

# A Plausibility Semantics for Abstract Argumentation Frameworks

Emil Weydert

Individual and Collective Reasoning Group  
ILIAS-CSC, University of Luxembourg

## Abstract

We propose and investigate a simple plausibility-based extension semantics for abstract argumentation frameworks based on generic instantiations by default knowledge bases and the ranking construction paradigm for default reasoning.<sup>1</sup>

## 1 Prologue

The past decade has seen a flourishing of abstract argumentation theory, a coarse-grained high-level form of defeasible reasoning introduced by Dung [Dung 95]. It is characterized by a top-down perspective which ignores the logical fine structure of arguments and focuses instead on logical (conflict, support, ...) or extra-logical (preferences, ...) relations between given black box arguments so as to identify reasonable argumentative positions. One way to address the complexity of enriched argument structures carrying interacting relations, and to identify the best approaches for evaluating Dung's basic attack frameworks as well as more sophisticated argumentation systems, is to look for deeper unifying semantic foundations allowing us to improve, compare, and judge existing proposals, or to develop new ones.

A major issue is to what extent an abstract account can adequately model concrete argumentative reasoning in the context of a sufficiently expressive, preferably defeasible logic. The instantiation of abstract frameworks by more fine-grained logic-based argument constructions and configurations is therefore an important tool for justifying or criticising abstract argumentation theories. Most of this work is however based on the first generation of nonmonotonic formalisms, like Reiter's default logic or logic programming. While these are closer to classical logic and the original spirit of Dung's approach, it is well known that they fail to model plausible implication. In fact, they are haunted by counterintuitive behaviour and violate major desiderata for default reasoning encoded in benchmark examples and rationality postulates [Mak 94]. For instance, the only way to deal even with simple instances of specificity reasoning are opaque ad hoc prioritization mechanisms.

<sup>1</sup>This is an improved - polished and partly revised - version of my ECSQARU 2013 paper. It adds a link to structured argumentation, refines the semantic instantiation concept, and discusses attacks between inference pairs.

The goal of the present work is therefore to supplement existing instantiation efforts with a simple ranking-based semantic model which interprets arguments and attacks by conditional knowledge bases. The well-behaved ranking construction semantics for default reasoning [Wey 96, 98, 03] can then be exploited to specify a new extension semantics for Dung frameworks which allows us to directly evaluate the plausibility of argument collections. Its occasionally unorthodox behaviour may shed a new light on basic argumentation-theoretic assumptions and concepts.

We start with an introduction to default reasoning based on the ranking construction paradigm. After a short look at abstract argumentation theory, we show how to interpret abstract argumentation frameworks by instantiating the arguments and characterizing the attacks with suitable sets of conditionals describing constraints over ranking models. Based on the concept of generic instantiations, i.e. using minimal assumptions, and plausibility maximization, we then specify a natural ranking-based extension semantics. We conclude with a simple algorithm, some instructive examples, and the discussion of several important properties.

## 2 Ranking-based default reasoning

We assume a basic language  $L$  closed under the usual propositional connectives, together with a classical satisfaction relation  $\models$  inducing a monotonic entailment relation  $\vdash \subseteq 2^L \times L$ . The model sets of  $(L, \models)$  are denoted by  $\llbracket \varphi \rrbracket = \{m \mid m \models \varphi\}$ , resp.  $\llbracket \Sigma \rrbracket = \bigcap_{\varphi \in \Sigma} \llbracket \varphi \rrbracket$  for  $\Sigma \subseteq L$ .  $\mathcal{B}_L$  is the boolean proposition algebra over  $\mathbb{B}_L = \{\llbracket \varphi \rrbracket \mid \varphi \in L\}$ . Let  $Cn(\Sigma) = \{\psi \mid \Sigma \vdash \psi\}$ .

Default inference is an important instance of nonmonotonic reasoning concerned with drawing reasonable but potentially defeasible conclusions from knowledge bases of the form  $\Sigma \cup \Delta$ , where  $\Sigma \subseteq L$  is a set of assumptions or facts, e.g. encoding knowledge about a specific state of affairs in the domain language  $L$ , and  $\Delta \subseteq L(\rightarrow, \rightsquigarrow)$  is a collection of conditionals expressing strict or exception-tolerant implicational information over  $L$ , which is used to guide defeasible inference.  $L(\rightarrow, \rightsquigarrow) = \{\varphi \rightarrow \psi \mid \varphi, \psi \in L\} \cup \{\varphi \rightsquigarrow \psi \mid \varphi, \psi \in L\}$  is the corresponding flat conditional language on top of  $L$ . In the following we will focus on finite  $\Sigma$  and  $\Delta$ .  $\Delta^{\rightarrow} = \{\varphi \rightarrow \psi \mid \varphi \rightarrow \psi, \varphi \rightsquigarrow \psi\}$  collects the material implications corresponding to the conditionals in  $\Delta$ .

The strict implication  $\varphi \rightarrow \psi$  states that  $\varphi$  necessarily

implies  $\psi$ , forcing us to accept  $\psi$  given  $\varphi$ . The default implication  $\varphi \rightsquigarrow \psi$  tells us that  $\varphi$  plausibly/normally implies  $\psi$ , and only recommends the acceptance of  $\psi$  given  $\varphi$ . The actual impact of a default depends of course on the context  $\Sigma \cup \Delta$  and the chosen nonmonotonic inference concept  $\vdash$ , which will be discussed later.

We can distinguish two perspectives in default reasoning: the autoepistemic/context-based one, and the plausibilistic/quasi-probabilistic one. The former is exemplified by Reiter's default logic, where defaults are usually modeled by normal default rules of the form  $\varphi : \psi/\psi$  (if  $\varphi$ , and it is consistent that  $\psi$ , then  $\psi$ ). A characteristic feature is that the conclusions are obtained by intersecting suitable equilibrium sets, known as extensions.

The alternative is to use default conditionals interpreted by some preferential or valuational semantics, e.g. System Z [Pea 90, Leh 92], or probabilistic ME-based accounts [GMP 93] (ME = maximum-entropy). For historical reasons and technical convenience (closeness to classical logic), the first approach has received most attention, especially in the context of argumentation. However, this ignores the fact that the conditional semantic paradigm has a much better record when it comes to the natural handling of benchmark examples and the satisfaction of rationality postulates [Mak 94]. It therefore seems promising to investigate whether such semantic-based accounts can also help to instantiate and evaluate abstract argumentation frameworks.

Our default conditional semantics for interpreting argumentation frameworks is based on the simplest plausibility measure concept able to reasonably handle independence and conditionalization, namely Spohn's ranking functions [Spo 88, 12], or more generally, ranking measures [Wey 94]. These are quasi-probabilistic belief valuations expressing the degree of surprise or implausibility of propositions. Integer-valued ranking functions were originally introduced by Spohn to model the iterated revision of graded plain belief. We will consider  $[0, \infty]_{\text{real}}$ -valued ranking measures<sup>2</sup>, where  $\infty$  expresses doxastic impossibility.

### Definition 2.1 (Real-valued ranking measures)

A map  $R : \mathcal{B}_L \rightarrow ([0, \infty], 0, \infty, +, \leq)$  is called a *real-valued ranking measure* iff  $R(\llbracket \mathbf{T} \rrbracket) = 0$ ,  $R(\llbracket \mathbf{F} \rrbracket) = R(\emptyset) = \infty$ , and for all  $A, B \in \mathbb{B}_L$ ,  $R(A \cup B) = \min_{\leq} \{R(A), R(B)\}$ .  $R(\cdot|\cdot)$  is the associated conditional ranking measure defined by  $R(B|A) = R(A \cap B) - R(A)$  if  $R(A) \neq \infty$ , else  $R(B|A) = \infty$ .  $R_0$  is the uniform ranking measure, i.e.  $R_0(A) = 0$  for  $A \neq \emptyset$ . If  $\mathcal{B} = \mathcal{B}_L$ , we will use the abbreviation  $R(\varphi) := R(\llbracket \varphi \rrbracket)$ .

For instance, the order of magnitude reading interprets ranking measure values  $R(A)$  as exponents of infinitesimal probabilities  $P(A) = p_A \varepsilon^{R(A)}$ , which explains the parallels with probability theory. The monotonic semantics of our conditionals  $\rightarrow, \rightsquigarrow$  is based on the satisfaction relation  $\models_{rk}$ . The corresponding truth conditions are

- $R \models_{rk} \varphi \rightarrow \psi$  iff  $R(\varphi \wedge \neg\psi) = \infty$ .
- $R \models_{rk} \varphi \rightsquigarrow \psi$  iff  $R(\varphi \wedge \psi) + 1 \leq R(\varphi \wedge \neg\psi)$ .

<sup>2</sup>Although for us, rational values would actually be sufficient.

That is, we assume that a strict implication  $\varphi \rightarrow \psi$  states that  $\varphi \wedge \neg\psi$  is doxastically impossible.

Note that we may replace  $\varphi \rightarrow \psi$  by  $\varphi \wedge \neg\psi \rightsquigarrow \mathbf{F}$ , i.e. it would be actually enough to consider  $L(\rightsquigarrow)$ . We use  $\leq$  with a threshold because this provides more discriminatory power and also guarantees the existence of minima for relevant ranking construction procedures. The exchangeability of arbitrary  $r, r' \neq 0, \infty$  by automorphisms allows us to focus, by convention, on the threshold 1. For  $\Delta \cup \{\delta\} \subseteq L(\rightarrow, \rightsquigarrow)$ , we set

$$\llbracket \Delta \rrbracket_{rk} = \{R \mid R \models_{rk} \Delta\}, \Delta \vdash_{rk} \delta \text{ iff } \llbracket \Delta \rrbracket_{rk} \subseteq \llbracket \delta \rrbracket_{rk}.$$

$\vdash_{rk}$  is monotonic and verifies the axioms and rules of preferential conditional logic and disjunctive rationality (threshold semantics: no rational monotony) for  $\rightsquigarrow$  [KLM 90].

But it is important to understand that the central concept in default reasoning is not some monotonic conditional logic for  $L(\rightarrow, \rightsquigarrow)$ , but a nonmonotonic meta-level inference relation  $\vdash$  over  $L \cup L(\rightarrow, \rightsquigarrow)$  specifying which conclusions  $\psi \in L$  can be plausibly inferred from finite  $\Sigma \cup \Delta \subseteq L \cup L(\rightarrow, \rightsquigarrow)$ . We write  $\Sigma \cup \Delta \vdash \psi$ , or alternatively  $\Sigma \vdash_{\Delta} \psi$ , and set  $C_{\Delta}^{\vdash}(\Sigma) = \{\psi \mid \Sigma \vdash_{\Delta} \psi\}$ .

The ranking semantics for plausibilistic default reasoning is based on nonmonotonic ranking choice operators  $\mathcal{I}$  which map each finite  $\Delta \subseteq L(\rightarrow, \rightsquigarrow)$  to a collection  $\mathcal{I}(\Delta) \subseteq \llbracket \Delta \rrbracket_{rk}$  of preferred ranking models of  $\Delta$ . A corresponding ranking-based default inference notion  $\vdash^{\mathcal{I}}$  can then be specified by

$$\Sigma \vdash_{\Delta}^{\mathcal{I}} \psi \text{ iff for all } R \in \mathcal{I}(\Delta), R \models_{rk} \Delta \wedge \Sigma \rightsquigarrow \psi.$$

Similarly, we can also define a monotonic inference concept characterizing the strict consequences.

$$\Sigma \vdash_{\Delta}^{\mathcal{I}} \psi \text{ iff for all } R \in \mathcal{I}(\Delta), R \models_{rk} \Delta \wedge \Sigma \rightarrow \psi.$$

If  $\mathcal{I}(\Delta) = \llbracket \Delta \rrbracket_{rk}$ ,  $\vdash_{\Delta}^{\mathcal{I}}$  is, modulo cosmetic details, equivalent to preferential entailment (System P) [KLM 90]. If  $\leq_{pt}$  describes the pointwise comparison of ranking measures, i.e.  $R \leq_{pt} R'$  iff for all  $A \in \mathbb{B}_L$   $R(A) \leq R'(A)$ , then  $\mathcal{I}(\Delta) = \{Min_{\leq_{pt}} \llbracket \Delta \rrbracket_{rk}\}$  essentially characterizes System Z [Pea 90]. Because these approaches fail to adequately deal with inheritance to exceptional subclasses, we introduced and developed the construction paradigm for default reasoning [Wey 96, 98, 03], which is a powerful strategy for specifying reasonable  $\mathcal{I}$  based on Spohn's Jeffrey-conditionalization for ranking measures. The resulting default inference notions are well-behaved and show nice inheritance features. The essential idea is that defaults do not only specify ranking constraints, but also admissible construction steps to generate them. In particular, for each default  $\varphi \rightsquigarrow \psi$ , we are allowed to uniformly shift upwards (make less plausible/increase the ranks of) the  $\varphi \wedge \neg\psi$ -worlds, which amounts to strengthen belief in the corresponding material implication  $\varphi \rightarrow \psi$ . If  $W$  is finite, this is analogous to specifying the rank of a world by adding a weight  $\geq 0$  for each default it violates. More formally, we define a shifting transformation  $R \rightarrow R + r[\rho]$  such that for each ranking measure  $R, \chi, \rho \in L$ , and  $r \in [0, \infty]$ , we set

$$(R + r[\rho])(\chi) = \min\{R(\chi \wedge \rho) + r, R(\chi \wedge \neg\rho)\}.$$

This corresponds to uniformly shifting  $\rho$  by  $r$ .

### Definition 2.2 (Constructibility)

Let  $\Delta = \{\varphi_i \rightsquigarrow / \rightarrow \psi_i \mid i \leq n\} \subseteq L(\rightarrow, \rightsquigarrow)$ . A ranking measure  $R'$  is said to be constructible from  $R$  over  $\Delta$ , written  $R' \in \text{Constr}(\Delta, R)$ , iff there are  $r_i \in [0, \infty]$  s.t.  $R' = R + \sum_{i \leq n} r_i [\varphi_i \wedge \neg \psi_i]$ .<sup>3</sup>

For instance, we obtain a well-behaved robust default inference relation, System J [Wey 96], just by setting  $\mathcal{I}_J(\Delta) = \text{Constr}(\Delta, R_0) \cap \llbracket \Delta \rrbracket_{rk}$ . To implement shifting minimization, we may strengthen System J by allowing proper shifting ( $r_i > 0$ ) only if the targeted ranking constraint interpreting a default  $\varphi_i \rightsquigarrow \psi_i$  is realized as an equality constraint  $R(\varphi_i \wedge \psi_i) + 1 = R(\varphi_i \wedge \neg \psi_i)$ . Otherwise, the shifting wouldn't seem to be justified in the first place.

### Definition 2.3 (Justifiable constructibility)

$R$  is called a justifiably constructible model of  $\Delta$ , written  $R \in \mathcal{I}_{jj}(\Delta)$  iff  $R \models_{rk} \Delta$ ,  $R = R_0 + \sum_{i \leq n} r_i [\varphi_i \wedge \neg \psi_i]$ , and for each  $r_j > 0$ ,  $R(\varphi_j \wedge \psi_j) + 1 = R(\varphi_j \wedge \neg \psi_j)$ .

It follows from a standard property of entropy maximization (ME) that the order-of-magnitude translation of ME, in the context of a nonstandard model of the reals with infinitesimals [GMP 93, Wey 95], to the ranking level always produces a canonical justifiably constructible ranking model  $R_{me}$ . We set  $\mathcal{I}_{me}(\Delta) = \{R_{me}^\Delta\}$ . Hence, if  $\Delta \not\models_{rk} \mathbf{F}$ ,  $R_{me}^\Delta \in \mathcal{I}_{jj}(\Delta) \neq \emptyset$ . If  $\mathcal{I}_{jj}(\Delta)$  is a singleton, we have therefore  $\vdash^{jj} = \vdash^{me}$ . This holds for instance for minimal core default sets  $\Delta$  [GMP 93], where no doxastically possible  $\varphi_i \wedge \neg \psi_i$ , i.e.  $\Delta \not\models_{rk} \varphi_i \wedge \neg \psi_i \rightsquigarrow \mathbf{F}$ , is covered by other  $\varphi_j \wedge \neg \psi_j$ . However, because of its fine-grained quantitative character,  $\vdash^{me}$  is actually representation-dependent, i.e. the solution depends on how we describe a problem in  $L$ , it is not invariant under boolean automorphisms of  $\mathcal{B}_L$ . Fortunately, there are two other natural representation-independent ways to pick up a canonical justifiably constructible model.

- System JZ is based on a natural canonical hierarchical ranking construction in the tradition of System Z and ensures justifiable constructibility [Wey 98, 03]. It constitutes a uniform way to implement the minimal information philosophy at the ranking level.
- System JJR is based on the fusion of the justifiably constructible ranking models of  $\Delta$ , i.e.  $\mathcal{I}_{j jr}(\Delta) = \{R_{j jr}^\Delta\}$ , where for all  $A \in \mathbb{B}_L$ ,  $R_{j jr}^\Delta(A) = \text{Min}_{\leq p_t} \mathcal{I}_{jj}(A)$ .  $\vdash^{j jr}$  may be of particular interest because its canonical ranking model is at least as plausible as every justifiably constructible one.

Note that for non-canonical  $\mathcal{I}_{jj}(\Delta)$ , it is possible that  $R_{j jr}^\Delta \notin \mathcal{I}_{jj}(\Delta)$ . We have  $\vdash^{jj} \subset \vdash^{me}, \vdash^{jz}, \vdash^{j jr}$ . Fortunately, for the generic default sets we will use to interpret abstract argumentation frameworks, all four turn out to be equivalent. To conclude this section, let us consider a simple example with a single JJ-model.

### Big birds example:

Non-flying birds are not inferred to be small.

<sup>3</sup>Similar ideas can be found in [BSS 00, KI 01].

$$\{B, \neg F\} \cup \{B \rightsquigarrow S, B \rightsquigarrow F, \neg S \rightsquigarrow \neg F\} \not\vdash^{jj} S$$

The canonical JJ/ME/JZ/JJR-model is then

$R = R_0 + 1[\neg F] + 2[\neg S \wedge F]$ . But  $R \not\models_{rk} B \wedge \neg F \rightsquigarrow S$  because  $R(B \wedge \neg F \wedge S) = R(B \wedge \neg F \wedge \neg S) = 1$

## 3 Abstract argumentation

The idea of abstract argumentation theory, launched by Dung [Dun 95], has been to replace the traditional bottom-up strategy, which models and exploits the logical fine structure of arguments, by a top-down perspective, where arguments become black boxes evaluated only based on knowledge about specific logical or extra-logical relationships connecting them. It is interesting to see that such a coarse-grained relational analysis often seems sufficient to determine which collections of instantiated arguments are reasonable. In addition to possible conceptual and computational gains, the abstract viewpoint offers furthermore a powerful methodological tool for general argumentation-theoretic investigations.

An abstract argumentation framework in the original sense of Dung is a structure of the form  $\mathcal{A} = (\mathbb{A}, \triangleright)$ , where  $\mathbb{A}$  is a collection of abstract entities representing arguments, and  $\triangleright$  is a possibly asymmetric binary attack relation modeling conflicts between arguments. To grasp the expressive complexity of real-world argumentation, several authors have extended this basic account to include further inferential or cognitive relations, like support links, preferences, valuations, or collective attacks. Our general definition<sup>4</sup> [Wey 11] for the first-order context is as follows.

**Definition 3.1 (Hyperframeworks)** A general abstract argumentation framework, or hyperframework (HF), is just a structure of the form  $\mathcal{A} = (\mathbb{A}, (\mathcal{R}_i)_{i \in I}, (\mathcal{P}_j)_{j \in J})$ , where  $\mathbb{A}$  is the domain of arguments, the  $\mathcal{R}_i$  are conflictual, and the  $\mathcal{P}_i$  non-conflictual relations over  $\mathbb{A}$ .  $B \subseteq \mathbb{A}$  is said to be conflict-free iff it does not instantiate a conflictual relation.

For instance, standard Dung frameworks  $(\mathbb{A}, \triangleright)$  carry one conflictual and no non-conflictual relations. The general inferential task in abstract argumentation is to identify reasonable evaluations of the arguments described by  $\mathcal{A}$ , e.g. to find out which sets of arguments describe acceptable argumentative positions. These are called extensions. In Dung's scenario, the extensions are  $E \subseteq \mathbb{A}$  obeying suitable acceptability conditions in the context of  $\mathcal{A}$ , the minimal requirement being the absence of internal conflicts. For instance,  $E$  is admissible iff it is conflict-free and each attacker of an  $a \in E$  is attacked by some  $b \in E$ .  $E$  is grounded/preferred iff it is minimally/maximally admissible, it is stage iff  $E \cup \triangleright'' E$  is maximal, semi-stable if it is also admissible, and stable iff  $\mathbb{A} - E = \triangleright'' E$ . Here  $\triangleright'' E$  is the relational image of  $E$ , i.e. the set of  $a \in \mathbb{A}$  attacked by some  $b \in E$ . In concrete decision contexts, we may however also want to exploit finer-grained assessments of arguments, like prioritizations or classifications. This suggests a more general semantic perspective [Wey 11].

**Definition 3.2 (Hyperextensions)** A hyperframework semantics is a map  $\mathcal{E}$  associating with each hyperframework

<sup>4</sup>A bit of an overkill for this paper, but we couldn't resist.

$\mathcal{A} = (\mathbb{A}, (\mathcal{R}_i)_{i \in I}, (\mathcal{P}_j)_{j \in J})$  of a given signature a collection  $\mathcal{E}(\mathcal{A})$  of distinguished evaluation structures expanding  $\mathcal{A}$ , of the form  $(\mathcal{A}, In^{\mathcal{A}}, (\mathcal{F}_h)_{h \in H})$ .  $In^{\mathcal{A}}$  is here a conflict-free subset of  $\mathbb{A}$ . The elements of  $\mathcal{E}(\mathcal{A})$  are called hyperextensions of  $\mathcal{A}$ .

$In^{\mathcal{A}}$  plays here the role of a classical extension, whereas the  $\mathcal{F}_h$  ( $h \in H$ ) express more sophisticated structures over arguments, e.g. a posteriori plausibility orderings, value predicates, or completions of framework relations considered partial. If  $H = \emptyset$ , we are back to Dung.

## 4 Concretizing arguments

Ideally, abstract argumentation frameworks should be reconstructible as actual abstractions of logic-based argumentation scenarios. Such an anchoring seems required to develop, evaluate, and apply the abstract models in a suitable way. In a first step, this amounts to instantiate the abstract arguments from the framework domain by logical entities representing concrete arguments, and to interpret the abstract framework structure by specific inferential or evaluational relationships fitting the conceptual intentions the abstract level tries to capture. In what follows we will sketch a natural hierarchy of instantiation layers, passing from more concrete, deep instantiations, to more abstract, shallow ones, with a focus on the intermediate level.

### Structured instantiations:

We start with logic-based structured argumentation over a defeasible conditional logic  $\mathcal{L}_\delta = (L \cup L(\Rightarrow, \rightsquigarrow), \vdash^\delta, \vdash^\delta)$ , with  $(L, \vdash)$  as a classical Tarskian background logic. For the moment, we do not impose any further a priori conditions on  $\mathcal{L}_\delta$ . But eventually we will turn to specific ranking-based default formalisms. In the context of  $\mathcal{L}_\delta$ , a concrete defeasible argument  $a$  for a claim  $\psi_a \in L$ , exploiting some given general knowledge base  $\Sigma \cup \Delta$ , is modeled by a finite rooted defeasible inference tree  $\mathcal{T}_a$  whose nodes  $s$  are tagged by local claims  $\eta_s \in L \cup L(\Rightarrow, \rightsquigarrow)$  such that

- the root node is tagged by  $\psi_a$ ,
- the leaf nodes are tagged by  $\eta_s \in \Sigma_a \cup \Delta_a \cup \Lambda$ , where  $\Lambda = \{\mathbf{T}\} \cup \{\varphi \Rightarrow \varphi, \varphi \rightsquigarrow \varphi \mid \varphi \in L\}$  (basic tautologies),
- the non-leaf nodes are tagged by  $\eta_s \in L$  s.t.  $\Gamma_s \vdash^\delta \eta_s$  where  $\Gamma_s$  is the set of claims from the children of  $s$ .

$\Sigma_a \cup \Delta_a$  is the contingent premise set of  $a$ , the premises being the claims of the leaf nodes. Within concrete arguments, the local justification steps, e.g. from  $\Gamma_s$  to  $\eta_s$ , are typically assumed to be elementary, like instances of modus ponens. To handle reasoning by cases, which holds for plausible implication, we may also apply the *disjunctive modus ponens* for  $\Rightarrow$  and  $\rightsquigarrow$ , e.g.

$$\Gamma_s = \{\varphi_1 \vee \dots \vee \varphi_n, \varphi_1 \rightsquigarrow \psi_1, \dots, \varphi_n \rightsquigarrow \psi_n\}$$

$$\vdash^\delta \psi_1 \vee \dots \vee \psi_n (= \eta_s).$$

If  $\Gamma_s \subseteq L \cup L(\Rightarrow)$ , we can replace  $\vdash^\delta$  by  $\vdash$  and obtain a strict inference step. For our purposes we may ignore the exact nature of the justification steps. Note that the correctness of local inference steps does not entail the

global correctness of the argument  $a$ . Consider for instance  $\Sigma_a \cup \Delta_a = \{\varphi\} \cup \{\varphi \rightsquigarrow \psi, \psi \rightsquigarrow \neg\varphi\}$ , which is consistent w.r.t.  $\vdash^\delta = \vdash^{\mathcal{I}}$ .

$$\{\varphi\} \cup \{\varphi \rightsquigarrow \psi\} \vdash^\delta \psi \text{ and } \{\psi\} \cup \{\psi \rightsquigarrow \neg\varphi\} \vdash^\delta \neg\varphi,$$

but  $\Sigma_a \cup \Delta_a \not\vdash^\delta \neg\varphi$ . This example looks odd because accepting the whole argument would require the acceptance of all its claims, which is blocked by  $\varphi, \neg\varphi \vdash \mathbf{F}$ . In fact, a natural requirement for an acceptable argument  $a$  would be

$$\text{Material consistency: } \Sigma_a \cup \Delta_a \not\vdash \mathbf{F}.$$

This means that the factual premises and the material implications corresponding to the conditional premises are classically consistent. Note that this condition is strictly stronger than  $\Sigma_a \cup \Delta_a \not\vdash^\delta \mathbf{F}$  because we typically have  $\{\mathbf{T} \rightsquigarrow \varphi, \neg\varphi\} \not\vdash^\delta \mathbf{F}$  whereas  $\{\mathbf{T} \rightarrow \varphi, \neg\varphi\} \vdash \mathbf{F}$ . However, in practice, without omniscience w.r.t. propositional logic, it may not be clear whether these global conditions are actually satisfied. Real arguments may well be inconsistent in the strong sense.

In structured argumentation, an argument tree has two functions: first, to describe and offer a prima facie justification for a claim, and secondly, to specify target points where other arguments may attack. It is essentially a computational tool which is intended to help identifying - or even defining - inferential relationships within a suitable defeasible conditional logic  $\mathcal{L}_\delta$ , and to help specifying attack relations to determine reasonable argumentative positions.

But what can we say about the semantic content of an argument represented by such a tree? What is an agent committed to if he accepts or believes a given argument, or a whole collection of arguments? Which tree attributes have to be known to specify this content? What is the meaning of attacks between arguments?

### Conditional instantiations:

Our basic idea is that, whatever the requirements for argumentation trees in the context of  $\mathcal{L}_\delta$ , and whatever the content of an argument  $a$  represented by such a tree  $\mathcal{T}_a$ , it should only depend on the collection of local claims  $\{\eta_s \mid s \text{ node of } \mathcal{T}_a\}$ , and more specifically, on the choice of the main claim  $\psi_a$ , the premise claims  $\Sigma_a \cup \Delta_a$ , and the intermediate claims  $\Psi_a$ . In fact, because the acceptance of a structured argument includes the acceptance of all its subarguments, we have to consider the main claims of the subarguments as well. So we can assume that the content of  $\mathcal{T}_a$  is fixed by the triple  $(\Sigma_a \cup \Delta_a, \Psi_a, \psi_a)$ . An agent accepting  $a$  obviously has to be committed to all the elements of the base  $\Sigma_a \cup \Delta_a \cup \Psi_a \cup \{\psi_a\}$ .

To be fully acceptable w.r.t.  $\mathcal{L}_\delta$ , the structured argument also has to be globally correct in the sense that all its local claims are actually defeasibly entailed by  $\Sigma_a \cup \Delta_a$ . In particular,  $\Sigma_a \cup \Delta_a \vdash^\delta \psi$  for each  $\psi \in \Psi_a \cup \{\psi_a\}$ . This requirement should also hold for each subargument  $b$  of  $a$ . But note that, because of defeasibility, this does not exclude that the premises  $\Sigma_b \cup \Delta_b$  of a subargument  $b$  could implicitly infer the negation of a local claim  $\psi_x$  external to  $b$ , as long as this conflicting inference is eventually overridden by the full premise set  $\Sigma_a \cup \Delta_a$ . It follows that the strengthening of a subargument by choosing a stronger claim could

undermine global correctness. But if the intermediate claims are always inferred and therefore implicitly present, we may actually drop  $\Psi_a$  and just consider for each globally correct argument  $a$  the finite inference pair  $(\Sigma_a \cup \Delta_a, \psi_a)$ .

Given a pure Dung framework  $\mathcal{A} = (\mathbb{A}, \triangleright)$  and the defeasible conditional logic  $\mathcal{L}_\delta$ , a structured instantiation  $I_{str}$  of  $\mathcal{A}$  maps each  $a \in \mathbb{A}$  to a globally correct argument tree  $\mathcal{T}_a$  over  $\mathcal{L}_\delta$ . On the most general level, we do not want to impose a priori further restrictions beyond inferential correctness. In practice one may however well decide to focus on specific argument trees, e.g. those using specific justification steps. Each  $I_{str}(a)$  specifies a correct inference pair  $I_{log}(a) = (\Sigma_a \cup \Delta_a, \psi_a)$ , which we call a conditional logical instantiation of  $a$  over  $\mathcal{L}_\delta$ .  $I_{log}$  specifies the intended logical content of an argument on the syntactic level. Note that it depends on the tree concept whether we can obtain all the correct inference pairs.

In monotonic argumentation, the consistency and minimality of the premise sets are standard assumptions. But within defeasible argumentation, a more liberal perspective may be preferable. For instance, on the structured level, we want to allow arguments claiming **F**. The reductio ad absurdum principle then offers a possibility to attack arguments from within. Consequently, we also have to accept instantiating inference pairs whose conclusion is **F**. On the other hand, material consistency, the existence of models of  $\Sigma_a$  which do not violate any conditional in  $\Delta_a$ , is a natural requirement in the context of argumentation theory. But we can replace it by a qualified version, restricted to those instances where  $\Sigma_a \cup \Delta_a$  is actually consistent.

What about minimality? First, it may obviously fail for inference pairs obtained by flattening argument trees. Of course, we could consider an additional minimization step where we replace each  $(\Sigma_a \cup \Delta_a, \psi_a)$  by all those  $(\Phi, \psi_a)$  with  $\Phi \subseteq \Sigma_a \cup \Delta_a$  and which are minimal s.t.  $\Phi \vdash^\delta \psi_a$ . Although this may be computationally costly, it could be theoretically appealing. However, minimality could also be questioned because by adding premises, a conclusion may successively get accepted, rejected, and accepted again, letting the character of the inferential support change between different levels of specificity, which calls for a discrimination between the corresponding inference pairs. Proponents of minimality object that these types of support could, perhaps, also be reproduced by suitable minimal  $(\Phi, \psi_a)$ . However, this assumption is not sustainable for ranking-based semantics for argumentation, because here the results may change if we restrict ourselves to minimal premise sets. In fact, shrinking  $\Sigma_a \cup \Delta_a$  to  $\Phi$  may actually increase the set of possible attacks. In particular, we could have attacks on all the minimal  $\Phi \vdash^\delta \psi_a$ , but none on  $\Sigma_a \cup \Delta_a \vdash^\delta \psi_a$ . Hence premise minimality may fail.

### Shallow instantiations:

Let us recall our task: exploiting a ranking semantics for default reasoning to provide a plausibilistic semantics for abstract argumentation. But inference pairs, which populate the conditional logical instantiation level, are still rather complex and opaque objects. To model argumentation frameworks and their semantics, we would here have to deal

with sets of sets of conditionals, whose inferential interactions may furthermore be hard to assess. We therefore prefer to start with simpler entities and to seek more abstraction.

Consider the main goal of an agent: to extract from argument configurations suitable beliefs, expressed in the domain language  $L$ , whose plausibility is semantically modeled by ranking measures over  $\mathbb{B}_L$ . Given an inference pair  $(\Sigma_a \cup \Delta_a, \psi_a)$  representing the full conditional logical content of an argument  $a$ , in addition to the main claim  $\psi_a$ , there are three relevant collections of formulas:  $\Sigma_a, C_\vdash(\Delta_a \cup \Sigma_a), C_{\vdash\sim}(\Delta_a \cup \Sigma_a)$  which represent resp. the premises, the strict, and the defeasible consequences. If the language is finitary, this gives us four  $L$ -formulas representing the relevant propositional  $L$ -content.

- $\varphi_a = \wedge C_\vdash(\Sigma_a)$  (premise content).
- $\theta_a = \wedge C_\vdash(\Delta_a \cup \Sigma_a)$  (strict content).
- $\delta_a = \wedge C_{\vdash\sim}(\Delta_a \cup \Sigma_a)$  (defeasible content).
- $\psi_a$  (main claim).

We have  $\delta_a \vdash \theta_a \vdash \varphi_a$ , and  $\delta_a \vdash \psi_a$  by inferential correctness.  $\delta_a$  specifies the strongest possible claim based on the information made available by the argument. For our semantic modeling purposes, we will assume that  $\psi_a = \delta_a$ . If we abstract away from the representational details, we arrive at our central concept: the *shallow semantic instantiation* of  $a$  extracted from the conditional logical instantiation  $I_{log}(a)$ .

$$I_{sem}(a) = (\llbracket \varphi_a \rrbracket, \llbracket \theta_a \rrbracket, \llbracket \psi_a \rrbracket).$$

In the following, we will sloppily denote  $I_{sem}(a)$  by  $(\varphi_a, \theta_a, \psi_a)$ . One should emphasize that these propositional semantic profiles are not intended to grasp the full nature of arguments, but only to reflect certain characteristics exploitable by suitable argumentation semantics. We observe that each proposition triple  $(\varphi, \theta, \psi)$  with  $\psi \vdash \theta \vdash \varphi$  can become a shallow instantiation. In fact, if  $I_{log}(a) = (\{\varphi, \varphi \rightarrow \theta, \varphi \rightsquigarrow \psi\}, \psi)$ , for standard  $\vdash^\delta$ , we obtain  $I_{sem}(a) = (\varphi, \theta, \psi)$ . In terms of ranking constraints, this gives us  $R(\varphi \wedge \neg\theta) = \infty$  and  $R(\varphi \wedge \psi) + 1 \leq R(\varphi \wedge \neg\psi)$ .

## 5 Concretizing attacks

One argument attacks another argument if accepting the first interferes with the inferential structure or goal of the second one. To avoid a counterattack, the premises of the attacked argument should also not affect the inferential success of the attacker, otherwise the presupposition of the attack could be undermined. In the following we will investigate attack relations between conditional logical resp. shallow semantic instantiations of abstract arguments. We start with the former. Let  $I_{log}(a) = (\Sigma_a \cup \Delta_a, \psi_a)$  and  $I_{log}(b) = (\Sigma_b \cup \Delta_b, \psi_b)$  be two correct inference pairs for  $\mathcal{L}_\delta$ . We distinguish two scenarios: unilateral and mutual attack. The idea is to say that  $(\Sigma_a \cup \Delta_a, \psi_a)$  unilaterally attacks  $(\Sigma_b \cup \Delta_b, \psi_b)$  iff the premises of both arguments together with  $\psi_a$  enforce the strict rejection of  $\psi_b$ , i.e.

$$\Sigma_b \cup \Delta_b \cup \Sigma_a \cup \Delta_a \cup \{\psi_a\} \vdash^\delta \neg\psi_b,$$

whereas the defeasible inference of  $\psi_a$  from the premises is preserved, i.e.

$$\Sigma_a \cup \Delta_a \cup \Sigma_b \cup \Delta_b \vdash^\delta \psi_a.$$

On the other hand,  $(\Sigma_a \cup \Delta_a, \psi_a)$  and  $(\Sigma_b \cup \Delta_b, \psi_b)$  attack each other iff they strictly reject each other's claims, i.e.

$$\Sigma_b \cup \Delta_b \cup \Sigma_a \cup \Delta_a \cup \{\psi_a\} \vdash^\delta \neg\psi_b, \text{ and}$$

$$\Sigma_a \cup \Delta_a \cup \Sigma_b \cup \Delta_b \cup \{\psi_b\} \vdash^\delta \neg\psi_a.$$

This holds for instance if their premise sets, resp. their claims, are classically inconsistent. This definition provides one of the strongest possible natural attack relations for inference pairs. Note that we have a self-attack iff the premise set is inconsistent, i.e.  $\Sigma_a \cup \Delta_a \vdash^\delta \mathbf{F}$ . To exploit the powerful semantics of ranking-based default reasoning, in what follows we will assume that  $\vdash^\delta = \vdash^{\mathcal{I}}$ , where  $\mathcal{I}$  is a ranking choice function.

How can we exploit the above approach to define attacks between shallow instantiations, e.g.  $I_{sem}(a) = (\varphi_a, \theta_a, \psi_a)$  and  $I_{sem}(b) = (\varphi_b, \theta_b, \psi_b)$ ? The corresponding inference pairs are  $I_{log}(a) = (\{\varphi_a, \varphi_a \rightarrow \theta_a, \varphi_a \rightsquigarrow \psi_a\}, \psi_a)$  and  $I_{log}(b) = (\{\varphi_b, \varphi_b \rightarrow \theta_b, \varphi_b \rightsquigarrow \psi_b\}, \psi_b)$ . For an unilateral attack from  $I_{log}(a)$  on  $I_{log}(b)$ , we must have

$$\mathcal{I}(\Delta_a \cup \Delta_b) \models_{rk} \varphi_a \wedge \varphi_b \wedge \psi_a \rightarrow \neg\psi_b, \text{ and}$$

$$\mathcal{I}(\Delta_a \cup \Delta_b) \models_{rk} \varphi_a \wedge \varphi_b \rightsquigarrow \psi_a.$$

This is, for instance, automatically realized if  $\psi_a \vdash \neg\psi_b$ ,  $\varphi_a \vdash \varphi_b$ , and we have logical independence elsewhere. For a bilateral attack, we may just drop the condition  $\varphi_a \vdash \varphi_b$ . However, we do not have to presuppose that all the attacks result from the logical structure induced by the instantiation. In fact, in addition to the instantiation-intrinsic attack relationships, there could be further attack links derived from a separate conditional knowledge base reflecting other known attacks.

From a given Dung framework  $\mathcal{A} = (\mathbb{A}, \triangleright)$  and a shallow instantiation  $I = I_{sem}$ , if we adopt the ranking semantic perspective and the above attack philosophy, we can induce a collection of conditionals specifying ranking constraints. For any  $a \in \mathbb{A}$ , the shallow inference pair supplies  $\varphi_a \rightarrow \theta_a$  (alternatively,  $\varphi_a \wedge \neg\theta_a \rightsquigarrow \mathbf{F}$ ) and  $\varphi_a \rightsquigarrow \psi_a$ . For every attack  $a \triangleright b$ , we get at least  $\psi_a \wedge \psi_b \rightsquigarrow \mathbf{F}$ . Note that this is a consequence of choosing maximal claims at the instantiation level. For each unilateral attack  $a \triangleright b$  we must add  $\varphi_a \wedge \varphi_b \rightsquigarrow \psi_a$  to preserve the inferential impact of  $a$  in the context of  $b$ . The resulting default base is

$$\begin{aligned} \Delta^{\mathcal{A}, I} = & \{\varphi_a \rightsquigarrow \psi_a, \varphi_a \rightarrow \theta_a \mid a \in \mathbb{A}\} \\ & \cup \{\psi_a \wedge \psi_b \rightsquigarrow \mathbf{F} \mid a \triangleright b \text{ or } b \triangleright a\} \\ & \cup \{\varphi_a \wedge \varphi_b \rightsquigarrow \psi_a \mid a \triangleright b, b \not\triangleright a\}. \end{aligned}$$

We observe that for each 1-loop, we get  $\psi_a \rightsquigarrow \mathbf{F}$  and  $\varphi_a \rightsquigarrow \psi_a$ , hence  $\Delta^{\mathcal{A}, I} \vdash^{\mathcal{I}} \neg\varphi_a$ . The doxastic impossibility of  $\varphi_a$  illustrates the paradoxical character of self-attacking arguments. The belief states compatible with an instantiated framework  $\mathcal{A}, I$  are here represented by the ranking models of  $\Delta^{\mathcal{A}, I}$ .

Conversely, we can identify for each instantiation  $I$  of  $\mathbb{A}$  and each collection of ranking measures  $\mathcal{R} \models_{rk} \Delta^{\mathcal{A}, I} = \{\varphi_a \rightsquigarrow \psi_a, \varphi_a \rightarrow \theta_a \mid a \in \mathbb{A}\}$  the attacks supported by all the  $R \in \mathcal{R}$ . Let  $\triangleright_I^{\mathcal{R}}$  be the resulting attack relation, that is, for each  $a, b \in \mathbb{A}$

$a \triangleright_I^{\mathcal{R}} b$  iff for all  $R \in \mathcal{R}$ ,  $R \models_{rk} \psi_a \wedge \psi_b \rightsquigarrow \mathbf{F}$  and

$$(R \models_{rk} \varphi_a \wedge \varphi_b \rightsquigarrow \psi_a \text{ or } R \not\models_{rk} \varphi_a \wedge \varphi_b \rightsquigarrow \psi_b).$$

The second disjunct is the result of an easy simplification. If  $a$  or  $b$  are self-reflective, we have  $a \triangleright_I^{\mathcal{R}} b$  because conditionals always hold if the premises are doxastically impossible. Because in this paper we will mainly consider canonical ranking choice functions, we are going to focus on  $\mathcal{R} = \{R\}$ , setting  $\triangleright_I^{\mathcal{R}} = \triangleright_I^R$ .

### Definition 5.1 (Ranking instantiation models)

Let the notation be as usual and  $\mathbb{A}^+ = \{a \in \mathbb{A} \mid a \not\triangleright a\}$ .  $(R, I)$  is called a ranking instantiation model (more sloppily, a ranking model) of  $\mathcal{A}$  iff

$$R \models_{rk} \Delta^{\mathbb{A}, I} = \{\varphi_a \rightsquigarrow \psi_a, \varphi_a \rightarrow \theta_a \mid a \in \mathbb{A}\},$$

and for all  $a, b \in \mathbb{A}^+$ ,  $a \triangleright b$  iff  $a \triangleright_I^R b$ . Let  $\mathcal{R}^{\mathcal{A}}$  be the collection of all the ranking instantiation models of  $\mathcal{A}$ .

That is, over the non-loopy arguments, the semantic-based attack relation  $\triangleright_I^R$  specified by  $R, I$  has to correspond exactly to the abstract attack relation  $\triangleright$ . The collection of ranking instantiation models is not meant to change if we add or drop attack links between self-reflective and other arguments because the details are absorbed by the impossible joint contexts. If  $\mathcal{A}$  and  $\mathcal{A}'$  share the same 1-loops and the same attack structure over the other arguments,  $\mathcal{R}^{\mathcal{A}} = \mathcal{R}^{\mathcal{A}'}$ .

It also important to observe, and we will come back to this, that each  $\mathcal{A} = (\mathbb{A}, \triangleright)$  admits many ranking instantiation models  $(R, I)$ , obtained by varying the ranking values or the proposition triples associated with the abstract arguments.

What can we say about classical types of attack? If we focus on the actual semantic content, rebuttal is characterized by incompatible defeasible consequents, and undermining by a defeasible consequent conflicting with an antecedent. In the ranking context, these two types of attacks can be modeled by constraints expressing necessities. The ranking characterizations are as follows. Recall that  $\psi_a \vdash \varphi_a$ ,  $\psi_b \vdash \varphi_b$ .

$a$  **rebuts**  $b$ :  $R(\psi_a \wedge \psi_b) = \infty$ , e.g. if  $\psi_a \vdash \neg\psi_b$ .

$a$  **undermines**  $b$ :  $R(\psi_a \wedge \varphi_b) = \infty$ , e.g. if  $\psi_a \vdash \neg\varphi_b$ .

In our simple semantic reading, undermining entails rebuttal because  $\psi_b \vdash \varphi_b$ . There are four qualitative attack configurations involving two arguments:  $\varphi_a \wedge \varphi_b$  being compatible with neither, one, or both of  $\psi_a, \psi_b$ . If  $a$  asymmetrically undermines  $b$ , we have  $R(\psi_a \wedge \varphi_b) = \infty$  and  $R(\psi_b \wedge \varphi_a) \neq \infty$ , hence  $R(\varphi_a \wedge \varphi_b) \neq \infty$ . This implies  $R \models_{rk} \varphi_a \wedge \varphi_b \wedge \psi_b \rightarrow \neg\psi_a$  and  $R \not\models_{rk} \varphi_a \wedge \varphi_b \rightsquigarrow \psi_a$ , i.e.  $b \triangleright_I^R a$  and  $a \not\triangleright_I^R b$  according to our attack semantics. It follows that undermining has no obvious ranking semantic justification if we stipulate that the defeasible claim entails the antecedent. Also note that rebuttal is entailed by symmetric and asymmetric attacks.

## 6 Ranking extensions

Ranking instantiation models offer new possibilities to identify reasonable argumentative positions. Let  $(R, I)$  be a model of the framework  $\mathcal{A} = (\mathbb{A}, \triangleright)$ . In the context of  $(R, I)$ , a minimal requirement for acceptable argument sets

$S \subseteq \mathbb{A}$  are coherent premises, i.e. the doxastic possibility of the joint strict contents  $\theta_S = \bigwedge_{a \in S} (\varphi_a \wedge \theta_a)$  w.r.t.  $R$ , which means  $R(\theta_S) \neq \infty$ . This excludes self-attacks, but not conflicts within  $S$ .  $S = \emptyset$  is by definition coherent because  $\theta_\emptyset = \mathbf{T}$ . Given that evidence  $\varphi_a$  should not be rejected without good reasons, the maximally coherent  $S \subseteq \mathbb{A}$  are of particular interest and constitute suitable background contexts when looking for extensions. Each  $E \subseteq S$  then specifies a proposition

$$\psi_{S,E} := \theta_S \wedge \bigwedge_{a \in E} \psi_a \wedge \bigwedge_{a \in \mathbb{A} - E} \neg \psi_a.$$

$\psi_{S,E}$  characterizes those worlds verifying the strict content of the  $a \in S$  and exactly the defeasible content of the  $a \in E$ . Because  $a \triangleright_I^R b$  implies  $R(\psi_a \wedge \psi_b) = \infty$ , any conflict  $a \triangleright b$  in  $E$  makes  $\psi_{S,E}$  impossible. Note however that  $R(\psi_{S,E}) = \infty$  may also result from non-binary conflicts involving multiple arguments, or a specific choice of logically dependent  $\varphi_a, \psi_a$ . What are the most reasonable extension candidates  $E \subseteq S \subseteq \mathbb{A}$  according to  $(R, I)$ ? One idea is to focus on those  $E$  which induce the most plausible  $\psi_{S,E}$  relative to  $\theta_S$  among all their maximal coherent supersets  $S$ .

**Definition 6.1 (Ranking extensions)** *Let  $(R, I)$  be a ranking instantiation model of  $\mathcal{A} = (\mathbb{A}, \triangleright)$ .  $E \subseteq \mathbb{A}$  is called a ranking-extension of  $\mathcal{A}$  w.r.t.  $(R, I)$  iff there is a maximal coherent  $S \subseteq \mathbb{A}$  with  $E \subseteq S$  and  $R(\psi_{S,E} | \theta_S) = 0$ .*

Observe that if  $S = \emptyset$  is the maximally coherent subset of  $\mathbb{A}$ , then  $R(\varphi_a) = \infty$  for each  $a \in \mathbb{A}$  and  $E = \emptyset$  is the only ranking extension. While the above definition looks rather decent, a cause of concern may be the great diversity of ranking models  $(R, I)$  available for any given  $\mathcal{A}$ . Consider for instance  $\mathcal{A} = (\{p, q, r\}, \{(p, q), (q, r)\})$ , i.e.  $p \triangleright q \triangleright r$ .  $\mathcal{A}$  together with a shallow instantiation  $I$  then induces ranking constraints described by the conditionals in

$$\begin{aligned} \Delta^{\mathcal{A}, I} = \{ & \psi_p \wedge \psi_q \rightsquigarrow \mathbf{F}, \psi_q \wedge \psi_r \rightsquigarrow \mathbf{F}, \varphi_p \wedge \varphi_q \rightsquigarrow \psi_p, \\ & \varphi_q \wedge \varphi_r \rightsquigarrow \psi_q, \varphi_p \rightsquigarrow \psi_p, \varphi_q \rightsquigarrow \psi_q, \varphi_r \rightsquigarrow \psi_r \}. \end{aligned}$$

If we assume that each set  $\{\varphi_x, \psi_x\}$  is logically independent from all the other  $\{\varphi_y, \psi_y\}$ , then  $\Delta^{\mathcal{A}, I}$  admits a unique justifiably constructible model, which therefore automatically must be the JZ- and JJR-model:  $R_{jz}^{\mathcal{A}, I}$ . In this example it is obtained by minimally shifting the violation areas of the conditionals.

$$R_{jz}^{\mathcal{A}, I} = R_0 + \infty[\psi_p \wedge \psi_q] + \infty[\psi_q \wedge \psi_r] + 1[\varphi_p \wedge \varphi_q \wedge \neg \psi_p] + 1[\varphi_q \wedge \varphi_r \wedge \neg \psi_q] + 1[\varphi_p \wedge \neg \psi_p] + 1[\varphi_q \wedge \neg \psi_q] + 1[\varphi_r \wedge \neg \psi_r].$$

Given that  $S = \mathbb{A}$  is coherent, there are eight extension candidates. For the doxastically possible alternatives,  $R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \{p, r\}}) = 2 < 3 = R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \{p\}}) = R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \{q\}}) < 4 = R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \{r\}}) < 5 = R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \emptyset}) < \infty$ . Because  $R_{jz}^{\mathcal{A}, I}(\varphi_S) = 2$ , we get  $R_{jz}^{\mathcal{A}, I}(\psi_{\mathbb{A}, \{p, r\}} | \varphi_S) = 0$ . The unique ranking extension is therefore  $\{p, r\}$ , which is also the standard Dung solution.

However, without any further constraints on the extension generating ranking instantiation model  $(R, I)$ , we could pick up as an alternative  $R = R_{jz}^{\mathcal{A}, I} + \infty[\psi_p \wedge \psi_r \wedge \varphi_q]$  such that  $R(\psi_p \wedge \psi_r \wedge \varphi_q) = \infty$ , resp. an  $I$  enforcing  $\psi_p \wedge \psi_r \wedge \varphi_q \vdash F$ . In both scenarios, the minima would

then become  $R(\psi_{\mathbb{A}, \{p\}}) = R(\psi_{\mathbb{A}, \{q\}}) = 3$ , imposing the ranking extensions  $\{p\}, \{q\}$ . Because of  $R(\psi_{\mathbb{A}, \{p, r\}}) = \infty$ , the standard extension  $\{p, r\}$  would necessarily be rejected. But this violates a hallmark of argumentation, namely the unconditional support of unattacked arguments, like  $p$ . This shows that we have to control the choice of ranking instantiation models to implement a reasonable ranking extension semantics.

The idea is now to choose on one hand, as our doxastic background, a well-justified canonical ranking measure model of the default base  $\Delta^{\mathcal{A}, I}$ , e.g. the JZ-model  $R_{jz}^{\mathcal{A}, I}$ , and on the other hand, implementing Ockham's razor, the simplest instantiations of the given framework  $\mathcal{A}$ . In particular, we stipulate that the instantiations of individual arguments should by default be logically independent. We emphasize that the goal here is just to interpret abstract argumentation frameworks with a minimal amount of additional assumptions, not to adequately model specific real-world arguments.

We can satisfy these desiderata by using disjoint vocabularies for instantiating different abstract arguments, and by relying on elementary instances of the defeasible modus ponens for the corresponding inference pairs. That is, we introduce for each  $a \in \mathbb{A}$  independent propositional atoms  $X_a, Y_a$ , and set  $I_{log}(a) = (\{X_a\} \cup \{X_a \rightsquigarrow Y_a\}, Y_a)$ . The corresponding shallow semantic instantiation is then  $I(a) = I_{sem}(a) = (\varphi_a, \varphi_a, \psi_a) = (X_a, X_a, X_a \wedge Y_a)$ . We call  $I$  a *generic instantiation*. Up to boolean isomorphism, it is completely characterized by the cardinality of  $\mathbb{A}$ .

If we fix a generic instantiation  $I$ , then the unique justifiably constructible ranking model of  $\Delta^{\mathcal{A}, I}$  is  $(a \triangleleft / \triangleright b : a \triangleright b \text{ or } b \triangleright a)$

$$\begin{aligned} R_{jz}^{\mathcal{A}} &= R_0 + \sum_{a \not\triangleright a} 1[\varphi_a \wedge \neg \psi_a] + \sum_{a \triangleright a} \infty[\varphi_a \wedge \neg \psi_a] + \\ & \quad \sum_{a \triangleright b} 1[\varphi_a \wedge \varphi_b \wedge \neg \psi_a] + \sum_{a \triangleleft \triangleright b} \infty[\psi_a \wedge \psi_b] \\ &= R_0 + \sum_{a \not\triangleright a} 1[X_a \wedge \neg Y_a] + \sum_{a \triangleright a} \infty[X_a \wedge \neg Y_a] + \\ & \quad \sum_{a \triangleright b} 1[X_a \wedge X_b \wedge \neg Y_a] + \sum_{a \triangleleft \triangleright b} \infty[X_a \wedge Y_a \wedge X_b \wedge Y_b]. \end{aligned}$$

Because the  $\{X_a, Y_a\}$  are logically independent for distinct  $a$ , and the defaults expressing an attack  $a \triangleright b$  just concern  $X_a \wedge X_b$ , only those  $X_a$  with  $a \triangleright a$  become impossible. In fact,  $\{\varphi_a \rightsquigarrow \psi_a, \psi_a \wedge \psi_a \rightsquigarrow \mathbf{F}\} \vdash_{rk} \varphi_a \rightsquigarrow \mathbf{F}$ . Hence, in line with intuition, the ranking instantiation model  $(R_{jz}^{\mathcal{A}}, I)$  will trivialize exactly the self-defeating arguments. Assuming genericity,  $\mathbb{A}^+ = \{a \in \mathbb{A} \mid a \not\triangleright a\}$  is then the unique maximal coherent subset of  $\mathbb{A}$ . We are now ready to specify our ranking-based evaluation semantics. Note that all the generic  $I$  are essentially equivalent.

### JZ-evaluation semantics (JZ-extensions):

$\mathcal{E}_{jz}(\mathcal{A}) = \{E \subseteq \mathbb{A} \mid E \text{ ranking extension w.r.t. } (R_{jz}^{\mathcal{A}, I}, I) \text{ for any/all generic } I\}$ .

There is actually a simple algorithm to identify the JZ-extensions using extension weights.

**Definition 6.2 (Extension weight)** *For each argumentation framework  $\mathcal{A} = (\mathbb{A}, \triangleright)$ , the extension weight function  $r_{\mathcal{A}} : 2^{\mathbb{A}} \rightarrow [0, \infty]$  is defined as follows: If  $E$  is conflict-free,*

$r_{\mathcal{A}}(E) = |\mathbb{A}^+ - E| + |\{a \in \mathbb{A}^+ - E \mid \exists b \in \mathbb{A}^+(a \triangleright b \wedge b \not\triangleright a)\}|$ , if not,  $r_{\mathcal{A}}(E) = \infty$ .

It is not too difficult to see that  $r_{\mathcal{A}}(E) = R_{jz}^{\mathcal{A},I}(\psi_{\mathbb{A}^+,E})$ . Hence,  $E \in \mathcal{E}_{jz}(\mathcal{A})$  iff  $r_{\mathcal{A}}(E) = \min\{r_{\mathcal{A}}(X) \mid X \subseteq \mathbb{A}\}$ . That is, the JZ-extensions are those where the sum of the number of non-reflective non-extension arguments and the number of one-sided attacks starting from them is minimal.

## 7 Examples and properties

To get a better understanding of the ranking extension semantics and its relation with other extension concepts, let us first take a look at how it handles some basic examples. Because of its uncommon semantic perspective and its partly quantitative character, we will observe some unorthodox behaviour. Under instantiation genericity, it is enough to compare  $R^{\mathcal{A},I}(\psi_{\mathbb{A}^+,E})$  for  $E \subseteq \mathbb{A}^+$ , or to focus on 1-loop-free frameworks. For each instance, we specify the domain  $\mathbb{A}$  and the full attack relation  $\triangleright$ .  $\psi_{\mathbb{A}^+, \{x_1 \dots x_n\}}$  is abbreviated by  $\psi_{x_1, \dots, x_n}$  resp.  $\psi_{\emptyset}$ .

**Simple reinstatement:**  $\{a, b, c\}$  with  $a \triangleright b \triangleright c$ .

The grounded extension  $\{a, c\}$  is the canonical result put forward by any standard acceptability semantics. The unique JJ-model, i.e. the JZ-model  $R$  of  $\Delta^{\mathcal{A},I}$ , satisfies  $R(\psi_a) = R(\psi_b) = 3, R(\psi_c) = 4, R(\psi_{a,c}) = 2$ , and  $R(\psi_{\emptyset}) = 5$ . The other candidates all get rank  $\infty$ . Because  $R(\psi_{a,c})$  is minimal,  $\{a, c\}$  is the only JZ-ranking extension, i.e.  $\mathcal{E}_{jz}(\mathcal{A}) = \{\{a, c\}\}$ .

**3-loop:**  $\{a, b, c\}$  with  $a \triangleright b \triangleright c \triangleright a$ .

Semantics under the admissibility dogm reject  $\{a\}, \{b\}, \{c\}$ , only  $\emptyset$  is admissible. But the JZ-model  $R$  verifies  $R(\psi_a) = R(\psi_b) = R(\psi_c) = 4 < 5 = R(\psi_{\emptyset})$ . Because all the alternatives are set to  $\infty$ , our ranking extensions are the maximal conflict-free sets  $\{a\}, \{b\}, \{c\}$ , i.e.,  $\mathcal{E}_{jz}$  clearly violates admissibility.

**Attack on 2-loop:**  $\{a, b, c\}$  with  $a \triangleright b \triangleright c \triangleright b$ .

We have  $R(\psi_{\emptyset}) = 4, R(\psi_a) = 2, R(\psi_b) = R(\psi_c) = 3, R(\psi_{a,c}) = 1$ , but  $\infty$  for the others. Here  $\mathcal{E}_{jz}(\mathcal{A}) = \{\{a, c\}\}$  picks up the canonical stable extension.

**Attack from 2-loop:**  $\{a, b, c\}$  with  $b \triangleright a \triangleright b \triangleright c$ .

We get  $R(\psi_{\emptyset}) = 4, R(\psi_a) = 3, R(\psi_b) = 2, R(\psi_c) = 3, R(\psi_{a,b}) = R(\psi_{b,c}) = \infty$ , and  $R(\psi_{a,c}) = 2$ .  $\mathcal{E}_{jz}(\mathcal{A}) = \{\{b\}, \{a, c\}\}$  thus collects the stable extensions.

**3,1-loop:**  $\{a, b, c\}$  with  $a \triangleright a \triangleright b \triangleright c \triangleright a$ .

$E = \emptyset$  is here the only admissible extension. The maximal coherent set is  $\mathbb{A}^+ = \{b, c\}$ , and we get  $R(\psi_b) = 1, R(\psi_c) = 2$ , as well as  $R(\psi_{\emptyset}) = 3$ . It follows that  $\mathcal{E}_{jz}(\mathcal{A}) = \{\{b\}\}$ , rejecting the stage extension  $\{c\}$ .

**3,2-loop:**  $\{a, b, c\}$  with  $b \triangleright a \triangleright b \triangleright c \triangleright a$ .

We have  $R(\psi_{\emptyset}) = 5, R(\psi_a) = 4, R(\psi_b) = 3$ , and

$R(\psi_c) = 3$ , i.e.  $\mathcal{E}_{jz}(\mathcal{A}) = \{\{b\}, \{c\}\}$ . But the stable extension  $\{b\}$  is the only admissible ranking extension.

The previous examples show that the ranking extension semantics  $\mathcal{E}_{jz}$  can diverge considerably from all the other major proposals found in the literature. It may look as if the main difference is its more liberal attitude towards some non-admissible, but still justifiable extensions. However, the semantics has an even more exotic flavour. Consider the following scenarios, where we indicate the minimal extension weights  $r_{\mathcal{A}}(E)$ .

**2-loop chain:**  $\{a, b, c\}, b \triangleright a \triangleright b \triangleright c \triangleright b$  :  
 $r(\{a, c\}) = 1 < 2 = r(\{b\})$ .

**Splitted 3-chain:**  $\{a, b, c, d\}, a \triangleright b \triangleright c, a \triangleright d \triangleright c$  :  
 $r(\{a, c\}) = r(\{b, d\}) = 4$ .

**Spoon:**  $\{a, b, c, d\}, a \triangleright b \triangleright c \triangleright d \triangleright c$  :  
 $r(\{a, d\}) = r(\{a, c\}) = r(\{b, d\}) = 3$ .

The first example documents the rejection of a stable extension, namely  $\{b\}$ . The second one illustrates the impact of quantitative considerations when dealing with a splitted variant of simple reinstatement. The third instance shows the coexistence of two stable extension with a non-admissible one. That is, even attack-free  $a$  can be questioned under certain circumstances. It follows that the above ranking semantic interpretation of argumentation frameworks deviates considerably from standard accounts and expectations. Let us now investigate how  $\mathcal{E}_{jz}$  handles some common principles for extension semantics.

**Isomorphy:**  $f : \mathcal{A} \cong \mathcal{A}'$  implies  $f'' : \mathcal{E}(\mathcal{A}') \cong \mathcal{E}(\mathcal{A})$ .

**Conflict-freedom:** If  $a, b \in E \in \mathcal{E}(\mathcal{A})$ , then  $a \not\triangleright b$ .

**CF-maximality:** If  $E \in \mathcal{E}(\mathcal{A})$ , then  $E$  is a maximal conflict-free subset of  $\mathbb{A}$ .

**Inclusion-maximality:** If  $E, E' \in \mathcal{E}(\mathcal{A})$  and  $E \subseteq E'$ , then  $E = E'$ .

**Reinstatement:** If  $E \in \mathcal{E}(\mathcal{A})$ ,  $a \in \mathbb{A}$ , and for each  $b \triangleright a$  there is an  $a' \in E$  with  $a' \triangleright b$ , then  $a \in E$ .

**Directionality:** Let  $\mathcal{A}_1 = (\mathbb{A}_1, \triangleright_1), \mathcal{A}_2 = (\mathbb{A}_2, \triangleright_2)$  be such that  $\mathbb{A}_1 \cap \mathbb{A}_2 = \emptyset, \triangleright_0 \subseteq \mathcal{A}_1 \times \mathcal{A}_2, \mathcal{A} = (\mathbb{A}_1 \cup \mathbb{A}_2, \triangleright_1 \cup \triangleright_0 \cup \triangleright_2)$ . Then we have  $\mathcal{E}(\mathcal{A}_1) = \{E \cap \mathbb{A}_1 \mid E \in \mathcal{E}(\mathcal{A})\}$ .

### Theorem 7.1 (Basic properties)

$\mathcal{E}_{jz}$  verifies isomorphy, conflict-freedom, inclusion maximality, and CF-maximality. It falsifies reinstatement and directionality.

The first four features are easy consequences of the  $\mathcal{E}_{jz}$ -specification. The violation of reinstatement directly follows from how the semantics handles 3-loops. The spoon example documents the failure of directionality if we set  $\mathbb{A}_1 = \{a, b\}$ . But directionality also fails for other prominent approaches, like the semi-stable semantics. Note however that it can be indirectly enforced by using  $\mathcal{E}_{jz}$  as the base function for an SCC-recursive semantics [BGG 05].



The following properties are inspired by the cumulativity principle for nonmonotonic reasoning. They state that if we drop an argument rejected by every extension, then this shouldn't add or erase skeptically supported arguments.

**Rejection cumulativity:**  $(\mathcal{A}|B: \mathcal{A}$  restricted to  $B)$

**Rej-Cut:** If  $a \notin \cup \mathcal{E}(\mathcal{A})$ , then  $\cap \mathcal{E}(\mathcal{A}|\mathbb{A} - \{a\}) \subseteq \cap \mathcal{E}(\mathcal{A})$ .

**Rej-CM:** If  $a \notin \cup \mathcal{E}(\mathcal{A})$ , then  $\cap \mathcal{E}(\mathcal{A}) \subseteq \cap \mathcal{E}(\mathcal{A}|\mathbb{A} - \{a\})$ .

Although our semantics relies on default inference notions verifying cumulativity at the level of  $\vdash_{\Delta}^{\mathcal{I}}$ , it nevertheless fails to validate the previous postulates.

**Theorem 7.2 (No rejection cumulativity)**

$\mathcal{E}_{jz}$  violates Rej-Cut and Rej-CM.

The counterexample for Rej-CUT is provided by  $b \triangleright c \triangleright a \triangleright b \triangleright a$ , because  $\{b\} \not\subseteq \{b\} \cap \{c\}$ . The one for Rej-CM is obtained by adding  $c \triangleright b$ . Here  $\{c\} \not\subseteq \{b\} \cap \{c\}$ .

Another idea for combining plausibilistic default reasoning and argumentation theory has been presented in [KIS 11]. It combines defeasible logic programming with a prioritization criterion based on System Z. While it handles some benchmarks better than the individual systems do, its heterogeneous character makes it hard to assess. It doesn't share our goal to seek a plausibilistic semantics for abstract argumentation and seems to produce different results even in the generic context.

## 8 Conclusions

We have shown how the ranking construction paradigm for default reasoning can be exploited to interpret abstract argumentation frameworks and to specify corresponding extension semantics by using generic argument instantiations and distinguished canonical ranking models. We have considered structured and conditional logical instantiations, defined attack between inference pairs, and after a further abstraction step, introduced simple semantic instantiations, which interpret arguments by triples of premise, strict, and defeasible content. While our basic ranking extension semantics  $\mathcal{E}_{jz}$  is intuitively appealing and has some interesting properties, it also exhibits a surprisingly unorthodox behaviour. This needs further exploration to see whether there are approaches which share the same semantic spirit but can avoid abnormalities conflicting with the standard argumentation philosophy. Actually, we have been able to develop an alternative semantics which seems to meet these demands, but it will have to be discussed elsewhere.

## 9 Bibliography

- BGG 05** P. Baroni, M. Giacomin, G. Guida. SCC-recursiveness: a general schema for argumentation semantics. *AIJ* 168:163-210, 2005.
- BSS 00** S. Benferhat, A. Saffiotti, P. Smets. Belief functions and default reasoning. *Artificial Intelligence* 122(1-2): 1-69, 2000.
- GMP 93** M. Goldszmidt, P. Morris, J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *IEEE Transact. Patt. Anal. and Mach. Int.* 15:220-232, 1993.

- KI 01** G. Kern-Isberner. Conditionals in nonmonotonic reasoning and belief revision, *LNAI 2087*. Springer, 2001.
- KIS 11** G. Kern-Isberner G.R. Simari. A Default Logical Semantics for Defeasible Argumentation. *Proc. of FLAIRS 2011*, AAAI Press, 2011.
- KLM 90** S. Kraus, D. Lehmann, M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. In *Artificial Intelligence*, 44:167-207, 1990.
- Mak 94** D. Makinson. General patterns of nonmonotonic reasoning. *Handbook of Logic in AI and LP*, vol. 3 (eds. Gabbay et al.): 35-110. Oxford University Press, 1994.
- Pea 90** J. Pearl. System Z: a natural ordering of defaults with tractable applications to nonmonotonic reasoning. *TARK 3*: 121-135. Morgan Kaufmann, 1990.
- Spo 88** W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. *Causation in Decision, Belief Change, and Statistics* (eds. W.L. Harper, B. Skyrms): 105-134. Kluwer, 1988.
- Spo 12** W. Spohn. *The Laws of Belief. Ranking Theory and Its Philosophical Applications*. Oxford University Press, Oxford 2012.
- Wey 94** E. Weydert. General belief measures. *UAI'94*, Morgan Kaufmann.
- Wey 95** E. Weydert. Defaults and infinitesimals. Defeasible inference by non-archimedeian entropy maximization. *UAI 95*: 540-547. Morgan Kaufmann, 1995.
- Wey 96** E. Weydert. System J - rev. entailment. *FAPR 96*:637-649. Springer, 1996.
- Wey 98** E. Weydert. System JZ - How to build a canonical ranking model of a default knowledge base. *KR 98*: 190-201. Morgan Kaufmann, 1998.
- Wey 03** E. Weydert. System JLZ - Rational default reasoning by minimal ranking constructions. *Journal of Applied Logic* 1(3-4): 273-308. Elsevier, 2003.
- Wey 11** E. Weydert. Semi-stable extensions for infinite frameworks. In *Proc. BNAIC 2012*: 336343.
- Wey 13** E. Weydert. On the Plausibility of Abstract Arguments. *ECSQARU 2013*, *LNAI 7958* (ed. L. van der Gaag): 522-533 Springer, 2013.