 2.14. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ and a *non-empty* set of extensions $\mathcal{E}$ the possible justification states of an argument $a \in \mathcal{A}$ according to $\mathcal{E}$ are defined by the following mutually exclusive conditions:

- $\forall E \in \mathcal{E}\ in(a,E)$, denoted as JS_I;
- $\forall E \in \mathcal{E}\ do(a,E)$, denoted as JS_D;
- $\forall E \in \mathcal{E}\ po(a,E)$, denoted as JS_P;
- $\exists E \in \mathcal{E} : do(a,E)$, $\exists E \in \mathcal{E} : po(a,E)$, and $\nexists E \in \mathcal{E} : in(a,E)$, denoted as JS_{DP};
- $\exists E \in \mathcal{E} : in(a,E)$, $\exists E \in \mathcal{E} : po(a,E)$, and $\nexists E \in \mathcal{E} : do(a,E)$, denoted as JS_{IP};
- $\exists E \in \mathcal{E} : in(a,E)$, $\exists E \in \mathcal{E} : do(a,E)$, and $\nexists E \in \mathcal{E} : po(a,E)$, denoted as JS_{ID};
- $\exists E \in \mathcal{E} : in(a,E)$, $\exists E \in \mathcal{E} : do(a,E)$, and $\exists E \in \mathcal{E} : po(a,E)$, denoted as JS_{IDP}.

Correspondences with "traditional" definitions of justification states are easily drawn. An argument is skeptically justified if and only if it is in the JS_I state, while credulous justification corresponds to the disjunction of the states JS_I, JS_{IP}, JS_{ID}, and JS_{IDP}. As to Definition 2.13, the state of justified corresponds to JS_I, the state of overruled to the disjunction of JS_D, JS_P, and JS_{DP}, while the state of defensible to the disjunction of JS_{IP}, JS_{ID}, and JS_{IDP}. Turning to Pollock's classification, it is easy to see that in the case of a unique-status semantics only JS_I, JS_D and JS_P may hold, which correspond to the state of undefeated, defeated outright and provisionally defeated, respectively. Other meaningful ways of defining aggregated justification states are investigated in [6].

[3] The case $a \in E \wedge E\mathcal{R}a$ is prevented by the conflict-free principle.

5 A review of extension-based argumentation semantics

Turning from general notions to actual approaches, we now examine several argumentation semantics proposed in the literature. From a historical point of view, it is possible to distinguish between:

- four "traditional" semantics, considered in Dung's original paper [13], namely *complete*, *grounded*, *stable*, and *preferred* semantics;
- subsequent proposals introduced by various authors in the literature, often to overcome some limitation or improve some undesired behavior of a traditional approach: we consider *stage*, *semi-stable*, *ideal*, $CF2$, and *prudent* semantics.

It is important to note that while the definitions of the semantics we will describe are formulated in the context of purely abstract argumentation frameworks, the underlying intuitions have commonalities with other (somewhat more concrete) formalizations in related contexts. In fact, as already mentioned, one of the main results of Dung's paper is showing that abstract argumentation frameworks are able to capture the properties of a large variety of more specific formalisms. This means that it is possible to define mappings from entities defined in a more specific formalism into an argumentation framework. In [13] mappings of this kind are provided for *stable marriage* problems, *default theories* in Reiter's default logic, logic programs with negation as failure, and Pollock's theory of defeasible reasoning.

Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation framework obtained through a mapping from a more specific formalism (e.g. a logic program). We have now two ways of deriving a "meaningful" set of subsets of $\mathcal{A}$: on one hand, we can apply a purely abstract semantics $\mathcal{S}$ to AF, obtaining the relevant set of extensions $\mathcal{E}_{\mathcal{S}}(AF)$. On the other hand, we can start from a meaningful concept in the underlying formalism (e.g. the set of models of a logic program) and then map it into a set $\mathcal{M}$ of subsets of $\mathcal{A}$ (continuing the example, by deriving the set of arguments corresponding to each model). An important question then arises: can interesting relationships be identified between $\mathcal{E}_{\mathcal{S}}(AF)$ and $\mathcal{M}$, given a suitable choice of the semantics $\mathcal{S}$? A strikingly affirmative answer is provided in [13]: it is shown that properly selecting $\mathcal{S}$ among the traditional grounded, preferred, and stable semantics one obtains sets of extensions $\mathcal{E}_{\mathcal{S}}(AF)$ which coincide with the sets of arguments corresponding to meaningful notions in the formalisms mentioned above. Specific indications about these coincidences will be given in the review of individual semantics. At a general level, they confirm that abstract argumentation semantics is a powerful analysis tool, able to focus on essential properties of a variety of formalisms and to shed light on their (possibly hidden) significant common features.

5.1 Complete semantics

We start our review by the notion of complete extension, as it lies at the heart of all traditional Dung's semantics. Actually the notion of complete extension is not associated with a notion of *complete semantics* in [13], but the term complete semantics has subsequently gained acceptance in the literature and will be used in the present analysis to refer to the properties of the set of complete extensions. Complete semantics is denoted as $\mathcal{CO}$.

The notion of complete extension is based on the principles of admissibility and reinstatement: a complete extension is a set which is able to defend itself and includes all arguments it defends, as stated by the following definition.

Definition 2.15. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $E \subseteq \mathcal{A}$ is a *complete extension* if and only if E is admissible and every argument of $\mathcal{A}$ which is acceptable wrt. E belongs to E, i.e. $E \in \mathcal{AS}(AF) \wedge \mathcal{F}_{AF}(E) \subseteq E$.

It is worth noting here that the empty set is admissible and that arguments not receiving attacks in an argumentation framework AF (called *initial arguments* and denoted as $\mathcal{IN}(AF)$ in the sequel) are acceptable wrt. the empty set (in fact $\mathcal{IN}(AF) = \mathcal{F}_{AF}(\emptyset)$). It can be shown that the following properties hold:

- $\forall AF\ \mathcal{E}_{\mathcal{CO}}(AF) \neq \emptyset$ (namely $\mathcal{CO}$ is universally defined);
- $\emptyset \in \mathcal{E}_{\mathcal{CO}}(AF)$ if and only if $\mathcal{IN}(AF) = \emptyset$;
- $\forall E \in \mathcal{E}_{\mathcal{CO}}(AF)\ \mathcal{IN}(AF) \subseteq E$.

Due to reinstatement, any complete extension not only includes the initial arguments, but also the arguments they defend, those which are in turn defended by them, and so on. More formally, for any argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, given a set $S \subseteq \mathcal{A}$, let $\mathcal{F}^1_{AF}(S) \triangleq \mathcal{F}_{AF}(S)$ and for $i > 1$, $\mathcal{F}^i_{AF}(S) \triangleq \mathcal{F}_{AF}(\mathcal{F}^{i-1}_{AF}(S))$. Then, it turns out that $\forall AF$, $\forall E \in \mathcal{E}_{\mathcal{CO}}(AF)$, $\forall i \geq 1$, $\mathcal{F}^i_{AF}(\emptyset) \subseteq E$.

In the example of Figure 2.1, $\{b\} = \mathcal{IN}(AF_{2.1})$ is clearly the only complete extension. Similarly in the example of Figure 2.3 the only complete extension is $\{b, c\}$. In the example of Figure 2.2 there are three complete extensions: $\emptyset$ (as there are no initial arguments), $\{a\}$, and $\{b\}$ (each one defending itself against the other). By similar considerations and the property of reinstatement it is easy to see that in the example of Figure 2.4 $\mathcal{E}_{\mathcal{CO}}(AF_{2.4}) = \{\emptyset, \{a, c\}, \{b\}\}$. The example of Figure 2.5 requires some more articulated considerations. First note that $\mathcal{IN}(AF_{2.5}) = \emptyset$ hence $\emptyset \in \mathcal{E}_{\mathcal{CO}}(AF_{2.5})$. Then note that all singletons except $\{c\}$ are admissible: to check whether they are complete extensions or not we have to resort to the reinstatement property. In particular, a defends c from b but not from d, so $\{a\}$ stands as a complete extension. On the other hand b defends d from its only attacker c, therefore $\{b, d\}$ is a complete extension, while $\{b\}$ is not. Argument d does not defend any other argument apart itself, thus $\{d\}$ is a complete extension. We have now to consider possibly larger admissible sets. It is easy to see that $\{a, c\}$ and $\{a, d\}$ are admissible (and attack all arguments they do not include). Summarizing we have $\mathcal{E}_{\mathcal{CO}}(AF_{2.5}) = \{\emptyset, \{a\}, \{d\}, \{b, d\}, \{a, c\}, \{a, d\}\}$.

We complete the treatment of $\mathcal{CO}$ by considering its ability to satisfy the properties which are not common to all semantics: we note that admissibility and reinstatement are enforced by definition, while the satisfaction of directionality is proved in

$$d \rightleftarrows c \longleftarrow b \rightleftarrows a$$

Fig. 2.5 $AF_{2.5}$: two mutual attacks.

[2]. As the examples above abundantly show, $\mathcal{CO}$ does not satisfy I-maximality, since a complete extension E_1 may be a proper subset of a complete extension E_2.

5.2 Grounded semantics

Grounded semantics, denoted as $\mathcal{GR}$, is a traditional unique-status approach whose formulation in argumentation-based reasoning (see e.g. the one proposed by Pollock in [16]) predates the following quite technical definition given in [13].

Definition 2.16. The grounded extension of an argumentation framework AF, denoted as $\mathrm{GE}(AF)$, is the least fixed point of its characteristic function $\mathcal{F}_{AF}$.

The underlying and preexisting informal intuition is however rather simple and can be directly put in relationship with some notions we already discussed in the context of complete semantics. The basic idea is that the (unique) grounded extension can be built incrementally starting from the initial unattacked arguments. Then the arguments attacked by them can be suppressed, resulting in a modified argumentation framework where, possibly, the set of initial arguments is larger. In turn the arguments attacked by the "new" initial arguments can be suppressed, and so on. The process stops when no new initial arguments arise after a deletion step: the set of all initial arguments identified so far is the grounded extension. An example with a graphical illustration of this incremental process is given in Figure 2.6, resulting in $\mathcal{E}_{\mathcal{GR}}(AF) = \{\{a,c\}\}$. To put it in other words, the grounded extension includes those and only those arguments whose defense is "rooted" in initial arguments (see [2] for a formal treatment of this notion, called *strong defense*). If there are no initial arguments the grounded extension is the empty set.

The reader should have noticed that the above construction corresponds to the iterated application of $\mathcal{F}_{AF}(\emptyset)$ already met in previous section: the set of arguments obtained up to each step i above coincides with $\mathcal{F}^i_{AF}(\emptyset)$ and the process stops when $\mathcal{F}^i_{AF}(\emptyset) = \mathcal{F}^{i+1}_{AF}(\emptyset)$ (i.e. a fixed point of $\mathcal{F}_{AF}$ is reached). As a counterpart to this intuitive correspondence, it is proved in [13] that:

i) the grounded extension is the least (wrt. set inclusion) complete extension: for any AF $\mathrm{GE}(AF) \in \mathcal{E}_{\mathcal{CO}}(AF)$ and $\forall E \in \mathcal{E}_{\mathcal{CO}}(AF)$ $\mathrm{GE}(AF) \subseteq E$;
ii) for any finite (and, more generally, finitary [13]) AF $\mathrm{GE}(AF) = \cup_{i=1,\dots,\infty}\mathcal{F}^i_{AF}(\emptyset)$.

Given the above explanations it should be now immediate to see that $\mathrm{GE}(AF_{2.1}) = \{b\}$, $\mathrm{GE}(AF_{2.3}) = \{b,c\}$, $\mathrm{GE}(AF_{2.2}) = \mathrm{GE}(AF_{2.4}) = \mathrm{GE}(AF_{2.5}) = \emptyset$.

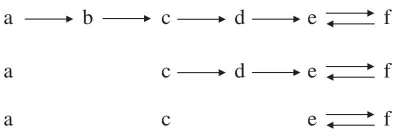

Fig. 2.6 $AF_{2.6}$: an example to illustrate grounded semantics

Besides the relationship with Pollock's approach [16], the grounded extension has been put in correspondence in [13] with the well-founded semantics of logic programs [20]. Turning to general semantics properties it is immediate to see that $\mathcal{GR}$ is universally defined and satisfies admissibility, reinstatement and I-maximality. It is proved in [2] that $\mathcal{GR}$ also satisfies directionality.

5.3 Stable semantics

Stable semantics, denoted as $\mathcal{ST}$, relies on a very simple (and easy to formalize) intuition: an extension should be able to attack all arguments not included in it. This leads to the notion of stable extension [13].

Definition 2.17. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $E \subseteq \mathcal{A}$ is a *stable extension* of AF if and only if E is conflict-free and $\forall a \in \mathcal{A}$, $a \notin E \Rightarrow E\mathcal{R}a$.

By definition, any stable extension E is also a complete extension (thus in particular $\mathrm{GE}(AF) \subseteq E$) and a maximal conflict-free set of AF. Referring to the examples seen so far, identifying stable extensions is straightforward: $\mathcal{E}_{\mathcal{ST}}(AF_{2.1}) = \{b\}$, $\mathcal{E}_{\mathcal{ST}}(AF_{2.2}) = \{\{a\},\{b\}\}$, $\mathcal{E}_{\mathcal{ST}}(AF_{2.3}) = \{b,c\}$, $\mathcal{E}_{\mathcal{ST}}(AF_{2.4}) = \{\{a,c\},\{b\}\}$, $\mathcal{E}_{\mathcal{ST}}(AF_{2.5}) = \{\{a,c\},\{a,d\},\{b,d\}\}$, $\mathcal{E}_{\mathcal{ST}}(AF_{2.6}) = \{\{a,c,e\},\{a,c,f\}\}$.

The simple intuition underlying stable semantics has significant counterparts in several contexts: it is proved in [13] that stable extensions can be put in correspondence with solutions of cooperative n-person games, solutions of the stable marriage problem, extensions of Reiter's default logic [19], and stable models of logic programs [15]. Stable semantics however has also a significant drawback: it is not universally defined as there are argumentation frameworks where no stable extensions exist. A simple example is provided in Figure 2.7: no conflict-free set is able to attack all other arguments in this case. While it has sometimes been claimed by supporters of stable semantics that situations where stable extensions do not exist are "pathological" in some sense, it has been shown in [13] that perfectly reasonable problems may be formalized with argumentation frameworks such that $\mathcal{E}_{\mathcal{ST}}(AF) = \emptyset$.

As to general properties, it is easy to see that I-maximality is enforced by definition and, since stable extensions are a subset of complete extensions, also admissibility and reinstatement are satisfied. $\mathcal{ST}$ is not directional, due to the fact that it is not universally defined. To see this, consider the example in Figure 2.8: we have that $\mathcal{E}_{\mathcal{ST}}(AF_{2.8}) = \{\{b\}\}$, however, considering the unattacked set $S = \{a,b\}$, we have $\mathcal{E}_{\mathcal{ST}}(AF_{2.8}\downarrow_S) = \{\{a\},\{b\}\}$. Therefore $\mathcal{AE}_{\mathcal{ST}}(AF_{2.8}, S) = \{\{b\}\} \neq \mathcal{E}_{\mathcal{ST}}(AF_{2.8}\downarrow_S)$.

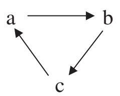

Fig. 2.7 $AF_{2.7}$: a three-length cycle

5.4 *Preferred semantics*

The “aggressive” requirement that an extension must attack anything outside it may be relaxed by requiring that an extension is as large as possible and able to defend itself from attacks. This is captured by the notion of preferred extension [13].

Definition 2.18. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $E \subseteq \mathcal{A}$ is a *preferred extension* of AF if and only if E is a maximal (wrt. set inclusion) element of $\mathcal{AS}(AF)$.

According to this definition every preferred extension E is also a complete extension (entailing that $\mathrm{GE}(AF) \subseteq E$); indeed, preferred extensions may be equivalently defined as maximal complete extensions. It follows that preferred semantics, denoted as $\mathcal{PR}$, is universally defined (as complete semantics is) and that, taking into account the treatment of the examples in Section 5.1, $\mathcal{E}_{\mathcal{PR}}(AF_{2.1}) = \{\{b\}\}$, $\mathcal{E}_{\mathcal{PR}}(AF_{2.2}) = \{\{a\},\{b\}\}$, $\mathcal{E}_{\mathcal{PR}}(AF_{2.3}) = \{\{b,c\}\}$, $\mathcal{E}_{\mathcal{PR}}(AF_{2.4}) = \{\{a,c\},\{b\}\}$, $\mathcal{E}_{\mathcal{PR}}(AF_{2.5}) = \{\{a,c\},\{a,d\},\{b,d\}\}$, and $\mathcal{E}_{\mathcal{PR}}(AF_{2.6}) = \{\{a,c,e\},\{a,c,f\}\}$. The reader may notice that in all these examples the set of preferred extensions coincides with the set of stable extensions. In general any stable extension is also a preferred extension but not vice versa. In the example of Figure 2.7 we have $\mathcal{E}_{\mathcal{PR}}(AF_{2.7}) = \{\emptyset\}$ while $\mathcal{E}_{\mathcal{ST}}(AF_{2.7}) = \emptyset$. In the example of Figure 2.8 there are two preferred extensions, namely $\{a\}$, $\{b\}$, but only one of them, namely $\{b\}$, is also stable.

The intuition underlying preferred semantics has a correspondence with preferential semantics of logic programs [12]. As to general semantics properties it is immediate to see that $\mathcal{PR}$ satisfies I-maximality, admissibility, reinstatement, while it is proved in [2] that it is directional. Given these facts, $\mathcal{PR}$ has often been regarded as the most satisfactory semantics in the context of Dung’s framework.

5.5 *Stage and semi-stable semantics*

Stage [21] and semi-stable [8] semantics, denoted respectively as $\mathcal{STA}$ and $\mathcal{SST}$, are based on the idea of prescribing the maximization not only of the arguments included in an extension but also of those attacked by it.

Definition 2.19. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ and a set $E \subseteq \mathcal{A}$ the *range* of E is defined as $E \cup E^+$, where $E^+ \triangleq \{a \in \mathcal{A} : E\mathcal{R}a\}$. E is a *stage extension* if and only if E is a conflict-free set with maximal (wrt. set inclusion) range. E is a *semi-stable extension* if and only if E is a complete extension with maximal (wrt. set inclusion) range.

As evident from Definition 2.19, the two semantics differ in the requirement on the sets whose range is maximal: being conflict-free for stage semantics, complete

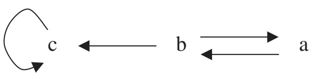

Fig. 2.8 $AF_{2.8}$: $\mathcal{ST}$ is not directional

extensions (or equivalently, admissible sets) for semi-stable semantics. Both $\mathcal{STA}$ and $\mathcal{SST}$ are clearly universally defined (differently from stable semantics), while coinciding with stable semantics (differently from preferred semantics) when stable extensions exist. In fact, for any stable extension E it holds that $E \cup E^+ = \mathcal{A}$. It follows that any complete extension (conflict-free set) which is not stable does not satisfy the range maximization requirement for argumentation frameworks where stable extensions exist, hence the coincidence of $\mathcal{ST}$ and $\mathcal{SST}$ ($\mathcal{STA}$) in these cases. If stable extensions do not exist, $\mathcal{SST}$ selects anyway as extensions some complete extensions and the maximization requirement restricts the choice to preferred extensions: $\forall AF\ \mathcal{E}_{\mathcal{SST}}(AF) \subseteq \mathcal{E}_{\mathcal{PR}}(AF)$. On the other hand, $\mathcal{STA}$ does not necessarily select admissible sets as extensions. Figure 2.9 shows an argumentation framework where $\mathcal{SST}$ and $\mathcal{STA}$ agree and do not coincide neither with $\mathcal{ST}$ nor with $\mathcal{PR}$: in fact $\mathcal{E}_{\mathcal{ST}}(AF_{2.9}) = \emptyset$, $\mathcal{E}_{\mathcal{PR}}(AF_{2.9}) = \{\{a\},\{b\}\}$, $\mathcal{E}_{\mathcal{SST}}(AF_{2.9}) = \mathcal{E}_{\mathcal{STA}}(AF_{2.9}) = \{\{b\}\}$. On the other hand, in the case of Figure 2.7 $\mathcal{E}_{\mathcal{PR}}(AF_{2.7}) = \mathcal{E}_{\mathcal{SST}}(AF_{2.7}) = \{\emptyset\}$ while $\mathcal{E}_{\mathcal{STA}}(AF_{2.7}) = \{\{a\},\{b\},\{c\}\}$.

It is easy to see that both $\mathcal{SST}$ and $\mathcal{STA}$ satisfy I-maximality, while only $\mathcal{SST}$ satisfies admissibility and reinstatement. $\mathcal{SST}$ and $\mathcal{STA}$ do not satisfy directionality, as $\mathcal{ST}$ does not.

5.6 *Ideal semantics*

Ideal semantics [14], denoted as $\mathcal{ID}$, provides a unique-status approach allowing the acceptance of a set of arguments possibly larger than in the case of $\mathcal{GR}$.

Definition 2.20. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ a set $S \subseteq \mathcal{A}$ is ideal if and only if S is admissible and $\forall E \in \mathcal{E}_{\mathcal{PR}}(AF)\ S \subseteq E$. The ideal extension, denoted as $\mathrm{ID}(AF)$, is the maximal (wrt. set inclusion) ideal set.

The definition of ideal set prescribes admissibility and skeptical justification under preferred semantics. From Section 5.2 we know that the grounded extension satisfies both requirements and is therefore an ideal set (this in particular implies that ideal semantics is universally defined). Ideal sets strictly larger than the grounded extension may exist, as shown by the example in Figure 2.10 where it holds that $\mathcal{E}_{\mathcal{PR}}(AF_{2.10}) = \{\{a,d\},\{b,d\}\}$: in this case the grounded extension is empty while $\mathrm{ID}(AF_{2.10}) = \{d\}$.

By definition $\mathcal{ID}$ satisfies I-maximality and admissibility. It is proved in [2] that $\mathcal{ID}$ also satisfies reinstatement and directionality.

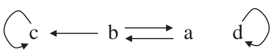

Fig. 2.9 $AF_{2.9}$: $\mathcal{SST}$ may not coincide with $\mathcal{PR}$ or $\mathcal{ST}$

5.7 $CF2$ semantics

$CF2$ semantics is defined in the frame of the SCC-recursive scheme [7] which is a general pattern for the definition of argumentation semantics based on the graph-theoretical notion of *strongly connected components* (SCCs) of an argumentation framework, i.e. the equivalence classes of arguments under the relation of mutual reachability via attack links. $CF2$ semantics can be roughly regarded as selecting its extensions among the maximal conflict-free sets of AF, on the basis of some topological requirements related to the decomposition of AF into strongly connected components. Examining in detail the definition of $CF2$ semantics is beyond the scope of this chapter: the interested reader may refer to [7].

Definition 2.21. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set $E \subseteq \mathcal{A}$ is an extension of $CF2$ semantics, i.e. $E \in \mathcal{E}_{CF2}(AF)$, if and only if

- $E \in \mathcal{MCF}(AF)$ if $|\mathrm{SCCS}_{AF}| = 1$
- $\forall S \in \mathrm{SCCS}_{AF}$ $(E \cap S) \in \mathcal{E}_{CF2}(AF{\downarrow}_{UP_{AF}(S,E)})$ otherwise

where $\mathcal{MCF}(AF)$ denotes the set of maximal conflict-free sets of AF, SCCS_{AF} denotes the set of strongly connected components of AF, and, for any $E, S \subseteq \mathcal{A}$, $UP_{AF}(S,E) = \{a \in S \mid \nexists b \in E : b \notin S, b\mathcal{R}a\}$.

The underlying idea consists in relying only on I-maximality and the conflict-free principle within a unique strongly connected component S, where all arguments (if more than one) both receive and deliver at least an attack. In particular it turns out that when AF consists of exactly one strongly connected component, the set of extensions prescribed by $CF2$ semantics exactly coincides with the set of maximal conflict-free sets of AF. This yields a uniform multiple-status treatment of cycles, which is not achieved by other semantics. In fact, it is easy to see that for any argumentation framework AF consisting of an even-length attack cycle it holds that $\mathcal{E}_{\mathcal{ST}}(AF) = \mathcal{E}_{\mathcal{PR}}(AF) = \mathcal{E}_{\mathcal{SST}}(AF) \neq \{\emptyset\}$. For instance we already know that $\mathcal{E}_{\mathcal{ST}}(AF_{2.2}) = \mathcal{E}_{\mathcal{PR}}(AF_{2.2}) = \mathcal{E}_{\mathcal{SST}}(AF_{2.2}) = \mathcal{MCF}(AF_{2.2}) = \{\{a\},\{b\}\}$. Similarly, in the case of a four-length cycle two extensions arise, each consisting of two arguments. On the other hand, for any argumentation framework AF consisting of an

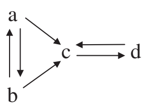

Fig. 2.10 $AF_{2.10}$: an example to illustrate ideal semantics

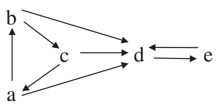

Fig. 2.11 $AF_{2.11}$: an example to illustrate $CF2$ semantics

odd-length attack cycle it holds that $\mathcal{E}_{\mathcal{ST}}(AF) = \emptyset$ and $\mathcal{E}_{\mathcal{PR}}(AF) = \mathcal{E}_{\mathcal{SST}}(AF) = \{\emptyset\}$. This disuniformity in the treatment of cycles gives rise to counterintuitive behaviors in some simple examples and has therefore been regarded as problematic in the literature [17]. Clearly $CF2$ semantics is able to overcome this limitation since $\mathcal{E}_{CF2}(AF) = \mathcal{MCF}(AF)$ for any argumentation framework AF consisting of an attack cycle, independently of its length. For instance, $\mathcal{E}_{CF2}(AF_{2.7}) = \{\{a\},\{b\},\{c\}\}$.

In a generic argumentation framework, $CF2$ extensions may be computed following the decomposition of AF into strongly connected components, which, due to a well-known graph-theoretical property, can be partially ordered according to the attack relation. Consider the example in Figure 2.11, which consists of two SCCs, namely $S_1 = \{a,b,c\}$ and $S_2 = \{d,e\}$. According to Definition 2.21 it must be the case that $E \cap S_1$ is a maximal conflict-free set of $AF_{2.11}\downarrow_{\{a,b,c\}}$. Thus we get three possible starting points for the construction of extensions, namely $\{a\}$, $\{b\}$, $\{c\}$. It has then to be noted that in any of these three cases d is attacked. Therefore in all cases $UP_{AF}(S_2,E)$ consists of the only argument e, yielding $E \cap S_2 = \{e\}$ as the only possibility. In summary $\mathcal{E}_{CF2}(AF_{2.11}) = \{\{a,e\},\{b,e\},\{c,e\}\}$. It can be seen that in all other examples considered so far $CF2$ extensions coincide with preferred extensions.

As to general properties, $CF2$ semantics is I-maximal, since its extensions are maximal conflict-free sets of AF, directional [2], and universally defined, since, for any AF, $\mathcal{MCF}(AF) \neq \emptyset$. On the other hand, the "desired" treatment of odd-length cycles entails that $CF2$ semantics gives up the traditional properties of admissibility and reinstatement as shown, for instance, by the example of Figure 2.7 where no extension is an admissible set and all violate the reinstatement property. It is proven however in [2] that the departure from the notion of reinstatement is not radical, since $CF2$ semantics satisfies two weaker versions of this property called $\mathcal{CF}$-reinstatement and weak reinstatement. As another confirmation that $CF2$ semantics has significant relationships with traditional semantics it can be recalled that any preferred extension is included in a $CF2$ extension, any stable extension is a $CF2$ extension and the grounded extension is included in any $CF2$ extension.

5.8 Prudent semantics

The family of prudent semantics [10] is introduced by considering a more extensive notion of attack in the context of traditional semantics. In particular, an argument a *indirectly attacks* an argument b if there is an odd-length attack path from a to b. For instance, in Figure 2.5 a indirectly attacks d, while in Figure 2.6 a indirectly attacks d and f, b indirectly attacks e, and c indirectly attacks f. The odd-length path need not be the shortest path and may include cycles: in Figure 2.10 both a and b indirectly attack d, while in Figure 2.11 a indirectly attacks e. A set S of arguments is free of indirect conflicts, denoted as $icf(S)$, if $\nexists a,b \in S$ such that a indirectly attacks b. The prudent version of the admissibility property and of several traditional notions of extension can then be defined.

Definition 2.22. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, a set of arguments $S \subseteq \mathcal{A}$ is *p(rudent)-admissible* if and only if $\forall a \in S$ a is acceptable wrt. S

and $icf(S)$. A set of arguments $E \subseteq \mathcal{A}$ is: a *preferred p-extension* if and only if E is a maximal (wrt. set inclusion) p-admissible set; a *stable p-extension* if and only if $icf(E)$ and $\forall a \in (\mathcal{A} \setminus E)\ E\mathcal{R}a$; a *complete p-extension* if and only if E is p-admissible and there is no $a \notin E$ such that a is acceptable wrt. E and $icf(E \cup \{a\})$.

Definition 2.23. Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, the function $\mathcal{F}^p_{AF} : 2^{\mathcal{A}} \to 2^{\mathcal{A}}$ such that for a given a set $S \subseteq \mathcal{A}$, $\mathcal{F}^p_{AF}(S) = \{a \mid a$ is acceptable with respect to $S \wedge icf(S \cup \{a\})\}$ is called the *p-characteristic function* of AF. Let j be the lowest integer such that the sequence $(\mathcal{F}^p_{AF})^i(\emptyset)$ is stationary for $i \geq j$: $(\mathcal{F}^p_{AF})^j(\emptyset)$ is the *grounded p-extension* of AF, denoted as GPE(AF).

The prudent versions of grounded, complete, preferred and stable semantics are denoted as $\mathcal{GRP}$, $\mathcal{COP}$, $\mathcal{PRP}$ and $\mathcal{STP}$, respectively.

It is possible to note that the adoption of indirect conflict has a different impact on the various notions of traditional semantics. The definition of stable prudent extension corresponds to the one of stable extension with an additional requirement. This means that, when $\mathcal{STP}$ is defined, its extensions are also stable (and thus have the same properties) but $\mathcal{D}_{\mathcal{STP}} \subsetneq \mathcal{D}_{\mathcal{ST}}$. For instance in the example of Figure 2.12 $\mathcal{STP}$ is not defined since the only stable extension, namely $\{a,d,e\}$ is not prudent due to the indirect attack from a to d. Analogously to the traditional version, the grounded prudent extension, denoted as GPE(AF), can be conceived as the result of the incremental application of a (p-)characteristic function starting from initial arguments. As $\mathcal{F}^p_{AF}$ is more restrictive than $\mathcal{F}_{AF}$ it follows that the grounded prudent extension is a possibly strict subset of the traditional grounded extension. This entails in particular that reinstatement is given up by $\mathcal{GRP}$, as it can be seen in the example of Figure 2.12 where GPE$(AF_{2.12}) = \{a,e\} \subsetneq$ GE$(AF_{2.12}) = \{a,d,e\}$ and d is not reinstated. In the context of complete and preferred prudent semantics a notable effect is that an initial argument may not be included in an extension. This is the case of a in the example of Figure 2.12 where $\mathcal{E}_{\mathcal{COP}}(AF_{2.12}) = \mathcal{E}_{\mathcal{PRP}}(AF_{2.12}) = \{\{a,e\},\{d,e\}\}$. Besides reinstatement this gives up also the directionality property, showing the neat departure of $\mathcal{COP}$ and $\mathcal{PRP}$ from the track of traditional semantics.

6 Advanced topics

This final section provides a quick overview and literature references for some advanced topics in the field of abstract argumentation, namely semantics agreement, skepticism relations, and semantics principles.

As to the first point, while different semantics proposals correspond to alternative intuitions which manifest themselves in distinct behaviors in some argumentation

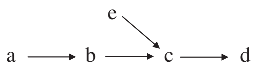

Fig. 2.12 $AF_{2.12}$: an example to illustrate prudent semantics

frameworks, there are also argumentation frameworks where the sets of extensions prescribed by different semantics coincide, as evident from several of the examples seen above. Topological properties of argumentation frameworks providing sufficient conditions for semantics agreement have been investigated in Dung's original paper: for instance the absence of attack cycles is sufficient to ensure agreement among grounded, stable and preferred semantics, while the absence of odd-length attack cycles is sufficient to ensure agreement between stable and preferred semantics. Subsequent results concerning topological families of argumentation frameworks relevant to agreement properties are reported in [11, 1], while a systematic set-theoretical analysis of agreement classes is provided in [5].

As to the notion of skepticism, it has often been used in informal ways to discuss semantics behavior, e.g. by observing that a semantics is "more skeptical" than another one. Intuitively, a semantics is more skeptical than another if it makes less committed choices about the justification of the arguments: a skeptical behavior tends to leave arguments in an "undecided" justification state, while a non-skeptical behavior corresponds to more "resolute" choices about acceptance or rejection of arguments. The issue of formalizing skepticism relations between sets of extension in terms of set theoretical properties and then comparing semantics according to skepticism has been first addressed in [6] and subsequently developed in [4].

Turning finally back to general properties of argumentation semantics, some of them, like admissibility and reinstatement, have regularly been considered in the literature [18], while the task of systematically defining a set of criteria for semantics evaluation and comparison has been undertaken only recently. Besides the principles we have explicitly discussed in Section 3, two families of *adequacy* properties (based on skepticism relations) are introduced in [2], where it is shown that, considering also these properties, none of the literature semantics discussed in this chapter is able to comply with all the desirable criteria. A novel semantics able to satisfy all of them has been proposed in [3]. At a different abstraction level, where argument structure and construction are explicitly dealt with, general *rationality postulates* for argumentation systems have been introduced in [9]. Exploring the definition of general principles for argumentation at different abstraction levels, investigating their relationships and analyzing their suitability for different application domains appear to be open and fruitful research directions. A related research issue concerns the identification of a generic definition scheme able to encompass into a unifying view a large variety of semantics. In this perspective, it is shown in [7] that all traditional semantics adhere to the parametric SCC-recursive scheme, where they differ simply by a *base function* which only specifies the sets of extensions of argumentation frameworks consisting of a single SCC.

The above mentioned results suggest that though there is a wide corpus of literature on abstract argumentation semantics, providing a rich variety of alternative approaches, the field is far from being "mature" and there is still large room for investigating both fundamental theoretical issues and their potential impact on practical applications.

References

1. Baroni, P., Giacomin, M.: Characterizing defeat graphs where argumentation semantics agree. In: G. Simari, P. Torroni (eds.) Proc. 1st Int. Workshop on Argumentation and Non-Monotonic Reasoning (ARGNMR07), pp. 33–48. Tempe, AZ (2007)
2. Baroni, P., Giacomin, M.: On principle-based evaluation of extension-based argumentation semantics. Artif. Intell. (Special issue on Argumentation in A.I.) **171**(10/15), 675–700 (2007)
3. Baroni, P., Giacomin, M.: Resolution-based argumentation semantics. In: P. Besnard, S. Doutre, A. Hunter (eds.) Proc. 2nd Int. Conf. on Computational Models of Argument (COMMA 2008), pp. 25–36. IOS Press, Toulouse, F (2008)
4. Baroni, P., Giacomin, M.: Skepticism relations for comparing argumentation semantics. Int. J. Approx. Reason. **50**(6), 854–866 (2009)
5. Baroni, P., Giacomin, M.: A systematic classification of argumentation frameworks where semantics agree. In: P. Besnard, S. Doutre, A. Hunter (eds.) Proc. 2nd Int. Conf. on Computational Models of Argument (COMMA 2008), pp. 37–48. IOS Press, Toulouse, F (2008)
6. Baroni, P., Giacomin, M., Guida, G.: Towards a formalization of skepticism in extension-based argumentation semantics. In: Proc. 4th Workshop on Computational Models of Natural Argument (CMNA 2004), pp. 47–52. Valencia, Spain (2004)
7. Baroni, P., Giacomin, M., Guida, G.: SCC-recursiveness: a general schema for argumentation semantics. Artif. Intell. **168**(1-2), 165–210 (2005)
8. Caminada, M.: Semi-stable semantics. In: P.E. Dunne, T. Bench-Capon (eds.) Proc. 1st Int. Conf. on Computational Models of Argument (COMMA 2006), pp. 121–130. IOS Press, Liverpool, UK (2006)
9. Caminada, M., Amgoud, L.: An axiomatic account of formal argumentation. In: Proc. 20th National Conf. on Artificial Intelligence (AAAI-05), pp. 608–613. AAAI Press, Menlo Park, CA (2005)
10. Coste-Marquis, S., Devred, C., Marquis, P.: Prudent semantics for argumentation frameworks. In: Proc. 17th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI 2005), pp. 568–572. IEEE Computer Society, Hong Kong, China (2005)
11. Coste-Marquis, S., Devred, C., Marquis, P.: Symmetric argumentation frameworks. In: Proc. 8th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005), pp. 317–328. Barcelona, E (2005)
12. Dung, P.M.: Negations as hypotheses: An abductive foundation for logic programming. In: K. Furukawa (ed.) Proc. 8th Int. Conf. on Logic Programming (ICLP 91), pp. 3–17. MIT Press, Paris, F (1991)
13. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. Artif. Intell. **77**(2), 321–357 (1995)
14. Dung, P.M., Mancarella, P., Toni, F.: A dialectic procedure for sceptical, assumption-based argumentation. In: P.E. Dunne, T. Bench-Capon (eds.) Proc. 1st Int. Conf. on Computational Models of Argument (COMMA 2006), pp. 145–156. IOS Press, Liverpool, UK (2006)
15. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. In: R.A. Kowalski, K. Bowen (eds.) Proc. 5th Int. Conf. on Logic Programming (ICLP 88), pp. 1070–1080. MIT Press, Cambridge, Massachusetts (1988)
16. Pollock, J.L.: How to reason defeasibly. Artif. Intell. **57**(1), 1–42 (1992)
17. Pollock, J.L.: Justification and defeat. Artif. Intell. **67**(2), 377–407 (1994)
18. Prakken, H., Vreeswijk, G.A.W.: Logics for defeasible argumentation. In: D.M. Gabbay, F. Guenthner (eds.) Handbook of Philosophical Logic, Second Edition. Kluwer Academic Publishers, Dordrecht (2001)
19. Reiter, R.: A logic for default reasoning. Artif. Intell. **13**(1–2), 81–132 (1980)
20. van Gelder, A., Ross, K., Schlipf, J.S.: The well-founded semantics for general logic programs. J. ACM **38**(3), 620–650 (1991)
21. Verheij, B.: Two approaches to dialectical argumentation: admissible sets and argumentation stages. In: Proc. 8th Dutch Conf. on Artificial Intelligence (NAIC'96), pp. 357–368. Utrecht, NL (1996)

http://www.springer.com/978-0-387-98196-3

Argumentation in Artificial Intelligence
Rahwan, I. (Ed.)
2009, XIII, 493 p., Hardcover
ISBN: 978-0-387-98196-3