Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

A Principle-based Analysis of Abstract Agent Argumentation Semantics

Liuwen Yu[1,2,3] , **Dongheng Chen**[4] , **Lisha Qiao**[1,2] , **Yiqi Shen**[5] , **Leendert van der Torre**[1,5]
[1]University of Luxembourg
[2]University of Bologna
[3]University of Turin
[4]Sun Yat-sen University
[5]Zhejiang University
liuwen.yu@uni.lu

Abstract

Abstract agent argumentation frameworks extend Dung's theory with agents, and in this paper we study four types of semantics for them. First, agent defense semantics replaces Dung's notion of defense by some kind of agent defense. Second, social agent semantics prefers arguments that belong to more agents. Third, agent reduction semantics considers the perspective of individual agents. Fourth, agent filtering semantics are inspired by a lack of knowledge. We study five existing principles and we introduce twelve new ones. In total, we provide a full analysis of fifty-two agent semantics and the seventeen principles.

1 Introduction

The two volumes of the Handbook of Formal Argumentation (Baroni et al. 2018; Gabbay et al. to appear) explain the central role of Dung's theory of abstract argumentation (Dung 1995) and many of its variants proposed over the past few decades. However, whereas several papers have proposed *agent-based* variants (Awad et al. 2017; Karanikolas, Bisquert, and Kaklamanis 2019; Gao et al. 2016), so far an overview of these variants is lacking. Moreover, the semantics of agent argumentation is related to merging argumentation frameworks (Coste-Marquis et al. 2007; Delobelle, Konieczny, and Vesic 2018; Caminada and Pigozzi 2011). We address the following research questions:

1. What kind of semantics can be defined for agent argumentation frameworks?
2. Which of the principles proposed in the literature (Baroni and Giacomin 2007; van der Torre and Vesic 2017) do not hold for such agent semantics?
3. What new principles can we define to distinguish the varieties of agent semantics?

For comparison, we distinguish four kinds of semantics for agent argumentation frameworks:

Agent defense approaches adapt Dung's notion of defense for argumentation semantics.

Social approaches (Leite and Martins 2011) are based on counting the number of agents and a reduction to preference-based argumentation (Amgoud and Cayrol 2002).

Agent reductions take the perspective of individual agents and create extensions accordingly (Giacomin 2017).

Filtering methods are inspired by the knowledge or trust of the agents (Arisaka, Satoh, and van der Torre 2017) and leave out some arguments or attacks, because they do not belong to any agents.

We make two important observations about the way the principle-based approach is used in formal argumentation in general, and in this paper in particular.

Minimality First, agent-based extensions typically introduce various aspects such as coalitions, knowledge, uncertainty, support, and so on. In line with common practice in the principle-based approach, this paper uses a minimal extension of Dung as a common core to these approaches. We only introduce an abstract set of agents, and we associate arguments with agents and nothing else.

Distinguishability Principles and axioms can be used in many ways. Often, they conceptualize the behavior of a system at a higher level of abstraction. Moreover, in the absence of a standard approach, principles can be used as a guideline for choosing the appropriate definitions and semantics depending on various needs. Therefore, in formal argumentation, principles are often more technical. The most discussed principles are admissibility, directionality and SCC decomposibility, which also play a central role in this paper. In this paper, we focus on principles distinguishing kinds of agent semantics.

The paper is organized as follows. In the following section we introduce agent argumentation frameworks. We discuss four kinds of semantics for agent argumentation frameworks in the next four sections. As for principles, traditional principles are introduced in the following section and thereafter variants of traditional principles are introduced. Then we introduce eight new agent principles. Finally, we discuss related work, future work and the conclusions of the paper. Due to space limitations, we only sketch a few proofs.

2 Agent Argumentation Framework

This section introduces agent argumentation frameworks. They generalize argumentation frameworks studied by Dung (1995), which are directed graphs, where the nodes are arguments, and the arrows correspond to the attack relation.

Definition 1 (Argumentation framework (Dung 1995))**.** *An* argumentation framework *(AF) is a pair $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A}$ is a set called arguments, and $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation over $\mathcal{A}$ called attack. For a set $S \subseteq \mathcal{A}$ and an argument $a \in \mathcal{A}$, we say that S attacks a if there exists $b \in S$ such that b attacks a, a attacks S if there exists $b \in S$ such that a attacks b, $a^- = \{b \in \mathcal{A} | b \text{ attacks } a\}$, and $S^-_{out} = \{a \in \mathcal{A} \setminus S | \ a \text{ attacks } S\}$.*

Dung's admissibility-based semantics is based on the concept of defense. A set of arguments defends another argument if they attack all its attackers.

Definition 2 (Admissible (Dung 1995))**.** *Let $\langle \mathcal{A}, \rightarrow \rangle$ be an AF. $E \subseteq \mathcal{A}$ is* conflict-free *iff there are no arguments a and b in E such that a attacks b. $E \subseteq \mathcal{A}$* defends *c iff for all arguments b attacking c, there is an argument a in E such that a attacks b. $E \subseteq \mathcal{A}$ is* admissible *iff it is conflict-free and defends all its elements.*

For their principle-based analysis, Baroni and Giacomin (2007) define semantics as a function from argumentation frameworks to sets of subsets of arguments.

Definition 3. (Dung semantics (Baroni and Giacomin 2007)) *Dung semantics is a function σ that associates with an argumentation framework $AF = \langle \mathcal{A}, \rightarrow \rangle$ a set of subsets of $\mathcal{A}$, and the elements of $\sigma(AF)$ are called extensions.*

Dung distinguishes between several definitions of extension.

Definition 4 (Extensions (Dung 1995))**.** *Let $\langle \mathcal{A}, \rightarrow \rangle$ be an AF. $E \subseteq \mathcal{A}$ is a* complete extension *iff it is admissible and it contains all the arguments it defends. $E \subseteq \mathcal{A}$ is a* grounded extension *iff it is the smallest complete extension (for set inclusion). $E \subseteq \mathcal{A}$ is a* preferred extension *iff it is the largest complete extension (for set inclusion). $E \subseteq \mathcal{A}$ is a* stable extension *iff it is conflict-free, and it attacks each argument which does not belong to E.*

Each kind of extension may be seen as an acceptability semantics that formally rules the argument evaluation process. In this article, we use $\sigma \in \{c, g, p, s\}$ to represent Dung semantics {complete, grounded, preferred, stable}.

Example 1 (Two conflicts)**.** *Consider the argumentation framework visualized on the left in Figure 1, where $\mathcal{A} = \{a, b, c, d\}$, $\rightarrow = \{a \rightarrow b, b \rightarrow a, c \rightarrow d, d \rightarrow c\}$. Each argument defends itself. There are nine admissible sets – $\{a\}, \{b\}, \{c\}, \{d\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \emptyset$ – which are all complete extensions. The grounded extension is $\emptyset$. The preferred extensions $\{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}$ are also stable extensions. For example, in an oft-used dinner scenario, we may choose between fish (a) or meat (b), and we may choose between eating at home (c) or going out (d), and these two choices are independent. In structured argumentation, these arguments may have a complex structure, providing the reasons for these conclusions, but in abstract argumentation we do not detail these reasons.*

An agent argumentation framework extends an argumentation framework with a set of agents and a relation associating arguments with agents. Note that an argument can belong to no agent, one agent, or multiple agents. This is the most general case. We briefly discuss restrictions in the further work section towards the end of this paper.

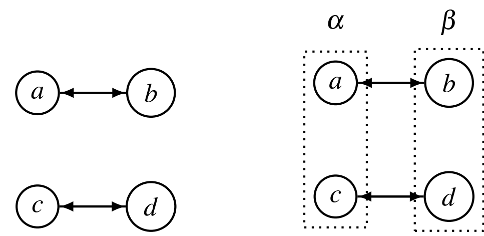

Figure 1: An AF and an AAF

We write $a \sqsubset \alpha$ for argument a belongs to agent α, or that agent α has argument a.

Definition 5 (Agent argumentation framework)**.** *An* agent argumentation framework *(AAF) is a 4-tuple $\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$ where $\mathcal{A}$ is a set of arguments, $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation over $\mathcal{A}$ called attack, $\mathcal{S}$ is a set of agents or sources, and $\sqsubset \subseteq \mathcal{A} \times \mathcal{S}$ is a binary relation associating arguments with agents. $\mathcal{A}_\alpha = \{a \in \mathcal{A} | a \sqsubset \alpha\}$ for all arguments that belong to agent α, $\mathcal{S}_a = \{\alpha | a \sqsubset \alpha\}$ for all agents that have argument a, $\rightarrow_a = \{x \rightarrow y | x = a \text{ or } y = a\}$ for the attack relations related to argument a, and $\sqsubset_\alpha = \{(a, \alpha) | a \sqsubset \alpha\}$ for the relation between agent α and its arguments.*

Example 2 (Two conflicts, continued from Example 1)**.** *Consider the agent argumentation framework visualized on the right in Figure 1. This figure should be read as follows. Each dashed box contains the arguments belonging to the same agent, $\mathcal{S} = \{\alpha, \beta\}$, and $\sqsubset = \{(a, \alpha), (b, \beta), (c, \alpha), (d, \beta)\}$. For example, Alice ($\alpha$) may hold the arguments for eating fish and staying at home, and Bob (β) may hold the arguments for eating meat and going outside.*

3 Agent Defense Semantics

We now introduce a new kind of defense for agent argumentation frameworks, which we call agent defense. Roughly, if an agent puts forward an argument, it can only be defended by arguments from the same agent. In extensions with coalitions, we may also consider agents of the same coalition defending each others' arguments (Qiao et al. 2021).

In individual agent defense, only a single agent can defend an argument, whereas in collective agent defense, a set of agents can do that.

Definition 6 (Agent Admissible)**.** *Let $\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$ be an AAF:*

- *$E \subseteq \mathcal{A}$ is* conflict-free *iff there are no arguments a and b in E such that a attacks b.*
- *$E \subseteq \mathcal{A}$ individually agent defends (agent defends$_1$) c iff there exists an agent α in $\mathcal{S}_c$ such that for all arguments b in $\mathcal{A}$ attacking c, there exists an argument a in $E \cap \mathcal{A}_\alpha$ such that a attacks b.*
- *$E \subseteq \mathcal{A}$ collectively agent defends (agent defends$_2$) c iff for all arguments b in $\mathcal{A}$ attacking c, there exists an agent α in $\mathcal{S}_c$ and an argument a in $E \cap \mathcal{A}_\alpha$ such that a attacks b.*

- $E \subseteq \mathcal{A}$ *is* agent admissible$_i$ *iff it is conflict-free and agent defends$_i$ all its elements, for i in* $\{1,2\}$.

The following example illustrates agent defense, and its role in so-called reinstatement. Though reinstatement is considered by many to be a desirable property, there is also a minority opinion that argues that reinstatement should not hold in general, c.f. the arguments and examples of Horty (2001). Example 3 shows that there is a middle way in this debate. Agent defense semantics allows for reinstatement if all the arguments belong to the same agent, but not if the arguments belong to distinct agents.

Example 3 (Reinstatement)**.** *Consider the agent argumentation framework visualized in Figure 2, where* $\mathcal{A} = \{a,b,c\}$, $\rightarrow = \{c \rightarrow b, b \rightarrow a\}$, $\mathcal{S} = \{\alpha,\beta,\gamma\}$ *and* $\sqsubset = \{(a,\alpha),(b,\beta),(c,\gamma)\}$. *Argument c defends argument a, but it does not agent defend it. For example, in the dinner scenario, Alice (α) may hold an argument in favor of eating meat, Bob (β) holds a better argument in favor of not eating meat but fish, and Cayrol (γ) holds an argument asking why fish is not an option (c). Assuming that Alice and Cayrol are not in a coalition, Cayrol does not agent defend the argument of Alice against the attacker of Bob.*

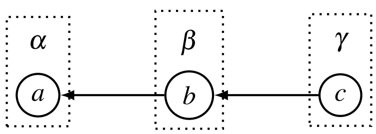

Figure 2: Agent Reinstatement

Definition 7 (Agent semantics)**.** *An agent semantics is a function δ that associates a set of subsets of $\mathcal{A}$ with an agent argumentation framework $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, and the elements of $\delta(AAF)$ are called agent extensions.*

We use Sem_1 and Sem_2 to represent agent semantics based on individual defense and collective defense respectively.

Definition 8 (Agent extensions)**.** *Let* $\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*:*

- $E \subseteq \mathcal{A}$ *is an* agent complete$_i$ extension *iff E is agent admissible$_i$ and it contains all the arguments it agent defends$_i$, for* $i \in \{1,2\}$.
- $E \subseteq \mathcal{A}$ *is an* agent grounded$_i$ extension *iff it is the smallest agent complete$_i$ extension (for set inclusion), for* $i \in \{1,2\}$.
- $E \subseteq \mathcal{A}$ *is an* agent preferred$_i$ extension *iff it is the largest agent complete$_i$ extension (for set inclusion), for* $i \in \{1,2\}$.
- $E \subseteq \mathcal{A}$ *is an* agent stable$_i$ extension *iff it is conflict-free and it attacks all the arguments in* $\mathcal{A} \backslash E$, *for* $i \in \{1,2\}$.

The following two examples illustrate agent extensions.

Example 4 (Two conflicts, continued from Example 2)**.** *Reconsider Figure 1. Each argument agent defends itself, therefore the agent complete extensions are the same as the complete extensions of the corresponding extensions of the argumentation framework without considering agents. The agent grounded, preferred and stable extensions are also the same as those of the argumentation framework.*

Example 5 (Reinstatement, continued from Example 3)**.** *Reconsider Figure 2. The individual and collective agent complete extension is* $\{c\}$. *It is also the unique individual and collective agent grounded and preferred extension. There is no agent stable extension. When the only accepted argument is c, it suggests a vegetarian dinner. Using stable semantics, no agreement is reached on dinner.*

The following example illustrates the difference between individual agent defense and collective agent defense. In particular, if a set of arguments individually agent defends another argument, then it also collectively agent defends it, but the example illustrates that the opposite does not always hold.

Example 6 (Collective defense)**.** *Consider the agent argumentation framework visualized in Figure 3, where* $\mathcal{A} = \{a,b_1,b_2,c_1,c_2\}$, $\rightarrow = \{c_1 \rightarrow b_1, b_1 \rightarrow a, c_2 \rightarrow b_2, b_2 \rightarrow a\}$, $\mathcal{S} = \{\alpha,\beta,\gamma\}$, $\sqsubset = \{(a,\alpha),(a,\beta),(b_1,\gamma),(b_2,\gamma),(c_1,\alpha),(c_2,\beta)\}$. *For example, in the dinner scenario, Alice and Bob argue in favor of eating meat, Cayrol has two better arguments for eating fish, but Alice argues why the first argument of Cayrol cannot be accepted, and Bob argues why the second argument of Cayrol cannot be accepted.*

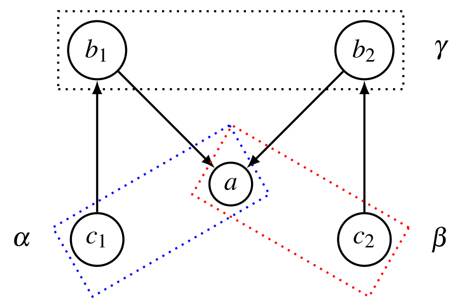

Figure 3: There is no single agent defending argument *a*

$\{c_1,c_2\}$ *collectively agent defend argument a, but they do not individually agent defend it. The agent admissible$_1$ extensions are* $\emptyset$, $\{c_1\}$, $\{c_2\}$ *and* $\{c_1,c_2\}$. *The only agent complete$_1$ extension is* $\{c_1,c_2\}$, *which is also the agent grounded$_1$ extension and the unique agent preferred$_1$ extension. There is no agent stable$_1$ extension. The agent admissible$_2$ extensions are* $\emptyset$, $\{c_1\}$, $\{c_2\}$, $\{c_1,c_2\}$ *and* $\{a,c_1,c_2\}$. *The only agent complete$_2$ extension is* $\{a,c_1,c_2\}$, *which is also the grounded$_2$ extension, the unique preferred$_2$ extension and stable$_2$ extension. Though Alice and Bob do not form a coalition in the sense that they defend each others' arguments, by using collective defense they can form a coalition in the sense that together they reinstate the argument in favor of eating meat.*

The following example illustrates another aspect of agent defense.

Example 7 (Agent defense)**.** *Consider Figure 4, where* $\mathcal{A} = \{a_1,a_2,b,c\}$, $\rightarrow = \{c \rightarrow b, b \rightarrow a_2, b \rightarrow a_1\}$, $\mathcal{S} = \{\alpha,\beta,\gamma\}$ *and* $\sqsubset = \{(a_1,\alpha),(a_2,\gamma),(b,\beta),(c,\alpha)\}$. *The unique individual (collective) agent complete extension, grounded extension and preferred extension is* $\{a_1,c\}$. *There is no stable extension. When we compute extensions using SCC-recursion,* *we first consider argument c, then argument b, and finally argument a_1 and a_2. When accepting c, we cannot simply remove b.*

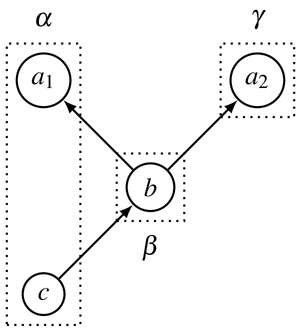

Figure 4: Agent defense

In the following three sections, we introduce several other kinds of semantics based on various kinds of reductions.

4 Social Agent Semantics

In this section, we introduce so-called social semantics, which is based on a reduction to preference-based argumentation for each argument, counting the number of agents that have the argument. It thus interprets agent argumentation as a kind of voting, as studied in social choice theory or judgment aggregation. It is not the only way to define social agent semantics, but given the formal setting we have adopted, it seems the simplest and most natural possibility.

We first give the definition of a preference-based argumentation framework.

Definition 9. (Preference-based argumentation framework) *A preference-based argumentation framework (PAF) is a 3-tuple $\langle \mathcal{A}, \rightarrow, \succ \rangle$ where $\mathcal{A}$ is a set of arguments, $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation, and $\succ$ is a partial order (irreflexive and transitive) over $\mathcal{A}$ called preference relation.*

Amgoud and Vesic (2014) introduce two different reductions of preference, while van der Torre and Vesic (2017) introduce two more. We refer to these papers for an explanation and motivation, and illustrate the difference between the reductions in Example 8 below.

Definition 10 (Reductions of PAF to AF (PR))**.** *Given a $PAF = \langle \mathcal{A}, \rightarrow, \succ \rangle$:*

- $PR_1(PAF) = \langle \mathcal{A}, \rightarrow' \rangle$, *where* $\rightarrow' = \{a \rightarrow' b | a \rightarrow b, b \not\succ a\}$.
- $PR_2(PAF) = \langle \mathcal{A}, \rightarrow' \rangle$, *where* $\rightarrow' = \{(a \rightarrow' b | a \rightarrow b, b \not\succ a \text{ or } b \rightarrow a, \text{ not } a \rightarrow b, a \succ b\}$.
- $PR_3(PAF) = \langle \mathcal{A}, \rightarrow' \rangle$, *where* $\rightarrow' = \{a \rightarrow' b | (a \rightarrow b, b \not\succ a \text{ or } a \rightarrow b, \text{ not } b \rightarrow a\}$.
- $PR_4(PAF) = \langle \mathcal{A}, \rightarrow' \rangle$, *where* $\rightarrow' = \{a \rightarrow' b | a \rightarrow b, b \not\succ a, \text{ or } b \rightarrow a, \text{ not } a \rightarrow b, a \succ b, \text{ or } a \rightarrow b, \text{ not } b \rightarrow a\}$.

In social agent semantics, an argument is preferred to another argument if it belongs to more agents. The reduction from AAF to PAF is used as an intermediary step for social agent semantics.

Definition 11 (Social Reductions of AAF to PAF (SAP))**.** *Given an $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, $SAP(AAF) = \langle \mathcal{A}, \rightarrow, \succ \rangle$ with $\succ = \{a \succ b || \mathcal{S}_a| > |\mathcal{S}_b|\}$.*

There are four definitions of social reduction, and σ is in $\{c, g, p, s\}$, thus, we have sixteen social agent semantics.

Definition 12 (Social Reductions of AAF to AF (SR))**.** *Given an $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, $SR_i(AAF) = PR_i(SAP(AAF))$, and PR_i is one of the four reductions of PAF to AF, where the semantics $\delta(AAF) = \sigma(SR_i(AAF)) = \sigma(PR_i(SAP(AAF)))$ for $i \in \{1, 2, 3, 4\}$.*

Example 8 (Social reasoning)**.** *Consider the agent argumentation framework (AAF) on the left in Figure 5, where $\mathcal{A} = \{a, b\}$, $\rightarrow = \{a \rightarrow b\}$, $\mathcal{S} = \{\alpha, \beta\}$ and $\sqsubset = \{(a, \alpha), (b, \alpha), (b, \beta)\}$. Argument b is preferred to argument a because it belongs to more agents. The preference-based argumentation framework (PAF) is visualized to the right of the AAF in Figure 5: $\mathcal{A} = \{a, b\}$, $\rightarrow = \{a \rightarrow b\}$, and $\succ = \{b \succ a\}$. To the right of PAF, there are four corresponding argumentation frameworks (AFs) after SR_1 to SR_4, the extensions of each are listed in Table 1.*

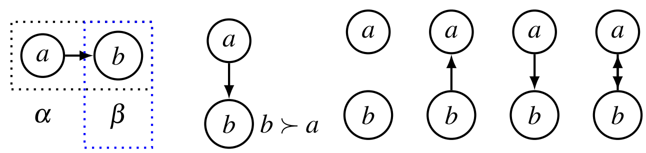

Figure 5: Social reduction

Sem.	$\mathbb{C}$	$\mathbb{G}$	$\mathbb{P}$	$\mathbb{S}$
SR_1	$\{\{a,b\}\}$	$\{\{a,b\}\}$	$\{\{a,b\}\}$	$\{\{a,b\}\}$
SR_2	$\{\{b\}\}$	$\{\{b\}\}$	$\{\{b\}\}$	$\{\{b\}\}$
SR_3	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$
SR_4	$\{\emptyset, \{a\}, \{b\}\}$	$\emptyset$	$\{\{a\}, \{b\}\}$	$\{\{a\}, \{b\}\}$

Table 1: The semantics of four corresponding argumentation frameworks (AFs) after SR_1 to SR_4. We refer to Dung's semantics as follows: Complete ($\mathbb{C}$), Grounded ($\mathbb{G}$), Preferred ($\mathbb{P}$), Stable ($\mathbb{S}$), and the same convention holds for all the others.

5 Agent Reduction Semantics

In this section, we introduce the third class of semantics. Agent reductions take the perspective of each agent and create extensions accordingly. In an abstract sense, an agent prefers its own arguments over the arguments of the other agents. It is again based on a reduction of agent argumentation frameworks to preference-based argumentation frameworks, just like social agent semantics, but now in a completely different way. One difference between social reductions in the previous section and the agent reductions in this section is that in the previous section, there is only reduction AF for every AAF, whereas in this section there is a set of such reductions, one for each agent, and then we take the union of all the reductions. Again, as in the previous section, the four kinds of reduction of preference-based argumentation frameworks lead to four kinds of agent reductions.

Definition 13 (Agent Reductions of AAF to PAF(AAP)). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*,* $AAP(AAF, \alpha) = \langle \mathcal{A}, \rightarrow, \succ \rangle$ *with* $\succ = \{a \succ b | a \sqsubset \alpha$ *and not* $b \sqsubset \alpha\}$*.*

As in social agent semantics, there are four definitions of agent reductions, and σ is in $\{c, g, p, s\}$. Thus, we have sixteen agent reduction semantics.

Definition 14 (Agent Reductions of AAF to AF (AR)). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, for* $\alpha \in \mathcal{S}$*,* PR_i *is one of the four reductions of PAF to AF, where the semantics* $\delta(AAF) = \sigma(AR_i(AAF)) = \sigma(\bigcup_{\alpha \in \mathcal{S}} PR_i(AAP(AAF, \alpha)))$ *for* $i \in \{1, 2, 3, 4\}$*. For* $AF_1 = \langle \mathcal{A}_1, \rightarrow_1 \rangle$ *and* $AF_2 = \langle \mathcal{A}_2, \rightarrow_2 \rangle$*, let* $AF_1 \cup AF_2 = \langle \mathcal{A}_1 \cup \mathcal{A}_2, \rightarrow_1 \cup \rightarrow_2 \rangle$*.*

Example 9 (Agent reduction). *Reconsider the AAF on the left in Figure 5. Firstly, consider the reduction for agent* β*. We have that argument b is preferred over argument a, thus, we get the same PAF as in Figure 5, though for a very different reason compared to that from social reduction. For agent* α*, the PAF makes all arguments equivalent, and the AF is simply the same as for the trivial reduction. To compute the agent extensions of the AAF, we take the union of the reductions for each agent. The AFs of* AR_i *are the union of the AFs of SRi in Table 1 with the AF in which a attacks b (the reduction for agent* α*). Thus,* $AR_1 = AR_3 = \langle \{a, b\}, \{a \rightarrow b\} \rangle$*, and* $AR_2 = AR_4 = \langle \{a, b\}, \{a \rightarrow b, b \rightarrow a\} \rangle$*. For instance, after* AR_1*, the AF of agent* α *is* $AR_1(AAF, \alpha) = \langle \{a, b\}, \{a \rightarrow b\} \rangle$*, while* $AR_1(AAF, \beta) = \langle \{a, b\}, \{\emptyset\} \rangle$*, so the union is* $\langle \{a, b\}, \{a \rightarrow b\} \rangle$*, and then we compute the extensions of this union. The result is Table 2 below for the sixteen agent reduction semantics we consider.*

Sem.	$\mathbb{C}$	$\mathbb{G}$	$\mathbb{P}$	$\mathbb{S}$
AR_1	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$
AR_2	$\{\emptyset, \{a\}, \{b\}\}$	$\emptyset$	$\{\{a\}, \{b\}\}$	$\{\{a\}, \{b\}\}$
AR_3	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$
AR_4	$\{\emptyset, \{a\}, \{b\}\}$	$\emptyset$	$\{\{a\}, \{b\}\}$	$\{\{a\}, \{b\}\}$

Table 2: The semantics of four corresponding argumentation frameworks (AF) after AR_1 to AR_4.

6 Agent Filtering Semantics

In this section, we introduce the fourth kind of semantics for agent argumentation frameworks. Agent filtering semantics remove arguments that do not belong to an agent (OrphanReduction), or they remove attacks that do not belong to an agent (NotBothReduction), where an attack belongs to an agent if both the attacker and the attacked argument belong to the agent.

Definition 15 (Agent Reductions of AAF to AF). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*:*

- *OrphanRemoval (OR):* $OR(AAF) = \langle \mathcal{A}', \rightarrow' \rangle$ *where* $\mathcal{A}' = \{a | \exists \alpha \in \mathcal{S}$ *such that* $a \sqsubset \alpha, \}, \rightarrow \cap \mathcal{A}' \times \mathcal{A}'$*.*
- *NotBothReduction (NBR):* $NBR(AAF) = \langle \mathcal{A}, \rightarrow' \rangle$ *where* $\rightarrow' = \{(a \rightarrow b | \exists \alpha \in \mathcal{S}$ *such that* $a \sqsubset \alpha$*, and* $b \sqsubset \alpha\}$*.*

Example 10 (Epistemic reasoning). *Consider the two AAFs in Figure 6. For the figure on the left, we may say that argument a is not known, as there is no agent that has it, and for the figure on the right, we may say that the attack is unknown, because there is no agent that has both arguments a and b. The filtering methods remove such unknown arguments (OrphanReduction) and unknown attacks (NotBothReduction).*

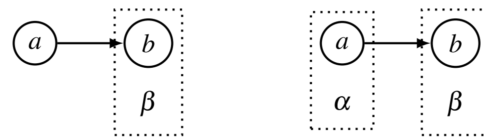

Figure 6: Unknown

7 Traditional Principles

In this section, we repeat six important principles from the literature. As the baseline for the principles, we also include Dung's semantics. It is based on the so-called trivial reduction, which simply ignores the agents and the relation between agents and arguments.

Definition 16 (Trivial Reduction (TR)). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle, TR(AAF) = \langle \mathcal{A}, \rightarrow \rangle$*.*

Principle 1 (Conflict-free (Baroni and Giacomin 2007)). *An agent semantics* δ *satisfies the conflict-free principle iff for every* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, for all* $E \in \delta(AAF)$*, there are no arguments a and b in E such that a attacks b.*

The conflict-free principle reflects the intuitive idea that an extension contains the arguments that can be accepted together, and that the conflicting arguments cannot be included in the same extension, while the admissibility principle reflects that all arguments are defended.

Principle 2 (Admissibility (Baroni and Giacomin 2007)). *An agent semantics* δ *satisfies the admissibility principle iff for every* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, every* $E \in \delta(AAF)$ *is admissible in* $\langle \mathcal{A}, \rightarrow \rangle$*.*

Directionality and SCC-recursiveness are introduced by Baroni, Giacomin, and Guida (2005). These principles reflect the idea that we can decompose an argumentation framework into sub-frameworks so that the semantics can be defined locally. For the directionality principle, they first introduce the definition of an unattacked set.

Definition 17 (Unattacked Set). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, a set* $\mathcal{U}$ *is unattacked iff there exists no* $a \in \mathcal{A} \backslash \mathcal{U}$ *such that a attacks an argument in* $\mathcal{U}$*. The set of unattacked sets in AAF is denoted as* $\mathcal{US}(AAF)$*.*

Definition 18 (Restriction). *Given an* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, and let* $\mathcal{U} \subseteq \mathcal{A}$ *be a set of arguments, the restriction of AAF to* $\mathcal{U}$ *is the agent abstract framework* $AAF{\downarrow_{\mathcal{U}}} = \langle \mathcal{U}, \rightarrow \cap \mathcal{U} \times \mathcal{U}, \mathcal{S}, \sqsubset \cap \mathcal{U} \times \mathcal{S} \rangle$*.*

Principle 3 (Directionality (Baroni and Giacomin 2007)). *An agent semantics* δ *satisfies the directionality principle iff for every* $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$*, for every* $\mathcal{U} \in \mathcal{US}(AAF)$*, it holds that* $\delta(AAF{\downarrow_{\mathcal{U}}}) = \{E \cap \mathcal{U} | E \in \delta(AAF)\}$*.*

Proposition 1. *Agent* $stable_1$ *semantics and agent* $stable_2$ *semantics (Def. 8) do not satisfy Principle 3.*

Proof. We use a counter-example to prove Proposition 1. Assume an $AAF = \langle\{a_1, a_2, a_3, b\}, \{b \rightarrow a_3, a_3 \rightarrow a_1, a_1 \rightarrow a_2, a_2 \rightarrow a_3\}, \{\alpha\}, \{b \sqsubset \alpha, a_1 \sqsubset \alpha, a_2 \sqsubset \alpha, a_3 \sqsubset \alpha\}\rangle$. The unattacked set of arguments is $U = \{b\}$. The stable extension of $(AAF \downarrow U)$ is $\{b\}$. However, there is no stable extension of this AAF. $\delta(\mathrm{AAF}{\downarrow_{\mathcal{U}}}) \neq \{E \cap \mathcal{U} | E \in \delta(AAF)\}$, thus, Agent $stable_1$ semantics and agent $stable_2$ semantics (Def. 8) do not satisfy Principle 3. □

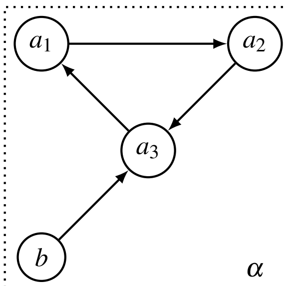

Figure 7: A counterexample to prove Proposition 1

The SCC-recursiveness is based on the notion of strongly connected components from graph theory.

Definition 19 (Strongly Connected Component)**.** *Let an* $AAF = \langle\mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset\rangle$. *The binary relation of path-equivalence between nodes, denoted as* $PE_{AAF} \subseteq (\mathcal{A} \times \mathcal{A})$*, is defined as follows:*

- *for every* $a \in \mathcal{A}, (a, a) \in PE_{AAF}$
- *given two distinct arguments* $a, b \in \mathcal{A}$*, we say that* $((a, b) \in PE_{AAF}$ *iff there is a path from a to b and a path from b to a.*

The strongly connected components of AAF are the equivalence classes of arguments under the relation of path-equivalence. The set of strongly connected components is denoted by $SCCS_{AAF}$.

Given an argument $a \in \mathcal{A}$, notation $SCCS_{AAF}(a)$ stands for the strongly connected component that contains a. In the particular case where the argumentation framework is empty, i.e., $AAF = \langle\emptyset, \emptyset, \emptyset, \emptyset\rangle$, we assume that $SCCS_{AAF} = \{\emptyset\}$. The choice of extensions of the antecedent strongly connected components determines a partition of the arguments of a strongly connected component S into three subsets: defeated (D), provisionally defeated (P) and undefeated (U) (Baroni, Giacomin, and Guida 2005).

In words, the set $D_{AAF}(S, E)$ consists of the arguments of S being attacked by E from outside S, the set $U_{AAF}(S, E)$ consists of the arguments in S that are not attacked by E from outside S and are defended by E and $P_{AFF}(S, E)$ consists of the arguments in S that are not attacked by E from outside S and are not defended by E.

Definition 20 (D, P, U, UP)**.** *Given an* $AAF = \langle\mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset\rangle$*, a set* $E \subseteq \mathcal{A}$ *and a strongly connected component* $S \in SCCS_{AAF}$

- $D_{AAF}(S, E) = \{a \in S | (E \cap S^-_{out})$ *attacks* $a\}$
- $P_{AAF}(S, E) = \{a \in S | (E \cap S^-_{out})$ *does not attack a and* $\exists b \in (S^-_{out} \cap a^-)$ *such that E does not attack* $b\}$.
- $U_{AAF}(S, E) = S \backslash (D_{AAF}(S, E) \cup P_{AAF}(S, E))$
- $UP_{AAF}(S, E) = U_{AAF}(S, E) \cup P_{AAF}(S, E)$.

We now present the notion of SCC-recursiveness, which was introduced by Baroni, Giacomin, and Guida (2005).

Principle 4. (SCC-recursiveness (Baroni, Giacomin, and Guida 2005)) *Agent semantics* δ *satisfies the SCC-recursiveness principle iff for every* $AAF = \langle\mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset\rangle$*, we have* $\delta(AAF) = \mathcal{G}(AAF, \mathcal{A})$*, where for every AAF and for every set* $\mathcal{C} \subseteq \mathcal{A}$*, the function* $\mathcal{G}(AAF, \mathcal{C}) \subseteq 2^{\mathcal{A}}$ *is defined as follows: for every* $E \subseteq \mathcal{A}$, $E \in \mathcal{G}(AAF, \mathcal{C})$ *iff*

- *when* $|SCCS_{AAF}| = 1$, $E \in \mathcal{B}(AAF, \mathcal{C})$,
- *otherwise,* $\forall S \in SCCS_{AAF}$, $(E \cap S) \in \mathcal{G}(AAF{\downarrow_{UP_{AAF}(S,E)}}, U_{AAF}(S, E) \cap \mathcal{C})$,

where $\mathcal{B}(AAF, \mathcal{C})$ *is a function called a base function that given an* $AAF = \langle\mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset\rangle$*, such that* $|SCCS_{AAF}| = 1$ *and a set* $\mathcal{C} \subseteq \mathcal{A}$ *gives a subset of* $2^{\mathcal{A}}$.

Baumann, Brewka, and Ulbricht (2020) introduce the modularization principle. By definition, AAF^E is the sub-framework of AAF obtained by removing the so-called range of E, the corresponding attacks, and the relation with agents.

Definition 21 (E-reduct)**.** *Given an* $AAF = \langle\mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset\rangle$ *and* $E \subseteq \mathcal{A}$*, let* $E^+ = \{a \in \mathcal{A} | E$ *attacks* $a\}$, $E^\oplus = E \cup E^+$ *and* $E^* = \mathcal{A} \backslash E^\oplus$. *The E-reduct of AAF is the* $AAF^E = \langle E^*, R \cap (E^* \times E^*), \mathcal{S}, \sqsubset \cap (\mathcal{S} \times E^*)\rangle$.

Principle 5 (Modularity)**.** *An agent semantics* δ *satisfies modularization if for any AAF, we have* $E \in \delta(AAF)$ *and* $E' \in \delta(AAF^E)$ *implies* $E \cup E' \in \delta(AAF)$.

The modularity principle is related to the robustness principles of Rienstra et al. (2020), which consider the addition and removal of arguments and attacks. We consider here only argument removal, which we call argument modularity.

Table 3 provides full analysis of the traditional five principles. The first line of the trivial reduction lists a well-known analysis of which of these principles hold for Dung's semantics. Unsurprisingly, several easy examples we have already discussed in this paper show that few of the traditional principles hold for agent semantics. This is particularly a problem for SCC-recursiveness and modularity, because we cannot apply the corresponding recursive algorithm to compute the semantics. In the next section, we therefore introduce some variants of admissibility, SCC-recursion and modularity that are based on agent defense.

8 Variants of Traditional Principles

The agent admissibility principle is a straightforward adaptation of the admissibility principle, in which defense is replaced by agent defense. Since there are two kinds of admissibility, one for individual defense and one for collective defense, we end up with two agent admissibility principles.

Sem.	P1	P2	P3	P4	P5
TR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CPS}$
Sem_1	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	×	×
Sem_2	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGP}$	×	×
SR_1	×	×	$\mathbb{CGP}$	×	×
SR_2	$\mathbb{CGPS}$	×	×	×	×
SR_3	$\mathbb{CGPS}$	×	$\mathbb{CGP}$	$\mathbb{CGPS}$	×
SR_4	$\mathbb{CGPS}$	×	×	×	×
AR_1	×	×	$\mathbb{CGP}$	×	×
AR_2	$\mathbb{CGPS}$	×	×	×	×
AR_3	$\mathbb{CGPS}$	×	$\mathbb{CGP}$	$\mathbb{CGPS}$	×
AR_4	$\mathbb{CGPS}$	×	×	×	×
OR	$\mathbb{CGPS}$	×	$\mathbb{CGP}$	$\mathbb{CGPS}$	$\mathbb{CPS}$
NBR	×	×	$\mathbb{CGP}$	×	$\mathbb{CPS}$

Table 3: Comparison of reductions and traditional principles. When a principle is never satisfied by a certain reduction for all semantics, we use the × symbol, and we use a question mark to represent an open problem. P1 refers to Principle 1, and the same convention holds for all the others.

Principle 6 (Agent Admissibility$_1$)**.** *An agent semantics δ satisfies the agent admissibility$_1$ principle iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, every $E \in \delta(AAF)$ is agent admissible$_1$.*

Principle 7 (Agent Admissibility$_2$)**.** *An agent semantics δ satisfies the agent admissibility$_2$ principle iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, every $E \in \delta(AAF)$ is agent admissible$_2$.*

Proposition 2. *AR_1 to AR_4 and SR_1 to SR_4 do not satisfy Principle 6 and 7 for complete semantics.*

Proof. We use the agent argumentation framework in Figure 2 as a counter-example to prove Proposition 2. $SAP(AAF) = AAP(AAF) = \langle \{a,b,c\}, \{a \rightarrow b, b \rightarrow c\}, \emptyset \rangle$. $AR_i(AAF) = SR_i(AAF) = \langle \{a,b,c\}, \{a \rightarrow b, b \rightarrow c\} \rangle$. The complete extension of $AR_i(AAF)$ and $SR_i(AAF)$ is $\{a,c\}$. However, a cannot agent defend c, and $\{a,c\}$ is not agent admissible. Thus, AR_1 to AR_4 and SR_1 to SR_4 do not satisfy Principle 6 and 7 for complete semantics. □

The agent SCC-recursiveness principles are also adapted by replacing defense with agent defense, and again we end up with two principles for individual and collective defense. What needs to be adapted is the definition of P, the provisionally defeated arguments. Roughly, P stands for the case that an argument is not defended against b in E outside of S. Likewise, AP stands for the case that an argument a is not agent defended$_i$ against b in E from outside S.

To define agent SCC-recursiveness, we define AD_i, AP_i, AU_i, and AUP_i under individual agent defense and collective agent defense.

Definition 22 (AD_i, AP_i, AU_i, AUP_i)**.** *Given an $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, a set $E \subseteq \mathcal{A}$ and a strongly connected component $S \in SCCS_{AAF}$, we define:*

- $AD_{iAAF}(S,E) = D_{iAAF}(S,E)$
- $AP_{1AAF}(S,E) = \{a \in S | (E \cap S^-_{out})$ *does not attack a, and* $\forall \alpha \in \mathcal{S}_a, \exists b \in (S^-_{out} \cap a^-)$ *such that* $E \cap \mathcal{A}_{\mathcal{S}_a}$ *does not attack b.*$\}$
- $AP_{2AAF}(S,E) = \{a \in S | (E \cap S^-_{out})$ *does not attack a and* $\exists b \in (S^-_{out} \cap a^-)$ *such that* $\forall \alpha$ *in* $\mathcal{S}_a, E \cap \mathcal{A}_\alpha$ *does not attack b.*$\}$
- $AU_{iAAF}(S,E) = S \backslash (AD_{iAAF}(S,E) \cup AP_{iAAF}(S,E))$
- $AUP_{iAAF}(S,E) = AU_{iAAF}(S,E) \cup AP_{iAAF}(S,E)$.

Principle 8 (Agent SCC-recursiveness$_1$)**.** *An agent semantics δ satisfies the agent SCC-recursiveness$_1$ principle iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, we have $\delta(AAF) = \mathcal{G}(AAF, \mathcal{A})$, where for every AAF and for every set $\mathcal{C} \subseteq \mathcal{A}$, the function $\mathcal{G}(AAF, \mathcal{C}) \subseteq 2^{\mathcal{A}}$ is defined as follows: for every $E \subseteq \mathcal{A}$, $E \in \mathcal{G}(AAF, \mathcal{C})$ iff*

- *when $|SCCS_{AAF}| = 1$, $E \in \mathcal{B}(AAF, \mathcal{C})$,*
- *otherwise, $\forall S \in SCCS_{AAF}$, $(E \cap S) \in \mathcal{G}(AAF\downarrow_{AUP_{1AAF}(S,E)}, AU_{1AAF}(S,E) \cap \mathcal{C})$,*

Principle 9 (Agent SCC-recursiveness$_2$)**.** *An agent semantics δ satisfies the agent SCC-recursiveness$_2$ principle iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, we have $\delta(AAF) = \mathcal{G}(AAF, \mathcal{A})$, where for every AAF and for every set $\mathcal{C} \subseteq \mathcal{A}$, the function $\mathcal{G}(AAF, \mathcal{C}) \subseteq 2^{\mathcal{A}}$ is defined as follows: for every $E \subseteq \mathcal{A}$, $E \in \mathcal{G}(AAF, \mathcal{C})$ iff*

- *when $|SCCS_{AAF}| = 1$, $E \in \mathcal{B}(AAF, \mathcal{C})$,*
- *otherwise, $\forall S \in SCCS_{AAF}$, $(E \cap S) \in \mathcal{G}(AAF\downarrow_{AUP_{2AAF}(S,E)}, AU_{2AAF}(S,E) \cap \mathcal{C})$,*

Table 4 shows the comparison between agent semantics and agent admissibility principles and agent SCC-recursion. This is important, since it proves that we can have an efficient SCC-recursiveness algorithm for the new agent semantics. The table also shows that for P7 and P9, collective defense implies individual defense. Finally, the table shows that the adapted principles, like the traditional ones, are not very useful for distinguishing between the reduction-based semantics, i.e. the social agent semantics, the agent reduction semantics, and the agent filtering semantics. Therefore, we introduce some new principles in the remainder of the paper.

9 New Agent Principles

In this section, we introduce eight new principles to distinguish agent semantics. Principle 10 says that if more agents adopt an argument that is accepted, this does not affect the extension.

Principle 10 (AgentAdditionPersistence)**.** *An agent semantics δ satisfies AgentAdditionPersistence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, $E \in \delta(AAF)$, $\alpha \in \mathcal{S}$ and $a \in E$, we have $E \in \delta(\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \cup (a, \alpha) \rangle)$.*

Proposition 3. *AR_1 to AR_4 and SR_1 to SR_4 do not satisfy Principle 10 for complete semantics.*

Proof. We use a counter-example to prove Proposition 3. Assume $AAF_1 = \langle \{a,b\}, \{a \rightarrow b, b \rightarrow a\}, \{\alpha, \beta\}, \{a \sqsubseteq \alpha, b \sqsubseteq \beta\} \rangle$, $AR_i(AAF_1) = SR_i(AAF_1) = \langle \{a,b\}, \{a \rightarrow b, b \rightarrow a\} \rangle$. Let $AAF_2 = \langle \{a,b\}, \{a \rightarrow b, b \rightarrow a\}, \{\alpha, \beta\}, \{a \sqsubseteq$

Sem.	P6	P7	P8	P9
TR	×	×	×	×
Sem_1	$\mathbb{CGPS}$	×	$\mathbb{CGPS}$	×
Sem_2	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$
SR_1	×	×	×	×
SR_2	×	×	×	×
SR_3	×	×	×	×
SR_4	×	×	×	×
AR_1	×	×	×	×
AR_2	×	×	×	×
AR_3	×	×	×	×
AR_4	×	×	×	×
OR	×	×	×	×
NBR	×	×	×	×

Table 4: Comparison of the reductions and agent admissibility principles, and agent SCC-recursion.

$\alpha, a \sqsubseteq \beta, b \sqsubseteq \beta\}\}$. $AR_i(AAF_2) = SR_i(AAF_2) = \langle\{a,b\},\{a \rightarrow b\}\rangle$ The complete extensions of AAF_1 are $\{a\}$ and $\{b\}$, while the complete extension of AAF_2 is $\{a\}$. Thus, AR_1 to AR_4 and SR_1 to SR_4 do not satisfy Principle 10 for complete semantics. □

Proposition 4. *OR satisfies Principle 10 and Principle 11 for all the semantics.*

Proof. Assume an $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, $OR(AAF) = \langle \mathcal{A}', \rightarrow' \rangle$. For any extension $E \in \delta(AAF), \forall a \in E$, there exists an agent α such that a $\sqsubseteq \alpha$. By definition, we find that any argument in the extension has at least one agent, so attaching more agents to AAF will not affect $OR(AAF)$. Thus, OR satisfies Principle 10 and Principle 11 for all the semantics. □

Principle 11 reflects the same idea as principle 10, but is based on the assumption that a is accepted in all extensions.

Principle 11 (AgentAdditionUniversalPersistence)**.** *An agent semantics δ satisfies AgentAdditionUniveralPerisitence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, for $\forall E \in \delta(AAF)$, $\alpha \in \mathcal{S}$ and $a \in E$, we have $\forall E \in \delta(\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \cup (a, \alpha)\rangle)$, $a \in E$.*

Principle 12 reflects a principle we expect to hold for all agent semantics. It reflects anonymity: if we permute the agents, it does not affect the extensions. It is analogous to language independence for arguments defined by Baroni and Giacomin (2007).

Principle 12 (PermutationPersistence)**.** *An agent semantics δ satisfies PermutationPersistence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$ and $AAF' = \langle \mathcal{A}, \rightarrow, \mathcal{S}', \sqsubset' \rangle$, and where $\mathcal{S}$ and $\mathcal{S}'$ are two different ordered sets with common elements, we have $\delta(AAF) = \delta(AAF')$.*

Principle 13 reflects that if the arguments of two agents do not attack each other, we can merge these agents into one single agent. It does not hold for agent defense semantics, because new agent defenses may be created.

Principle 13 (MergeAgent)**.** *An agent semantics δ satisfies MergeAgent iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, $\exists \alpha, \beta \in \mathcal{S}$, for $\forall a \in \mathcal{A}_\alpha$ and $\forall b \in \mathcal{A}_\beta$, a does not attack b, b does not attack a, we have AAF' by changing $\forall$ $a \sqsubset \alpha$ to $a \sqsubset \beta$, and $\delta(AAF) = \delta(AAF')$.*

Principle 14 reflects that if two agents have the same arguments, we can remove one of these agents without changing the extensions. This represents the opposite of social semantics, where the number of the agents makes a difference.

Principle 14 (RemovalAgentPersistence)**.** *An agent semantics δ satisfies RemovalAgentPersistence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, for $\mathcal{S}_\alpha = \mathcal{S}_\beta$, we have $\delta(\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle) = \delta(\langle \mathcal{A}, \rightarrow, \mathcal{S} \backslash \alpha, \sqsubset \setminus \sqsubset_\alpha \rangle) = \delta(\langle \mathcal{A}, \rightarrow, \mathcal{S} \backslash \beta, \sqsubset \setminus \sqsubset_\beta \rangle)$.*

Principle 15 is inspired by social agent semantics. It states that for two argumentation frameworks with the same arguments and attacks, if for every argument the number of agents holding that argument is the same, then the extensions are the same.

Principle 15 (AgentNumberEquivalence)**.** *An agent semantics δ satisfies AgentNumberEquivalence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$ and an $AAF' = \langle \mathcal{A}, \rightarrow, \mathcal{S}', \sqsubset' \rangle$, for $\forall a \in \mathcal{A}$, $|\mathcal{S}_a| = |\mathcal{S}'_a|$, we have $\delta(AAF) = \delta(AAF')$.*

Principle 16 is inspired by agent reduction semantics. It states that if the set of the arguments of an agent is conflict-free, then there is an extension containing those arguments.

Principle 16 (Conflict-freeInvolvement)**.** *An agent semantics δ satisfies Conflict-freeInvolvement iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, for $\forall \alpha \in \mathcal{S}, \mathcal{A}_\alpha$ is conflict-free, there is an E, we have $\mathcal{A}_\alpha \subseteq E$.*

Principle 17 is inspired by OrphanReduction semantics. It states that if we have arguments that do not belong to any agents, then they can be removed from the framework without affecting the extensions.

Principle 17 (RemovalArgumentPersistence)**.** *An agent semantics δ satisfies RemovalArgumentPersistence iff for every $AAF = \langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle$, and for $\nexists \alpha \in \mathcal{S}$ and $a \sqsubset \alpha$, we have $\delta(\langle \mathcal{A}, \rightarrow, \mathcal{S}, \sqsubset \rangle) = \delta(\langle \mathcal{A} \backslash a, \rightarrow \setminus \rightarrow_a), \mathcal{S}, \sqsubset \rangle)$.*

Most of the principles are independent. In particular, Principle 3 (for short P3), P5, P12, P13, P14, P15, P16, and P17 are independent. On the other hand, other principles have inner relationship among themselves. For example, if a semantic satisfies P2, then it must satisfy P1. We denote this observation as P2 ⇒ P1. And all the observations are listed as: P2 ⇒ P1, P4 ⇒ P8 ⇒ P9, P6 ⇒ P7, P10 ⇒ P11.

In the resulting Table 5, all agent semantics satisfy P12. Perhaps surprisingly, both social agent semantics and agent reduction semantics does not satisfy P10, while trivial reduction semantics, social agent semantics and agent filtering semantics satisfy P13. Moreover, all agent semantics except the social agent semantics satisfy P14. No semantics satisfies P16. As expected, only OrphanRemoval satisfies P17. The only semantics that are not distinguished yet concern the use of different preference reductions, or different Dung semantics. To distinguish them, the principles proposed in preference-based argumentation and in Dung's semantics can be used. In that sense, the principle-based anal-

Sem.	P10	P11	P12	P13	P14	P15	P16	P17
TR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\times$
Sem_1	$\mathbb{S}$	$\mathbb{S}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
Sem_2	$\mathbb{S}$	$\mathbb{S}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
SR_1	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$
SR_2	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$
SR_3	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$
SR_4	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$
AR_1	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
AR_2	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
AR_3	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
AR_4	$\times$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$
OR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\mathbb{CGPS}$
NBR	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\mathbb{CGPS}$	$\times$	$\times$	$\times$

Table 5: Comparison between the resolutions and new agent principles.

ysis in this paper is complementary to the principle-based analysis in the other areas.

10 Related Work

Our work builds on a rich literature on formal argumentation and dialogue, and we can mention here only a few of the most directly related papers.

From the four kinds of agent argumentation semantics introduced in this paper, we are not aware of other approaches that adapt Dung's basic concepts directly, as we have done with individual and collective agent defense. There are other variants of semantics that adapt these notions, such as weak defense for weak admissibility semantics (Baumann, Brewka, and Ulbricht 2020), but that is not based on the agent metaphor.

The most related work is in social agent semantics. Leite and Martins (2011) introduce an abstract model of argumentation where agents can vote in favor of and against an issue. They define an abstract argumentation framework as a triple $\langle \mathcal{A}, \mathcal{R}, \mathcal{V} \rangle$, where $\mathcal{V} \to N \times N$ is a total function mapping each argument to its number of positive (Pro) and negative (Con) votes. Our paper, on the other hand, only considers positive votes. Caminada and Pigozzi (2011) capture the notion that individual members need to defend the collective decision in order to reach a compatible outcome, and propose to address judgment aggregation by combining different individual evaluations of the situation represented by an argumentation framework. Hunter, Polberg, and Thimm (2020) take an epistemic approach to probabilistic argumentation, where the arguments are believed or not believed in terms of different degrees, providing an alternative to the subtle standard Dung framework.

Concerning agent reduction semantics, several authors build on the local functions introduced by Baroni, Giacomin, and Guida (2005), and further developed by Baroni et al. (2014). Giacomin (2017) shows how to use this theory in multi-agent systems. The results in these papers indicate that such generalizations often become equivalent to Dung; and Arisaka, Satoh, and van der Torre (2017) extend the agent argumentation frameworks with coalitions among the agents. Rienstra et al. (2011) consider the case where the agents may have different semantics, for example one agent uses grounded semantics and another agent uses preferred semantics. Furthermore, Kontarinis and Toni (2015) analyse the identification of the malicious behavior of agents in the form of bipolar argumentation frameworks, which together with the work of Panisson et al. (2018) may inspire work on agent reduction semantics based on trustfulness.

In this paper, we build on the principle-based approach to preference-based argumentation developed by Amgoud and Cayrol (2002) together with several co-authors over the past fifteen years. In particular, the work of Amgoud and Vesic (2014) and the work of Kaci, van derTorre, and Villata (2018) have inspired us, although the principles discussed in our paper are mostly different from those studied in preference-based argumentation. In earlier work, two of the authors have related their axiomatic approach to the analysis of bipolar argumentation (Yu and Van der Torre 2020), and there are also close relations with the study of robustness principles (Rienstra et al. 2020).

11 Future Work

Prakken (2018) distinguishes between argumentation as inference and argumentation as dialogue. Abstract agent argumentation can bring elements of argumentation as dialogue into the foundations of argumentation as inference, and may help to bridge the gap between the two branches.

Within the formal setting we have adopted in this paper, many topics of further work present themselves. As always with the principle-based approach, we can introduce more semantics, for example by combining the ideas of the four classes, guided by the existing principles. We can also study more principles. We can find relations with other branches of logic and reasoning such as axiomatic approaches in social choice. Moreover, we can try to use principles to address the standard challenges of abstract argumentation, namely relating the abstract model to more structured forms of argumentation, and applying abstract argumentation, for instance, to legal reasoning.

For example, there is no agent semantics satisfying principle P16. We can define new semantics as a variant of Definition 14, which combines agent reductions in a new way. Instead of combining the frameworks, we can take the union of all the extensions to the individual frameworks: $\delta(AAF) = \sigma(\bigcup_{\alpha \in \mathbb{S}} PRi(AAP(AAF, \alpha)))$ for $i \in \{1,2,3,4\}$. $\delta'(AAF) = \bigcup_{\alpha \in \mathbb{S}} \sigma(PRi(AAP(AAF, \alpha)))$ for $i \in \{1,2,3,4\}$. If we use the definition of δ', Table 6 changes as follows.

Sem.	$\mathbb{C}$	$\mathbb{G}$	$\mathbb{P}$	$\mathbb{S}$
AR_1	$\{\{a\},\{a,b\}\}$	$\{\{a\},\{a,b\}\}$	$\{\{a\},\{a,b\}\}$	$\{\{a\},\{a,b\}\}$
AR_2	$\{\{a\},\{b\}\}$	$\{\{a\},\{b\}\}$	$\{\{a\},\{b\}\}$	$\{\{a\},\{b\}\}$
AR_3	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$	$\{\{a\}\}$
AR_4	$\{\emptyset,\{a\},\{b\}\}$	$\{\{a\},\emptyset\}$	$\{\{a\},\{b\}\}$	$\{\{a\},\{b\}\}$

A regular topic in abstract argumentation is to search for fragments with good computational properties, such as symmetric attack relations. Also with agent argumentation, we can study frameworks where: every argument is associated with at least one agent, every argument is associated with at most one agent, there are at most two agents, symmetric attack is possible, the arguments of each agent are conflict free, and so on.

Moreover, concerning the use of reductions in abstract argumentation, our paper raises the question of whether we can find a reduction to Dung's argumentation frameworks for agent defense semantics. While we have presented such reductions for all the other kinds of agent semantics, we have not yet found such a reduction for agent defense semantics. For such a reduction, we might also add auxiliary arguments, or we may introduce arguments for each pair of argument and agent.

Finally, one of the main challenges in the area of formal argumentation is the gap between abstract argumentation and dialogue. Caminada (2017) presents semantics of abstract argumentation that can be interpreted with regard to structured discussion in order to fill this gap. However, how to implement abstract argumentation with dynamic agent dialogue is still an open question.

12 Conclusion

As common in the principle-based approach to argumentation semantics, we have selected principles that distinguish agent semantics. In addition, we have added some principles that reflect important properties of agent semantics and that can be used to guide the development of future agent semantics. To be more specific, combining ideas from earlier work in abstract argumentation, our principle-based analysis has revealed several original insights. For example, a new twist was given to the fundamental role of defense and reinstatement in Dung's theory in the context of agent defense semantics.

Moreover, a new variant of SCC-recursiveness has been introduced, leading to an SCC-recursive algorithm for agent defense semantics. Since the other approaches are based on reductions, the traditional SCC-recursive algorithms can be used. Finally, the future work section illustrates that our formal framework not only serves as a tool for organizing existing work in the area, but also provides a solid foundation for further work in this direction.

There is quite some variety in agent semantics. In this paper, the priority and hierarchy of semantics and which kind of semantics is used depends on its application. Also they can be combined. For example, we can use both filtering and agent defense to remove unknown arguments or unknown attacks, and defend an argument put forward by an agent. We can use the principles to elect the most suitable semantics for an application, and in fact that is one of the most important uses of principles. For example, Principle 17 (Removal Argument Persistence) can help us to elect agent filtering semantics for an application.

Acknowledgments

We thank Beishui Liao for providing valuable comments on an earlier draft of this paper. Liuwen Yu has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie ITN EJD grant agreement. No 814177. Leendert van der Torre acknowledges financial support from the Fonds National de la Recherche Luxembourg (INTER/Mobility/19/13995684/DLAl/van der Torre)

References

Amgoud, L., and Cayrol, C. 2002. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Approximate Reasoning* 29(2):125–169.

Amgoud, L., and Vesic, S. 2014. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning* 55(2):585–606.

Arisaka, R.; Satoh, K.; and van der Torre, L. W. N. 2017. Anything you say may be used against you in a court of law - abstract agent argumentation (triple-a). In Pagallo, U.; Palmirani, M.; Casanovas, P.; Sartor, G.; and Villata, S., eds., *AI Approaches to the Complexity of Legal Systems*, volume 10791, 427–442. Springer.

Awad, E.; Booth, R.; Tohmé, F.; and Rahwan, I. 2017. Judgement aggregation in multi-agent argumentation. *J. Log. Comput.* 27(1):227–259.

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

Baroni, P.; Boella, G.; Cerutti, F.; Giacomin, M.; Van Der Torre, L.; and Villata, S. 2014. On the input/output behavior of argumentation frameworks. *Artificial Intelligence* 217:144–197.

Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds. 2018. *Handbook of Formal Argumentation*, volume 1. College Publications.

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

Baumann, R.; Brewka, G.; and Ulbricht, M. 2020. Comparing weak admissibility semantics to their dung-style counterparts–reduct, modularization, and strong equivalence in abstract argumentation. In *Proceedings of the In-*

ternational Conference on Principles of Knowledge Representation and Reasoning, volume 17, 79–88.

Caminada, M., and Pigozzi, G. 2011. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems* 22(1):64–102.

Caminada, M. 2017. Argumentation semantics as formal discussion. *Journal of Applied Logics* 4(8):2457–2492.

Coste-Marquis, S.; Devred, C.; Konieczny, S.; Lagasquie-Schiex, M.; and Marquis, P. 2007. On the merging of dung's argumentation systems. *Artif. Intell.* 171(10-15):730–753.

Delobelle, J.; Konieczny, S.; and Vesic, S. 2018. On the aggregation of argumentation frameworks: operators and postulates. *J. Log. Comput.* 28(7):1671–1699.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77:321–357.

Gabbay, D.; Giacomin, M.; Simari, G.; and Thimm, M., eds. to appear. *Handbook of Formal Argumentation*, volume 2. College Publications.

Gao, Y.; Toni, F.; Wang, H.; and Xu, F. 2016. Argumentation-based multi-agent decision making with privacy preserved. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 1153–1161.

Giacomin, M. 2017. Handling heterogeneous disagreements through abstract argumentation (extended abstract). In An, B.; Bazzan, A. L. C.; Leite, J.; Villata, S.; and van der Torre, L. W. N., eds., *PRIMA 2017: Principles and Practice of Multi-Agent Systems*, volume 10621 of *Lecture Notes in Computer Science*, 3–11. Springer.

Horty, J. F. 2001. Argument construction and reinstatement in logics for defeasible reasoning. *Artif. Intell. Law* 9(1):1–28.

Hunter, A.; Polberg, S.; and Thimm, M. 2020. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *Artif. Intell.* 281:103236.

Kaci, S.; van der Torre, L. W. N.; and Villata, S. 2018. Preference in abstract argumentation. In Modgil, S.; Budzynska, K.; and Lawrence, J., eds., *Computational Models of Argument - Proceedings of COMMA 2018, Warsaw, Poland, 12-14 September 2018*, volume 305 of *Frontiers in Artificial Intelligence and Applications*, 405–412. IOS Press.

Karanikolas, N.; Bisquert, P.; and Kaklamanis, C. 2019. A voting argumentation framework: Considering the reasoning behind preferences. In *ICAART 2019-11th International Conference on Agents and Artificial Intelligence*, volume 1, 42–53. SCITEPRESS-Science and Technology Publications.

Kontarinis, D., and Toni, F. 2015. Identifying malicious behavior in multi-party bipolar argumentation debates. In Rovatsos, M.; Vouros, G. A.; and Julián, V., eds., *Multi-Agent Systems and Agreement Technologies - 13th European Conference*, volume 9571 of *Lecture Notes in Computer Science*, 267–278. Springer.

Leite, J., and Martins, J. G. 2011. Social abstract argumentation. In Walsh, T., ed., *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, 2287–2292. IJCAI/AAAI.

Panisson, A. R.; Parsons, S.; McBurney, P.; Bordini, R. H.; et al. 2018. Choosing appropriate arguments from trustworthy sources. In *COMMA*, 345–352.

Prakken, H. 2018. Historical overview of formal argumentation. In *Handbook of Formal Argumentation*, 75–143. College Publications.

Qiao, L.; Yu, Y. S. L.; Liao, B.; and van der Torre, L. 2021. Arguing coalitions in abstract argumentation. *Logics for New-Generation AI* 93.

Rienstra, T.; Perotti, A.; Villata, S.; Gabbay, D. M.; and van der Torre, L. 2011. Multi-sorted argumentation. In *International Workshop on Theorie and Applications of Formal Argumentation*, 215–231. Springer.

Rienstra, T.; Sakama, C.; van der Torre, L.; and Liao, B. 2020. A principle-based robustness analysis of admissibility-based argumentation semantics. *Argument & Computation* 1–35.

van der Torre, L., and Vesic, S. 2017. The principle-based approach to abstract argumentation semantics. *If-CoLog Journal of Logics and Their Applications*.

Yu, L., and Van der Torre, L. 2020. A principle-based approach to bipolar argumentation. In *NMR 2020 Workshop Notes*, 227.