Monotonic and Nonmonotonic Inference for Abstract Argumentation

Richard Booth[1] **Souhila Kaci**[2] **Tjitze Rienstra**[1,2] **Leendert van der Torre**[1]

[1]**Université du Luxembourg**
6 rue Richard Coudenhove-Kalergi, Luxembourg
richard.booth/tjitze.rienstra/leon.vandertorre@uni.lu

[2]**LIRMM (CNRS/Université Montpellier 2)**
161 rue Ada, Montpellier, France
souhila.kaci@lirmm.fr

Abstract

We present a new approach to reasoning about the outcome of an argumentation framework, where an agent's reasoning with a framework and semantics is represented by an inference relation defined over a logical labeling language. We first study a monotonic type of inference which is, in a sense, more general than an acceptance function, but equally expressive. In order to overcome the limitations of this expressiveness, we study a non-monotonic type of inference which allows *counterfactual* inferences. We precisely characterize the classes of frameworks distinguishable by the non-monotonic inference relation for the admissible semantics.

1 Introduction

An argumentation framework (Dung 1995) (or framework, for short) consists of a set of arguments, whose content may be left unspecified, together with an attack relation encoding conflict between arguments. Given a framework, a semantics specifies which sets of arguments (called extensions) are rationally acceptable. This formalism captures many different types of reasoning considered in the area of AI.

In many applications, a framework somehow represents (part of) an agent's belief state. Beliefs are then formed on the basis of acceptable sets of arguments. For example, a 'grounded reasoner' forms beliefs on the basis of the framework's grounded extension, a 'preferred reasoner' on the basis of the preferred extensions, and so on. There is a problem with this account, however. We demonstrate this using the frameworks shown in figure 1. The problem is that, under the admissible semantics, F_3, F_4, F_5 and F_6 have the same set of extensions. Even worse, under most other semantics, they *all* have the same set of extensions. Thus, the beliefs of an agent whose belief state consists of one of them are equivalent to the beliefs of an agent whose belief state consists of any of the other ones. Despite the obvious differences, the six frameworks are equivalent, in this sense.

A more appropriate notion of equivalence is *strong equivalence* (Oikarinen and Woltran 2011). Given a semantics, two frameworks are said to be strongly equivalent if their extensions are the same given every possible addition of new arguments and attacks. Indeed, strong equivalence allows us to see that the six frameworks in figure 1 are different. But still, it leaves open the question of how to form beliefs on the basis of a framework, so that different frameworks can be meaningfully distinguished, even if their extensions are the same. This is the problem we address in this paper.

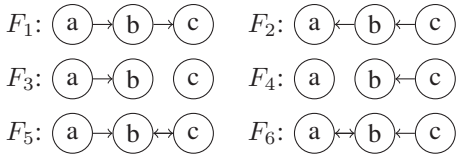

Figure 1: Six argumentation frameworks

To solve this we propose a new approach to reasoning about the outcome of a framework. The basis is a logical *labeling language*, where formulas assign labels (in, out, undecided) to arguments (Caminada 2006). A framework F (given a semantics s) then corresponds to one of two types of inference relations $\vdash_s^F$ or $\sim_s^F$, defined over the labeling language. Each of them represents a particular type of reasoning that an agent may perform, where $\phi \vdash_s^F \psi$ (or $\phi \sim_s^F \psi$) means that from ϕ, together with the knowledge encoded by F under the semantics s, the agent infers ψ. In contrast with the first type, which is monotonic, the second type is non-monotonic and embodies a *counterfactual* mode of inference. That is, it gives useful conclusions, even if the premise is normally false, e.g., stating that a is rejected, while a is always accepted. We show that with this type of inference we can distinguish different frameworks in a meaningful way, even if their extensions are the same. A possible application of our theory is in dialogues where, for example, agents must reason about about their framework, in order to decide how to persuade an opponent.

The outline of this paper is as follows. After presenting the necessary basics in section 2, we present in section 3 the logical labeling language, characterize a number of semantics in this language, and present a first type of (monotonic) inference relation. We then discuss the limitations of this relation, which are addressed in section 5, where we present our second, non-monotonic type of inference. Before concluding, we discuss in section 6 some properties and give a precise characterization of the expressiveness of the non-monotonic admissible inference relation.

2 Preliminaries

An argumentation framework (or framework, for short) consists of a set of arguments and an attack relation between arguments (Dung 1995). Formally:

Definition 1 *A framework F is a pair (A_F, R_F) where A_F is a set of* arguments *and $R_F \subseteq A_F \times A_F$ an* attack *relation. If $(x, y) \in R_F$, we say that x* attacks *y.*

We assume that for every F, A_F is a finite subset of a fixed set $\mathcal{U}$ of arguments and denote the set of all frameworks by $\mathcal{F}$. The attack relation encodes conflict between arguments: if x attacks y then they cannot be accepted together. The direction of the attack, that need not be symmetric, gives rise to a 'dialectical arrangement' of the arguments involved. That is, if x attacks y then x 'opposes' or 'argues against' y, while y does not necessarily argue against x. The goal is to select sets of arguments, called *extensions*, that are rationally acceptable. A *semantics* embodies a set of conditions that an extension must satisfy. The most studied ones are defined as follows:

Definition 2 *Let F be a framework. An* extension *of F is a set $E \subseteq A_F$. We say that E is* conflict-free *iff $\nexists x, y \in E$ s.t. $(x, y) \in R_F$. An argument $x \in A_F$ is* defended *by E iff $\forall (y, x) \in R_F$, $\exists z \in E$ s.t. $(z, y) \in R_F$. Given an extension E, we define Def(E) by Def(E) $= \{x \in A_F \mid E \text{ defends } x\}$. An extension $E \subseteq A_F$ is said to be:*

- admissible *iff E is conflict free and $E \subseteq$ Def(E).*
- complete *iff E is conflict free and $E =$ Def(E).*
- stable *iff E is admissible and $\forall x \in A_F \setminus E$, $\exists y \in E$ s.t. $(y, x) \in R_F$.*
- preferred *iff E is maximal (w.r.t. set inclusion) among the set of admissible labelings of F.*
- grounded *iff E is minimal (w.r.t. set inclusion) among the set of complete labelings of F.*

A semantics is usually captured by an *acceptance function*, that takes as input a framework and returns a set of extensions:

Definition 3 *An* acceptance function *is a function $\mathcal{E}_s : \mathcal{F} \to 2^{2^{\mathcal{U}}}$, where $s \in \{Ad, Co, St, Pr, Gr\}$ is called the* semantics *of $\mathcal{E}_s$. An acceptance function $\mathcal{E}_s$ satisfies the condition $\forall E \in \mathcal{E}_s(F)$, $E \subseteq A_F$. For $s \in \{Ad, Co, St, Pr, Gr\}$, $\mathcal{E}_s(F)$ is defined to return the set of all admissible, complete, etc. extensions of F.*

Given a framework F, semantics s and argument $x \in A_F$, two types of inference are usually considered. The first is *skeptical acceptance*, which amounts to checking whether *all* s-extensions include x; and the second is *credulous acceptance*, which amounts to checking whether *at least one* s-extension includes x. The result $\mathcal{E}_s(F)$ is always non-empty and may contain more than one extension, with two exceptions: $\mathcal{E}_{Gr}(F)$ always returns a single extension (which is equivalent to the set of arguments skeptically accepted under the complete semantics) and $\mathcal{E}_{St}(F)$ may be empty.

3 Monotonic inference for argumentation

In this section we present a number of inference relations that characterize the different argumentation semantics mentioned in the previous section. Given a finite set of arguments $A \subseteq \mathcal{U}$, the language $\mathcal{L}_A$ is generated by the following BNF: (where $x \in A$).

$$\phi := \mathtt{in}_x \mid \mathtt{out}_x \mid \mathtt{u}_x \mid \neg\phi \mid \phi \vee \phi$$

Here, $\mathtt{in}_x$ means that the argument x is accepted, $\mathtt{out}_x$ that it is rejected, and $\mathtt{u}_x$ that it is undecided—that is, neither rejected nor accepted. We also use the connectives $\wedge, \rightarrow, \leftrightarrow$ defined, as usual, in terms of $\neg$ and $\vee$, and the symbols $\bot, \top$ for contradiction and tautology. A *model* over $\mathcal{L}_A$ is a triple $m = (I_m, O_m, U_m)$ with $I_m, O_m, U_m \subseteq A$, $I_m \cup O_m \cup U_m = A$ and $I_m \cap O_m = I_m \cap U_m = U_m \cap O_m = \emptyset$. We denote the set of all models over $\mathcal{L}_A$ by $\mathcal{M}_A$. What we call a model is often called a *labeling* in the labeling-based semantics (Caminada 2006). Like in labeling-based semantics, we say that arguments in I_m, O_m and U_m are, respectively, (labeled) *in* and *out* and *undecided* in m. However, in line with the more classical formalization we have in mind, we call m model rather than a labeling. Interpretation is defined as follows:

Definition 4 *The satisfaction relation $\models \subseteq \mathcal{M}_{\mathcal{A}} \times \mathcal{L}_A$ is defined by: $m \models \mathtt{in}_x$ iff $x \in I_m$; $m \models \mathtt{out}_x$ iff $x \in O_m$; $m \models \mathtt{u}_x$ iff $x \in U_m$; $m \models \phi \vee \psi$ iff $m \models \phi$ or $m \models \psi$; and $m \models \neg\phi$ iff $m \not\models \phi$.*

As usual, we define $[\phi]$ by $[\phi] = \{m \mid m \models \phi\}$, write $\phi \models \psi$ if and only if $[\phi] \subseteq [\psi]$ and $\models \phi$ if and only if $m \models \phi$ for all m in $\mathcal{M}_A$. We first define four special formulas that characterize, given a framework F, the conditions of conflict-freeness, admissibility, completeness and stability:

Definition 5 *Given a framework F, we define:*

- *the* attack restriction α_F *by* $\alpha_F = \wedge_{(y,x) \in R_F} \alpha(x, y)$, *where* $\alpha(x, y)$ *is defined by* $\alpha(x, y) = (\mathtt{in}_y \rightarrow \mathtt{out}_x) \wedge (\mathtt{in}_x \rightarrow \mathtt{out}_y)$.
- *the* admissible restriction β_F *by* $\beta_F = \wedge_{x \in A_F} \beta_F(x)$, *where* $\beta_F(x)$ *is defined by* $\beta_F(x) = \mathtt{out}_x \rightarrow (\vee_{(y,x) \in R_F} \mathtt{in}_y)$.
- *the* complete restriction γ_F *by* $\gamma_F = \wedge_{x \in A_F} \gamma_F(x)$, *where* $\gamma_F(x)$ *is defined by* $\gamma_F(x) = (\wedge_{(y,x) \in R_F} \mathtt{out}_y) \rightarrow \mathtt{in}_x$.
- *the* stable restriction σ_F *by* $\sigma_F = \wedge_{x \in A_F} \sigma(x)$, *where* $\sigma(x)$ *is defined by* $\sigma(x) = \mathtt{in}_x \vee \mathtt{out}_x$.

In an attack restriction, $\alpha(x, y)$ states that if y (resp. x) is labeled in, then x (resp. y) is labeled out. In an admissibility or completeness restriction, $\beta_F(x)$ says that if we reject x, we must do so for a reason (i.e., x must have an in-labeled attacker); and $\gamma_F(x)$ says that if we have no reason not to accept x (i.e., all attackers are out) then we must accept it. Notice that the reverse conditions of those expressed by $\beta_F(x)$ and $\gamma_F(x)$ are already implied by α_F. Finally, $\sigma(x)$ simply states that x is either in or out. We have the following correspondences with abstract argumentation semantics:

Proposition 1 *For all $F \in \mathcal{F}$, we have:*

- $\{I_m \mid m \in [\alpha_F \wedge \beta_F]\} = \mathcal{E}_{Ad}(F)$,

- $\{I_m \mid m \in [\alpha_F \wedge \beta_F \wedge \gamma_F]\} = \mathcal{E}_{Co}(F)$.
- $\{I_m \mid m \in [\alpha_F \wedge \beta_F \wedge \gamma_F \wedge \sigma_F]\} = \mathcal{E}_{St}(F)$.

We will call a model satisfying $\alpha_F \wedge \beta_F$, $\alpha_F \wedge \beta_F \wedge \gamma_F$ and $\alpha_F \wedge \beta_F \wedge \gamma_F \wedge \sigma_F$ an *admissible*, *complete* and *stable* model, respectively. Note that our definition of admissible differs from the one used by Caminada and Gabbay (2009), which does not establish a 1-to-1 correspondence between admissible sets and labelings. Our definition does and is equivalent to what they call a *JV-labeling*, which is based on an earlier proposal by Jacobovits and Vermeir (1999).

As we said, the different argumentation semantics are usually represented by acceptance functions that take as input a framework and return a set of extensions. Conditional acceptance functions, which take as additional input a condition, have been studied as well (Booth et al. 2012). Here we take another approach: Given a framework F, each semantics s corresponds to an inference relation parametrized by F and s, i.e., $\vdash_s^F$. The framework and semantics act as a set of background assumptions that are taken for granted when establishing whether some conclusion follows from a premise, such that $\phi \vdash_s^F \psi$ means that ψ follows from ϕ, together with the knowledge encoded by F under the semantics s. Formally:

Definition 6 *Given a framework F, we define the admissible, complete and stable inference relations $\vdash_{Ad}^F$, $\vdash_{Co}^F$ and $\vdash_{St}^F$ by:*

- $\phi \vdash_{Ad}^F \psi$ *iff* $\phi \wedge \alpha_F \wedge \beta_F \models \psi$,
- $\phi \vdash_{Co}^F \psi$ *iff* $\phi \wedge \alpha_F \wedge \beta_F \wedge \gamma_F \models \psi$,
- $\phi \vdash_{St}^F \psi$ *iff* $\phi \wedge \alpha_F \wedge \beta_F \wedge \sigma_F \models \psi$.

The semantics discussed above are compositional, in the sense that we can define a property for each argument x separately (i.e., using $\beta_F(x)$, $\gamma_F(x)$ and $\sigma_F(x)$) such that, when *all* arguments satisfy this property, our models correspond to extensions under the respective semantics. It seems difficult to do this for the preferred semantics. The most natural way to define preferred inference is by delimiting the set of models we range over, when deciding whether a conclusion follows from a premise. The models that we range over are the *preferred models*, i.e., admissible models that are maximal with respect to the in labeled arguments. Formally:

Definition 7 *Given a set of models M, the set of* preferred models *of M , denoted by $Pr[M]$, is defined by $Pr[M] = \{m \in M \mid \nexists m' \in M, I_{m'} \supset I_m\}$.*

Proposition 2 *For all $F \in \mathcal{F}$, we have $\{I_m \mid m \in Pr[\alpha_F \wedge \beta_F]\} = \mathcal{E}_{Pr}(F)$.*

Definition 8 *Given a framework F we define the preferred inference relation $\vdash_{Pr}^F$ by $\phi \vdash_{Pr}^F \psi$ if and only if $Pr[\alpha_F \wedge \beta_F] \cap [\phi] \subseteq [\psi]$.*

Note that the admissible, complete and stable inference relations are instances of what Makinson calls *pivotal assumption* relations, while the preferred inference relation is a *pivotal valuation* relation (Makinson 2005). They act, in Makinson's discussions, as a conceptual bridge between monotonic and non-monotonic inference relations. These relations are supraclassical (i.e., $\phi \models \psi$ implies $\phi \vdash_s^F \psi$) but still monotonic.

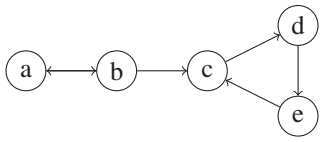

Figure 2: An argumentation framework

There are many ways to formalize abstract argumentation using logic, and ours is only one of them. Related work in this area include reductions of frameworks to propositional logic formulas, with the goal of finding extensions by checking whether an extension corresponds to a model of a formula, or by checking that a formula is satisfiable (Besnard and Doutre 2004); reductions of frameworks to logic programs under the answer-set semantics (Egly, Gaggl, and Woltran 2010); a study of a logical language consisting of attack and defense connectives (Boella, Hulstijn, and van der Torre 2005); and a formalization of fragments of argumentation theory using modal logic (Grossi 2010). Our approach simply provides us with a method to reason about the acceptability of arguments of a framework under some semantics s. For example, we can reason about skeptical and credulous acceptance:

Proposition 3 *Given a framework F, semantics $s \in \{Ad, Co, St, Pr\}$ and argument $x \in A_F$, we have that x is skeptically accepted iff $\top \vdash_s^F \mathtt{in}_x$ and x is credulously accepted iff $\top \nvdash_s^F \neg\mathtt{in}_x$.*

On the one hand, the representation of a semantics by a monotonic inference relation is more general than an acceptance function. This is demonstrated by the following example, where the inferences go beyond establishing just skeptical or credulous acceptance.

Example 1 *Let F be the framework shown in figure 2. We will denote a model m as a sequence $(l_a l_b l_c l_d l_e)$ where each l_x is the label of x. We use a similar notation in later examples. We have $[\alpha_F \wedge \beta_F] = \{(UUUUU), (IOUUU), (OIOUU), (OIOIO)\}$, $[\alpha_F \wedge \beta_F \wedge \gamma_F] = \{(UUUUU), (IOUUU), (OIOIO)\}$ and $[\alpha_F \wedge \beta_F \wedge \gamma_F \wedge \sigma_F] = \{(OIOIO)\}$. We can infer, e.g., $\top \vdash_{Co}^F \mathtt{out}_c \vee \mathtt{u}_c$, $\mathtt{out}_a \vdash_{Co}^F \mathtt{in}_d$, $\mathtt{in}_a \vdash_{St}^F \bot$ and $\top \vdash_{Pr}^F \mathtt{in}_a \vee \mathtt{in}_b$.*

On the other hand, all we can learn from $\vdash_s^F$ can also be learnt by just looking at the set $\mathcal{E}_s(F)$, and vice versa. Formally, we can state this as follows:

Proposition 4 *For all $F, G \in \mathcal{F}$ and all $s \in \{Ad, Co, St, Pr\}$, $\mathcal{E}_s(F) = \mathcal{E}_s(G)$ if and only if $\vdash_s^F = \vdash_s^G$.*

Thus, in terms of expressive power—that is, the classes of frameworks that we can distinguish—we lost nothing and gained nothing. In the next section, we will take a closer look at the limits of this expressiveness. After this, we will present a number of non-monotonic inference relations, that generalize the monotonic relations presented here, and that improve upon the limits of their expressiveness.

Before we move on, however, a remark about monotonicity seems to be in order. As we already pointed out, all the inference relations presented so far are monotonic. That is, for all $s \in \{Ad, Co, St, Pr\}$, we have $\phi \mid\!\sim_s^F \psi$ implies $\phi \wedge \chi \mid\!\sim_s^F \psi$. We can also look at (non-)monotonicity w.r.t. adding arguments and attacks. Monotonicity w.r.t. arguments (resp. attacks) then holds for a relation $\mid\!\sim_s^F$ iff $A_F \subseteq A_G$ (resp. $R_F \subseteq R_G$) implies $\mid\!\sim_s^F \subseteq \mid\!\sim_s^G$. We then find that for all $s \in \{Ad, Co, St, Pr\}$, $\mid\!\sim_s^F$ is monotonic w.r.t. adding arguments and—more interestingly—non-monotonic w.r.t. adding attacks. In this paper we will not look at this type of non-monotonicity.

4 Expressivity of monotonic inference

We are interested in the expressive power of the semantics introduced in the previous section. In terms of inference relations, we should then ask: What are the classes of frameworks that the inference relation can distinguish? Stated formally, given a relation $\mid\!\sim_s^F$, two frameworks F and G belong to the same indistinguishable class if $\mid\!\sim_s^F = \mid\!\sim_s^G$. We do not aim at making formal statements about this expressive power, but suffice in briefly showing its inadequacy.

Consider the six frameworks shown in figure 1. They all encode different relations between the arguments. For example, in framework 1, b is out because a is in. In framework 2, however, b is out because c is in, and not because a is in! For complete, preferred and stable monotonic inference, however, all six frameworks are indistinguishable. That is, for all $s \in \{Co, St, Pr\}$ and $i, j \in \{1, \ldots, 6\}$, we have $\mid\!\sim_s^{F_i} = \mid\!\sim_s^{F_j}$. Thus, the monotonic inference relations and corresponding semantics fail to distinguish the differences between the six frameworks. Admissible inference fares better: all frameworks are distinguishable for $\mid\!\sim_{Ad}^F$, except for 5 and 6, i.e., $\mid\!\sim_{Ad}^{F_5} = \mid\!\sim_{Ad}^{F_6}$. In sum, none of the inference relations is able to distinguish all six frameworks, and it is easy to come up with more examples of indistinguishable frameworks (e.g., none of the relations can distinguish any of the six frameworks in figure 1 with or without a self-attack added to b).

We can understand this limitation when we realize that these inference relations cannot deal with what we may call *counterfactual inference*. We can distinguish the frameworks 1 and 2 under the complete semantics, for example, if we could check what follows if a is out. But a being out cannot be satisfied together with $\beta_F \wedge \gamma_F$. That is, the semantics has ruled out all models in which a is out, so we cannot verify what the label of b and c would have been, if a would have been out.

In the next section, we present a number of non-monotonic generalizations of the inference relations presented in section 3. They allow us to reason counterfactually, and thus improve upon the limited expressivity of their monotonic counterparts.

5 Non-monotonic inference for argumentation

The non-monotonic generalizations of the inference relations that we present in this section are defined using the preferential model semantics, where ψ follows from ϕ iff ψ is true in the most preferred models of ϕ (Kraus, Lehmann, and Magidor 1990). Here, 'most preferred' is understood as most admissible, complete, stable or preferred in the sense of 'better' satisfying the constraints of the respective semantics. The preferential model semantics for argumentation has been considered before in (Roos 2010). The approach there is to define a preference relation over conflict free sets of arguments based directly on the attack relation. The most preferred states are then maximal conflict free sets and additional criteria for selecting preferred states give rise to different semantics.

Our approach is to prefer models that satisfy e.g. $\beta_F(x)$ or $\beta_F(x) \wedge \gamma_F(x)$ for maximal sets of arguments. This corresponds intuitively to the idea of counterfactual inference, mentioned earlier. Namely, in order to know what follows from ϕ even though, in accordance with the semantics, all models that satisfy ϕ have been ruled out, we must look at what follows from the models that satisfy the constraints imposed by the semantics 'as much as possible'. We use the following notation:

Definition 9 *Given a framework F and model $m \in \mathcal{M}_{A_F}$ we define β_F^m by $\beta_F^m = \{x \in A_F \mid m \models \beta_F(x)\}$ and γ_F^m by $\gamma_F^m = \{x \in A_F \mid m \models \gamma_F(x)\}$.*

We denote the relations by $\mid\!\sim_s^F$, for $s \in \{Ad, Co, St, Pr\}$. We now proceed with the formal definition of the non-monotonic inference relations. After this, we turn again to the matter of expressivity, and we discuss some properties that the non-monotonic inference relations satisfy.

Non-monotonic admissible inference

Admissible models are conflict-free (i.e., satisfying α_F) models that satisfy $\beta_F(x)$ for all $x \in A_F$. Thus, 'more admissible' means that $\beta_F(x)$ is satisfied for more arguments and, given two conflict-free models m, m', we say that m is 'at least as admissible' as m' if and only if $\beta_F^m \supseteq \beta_F^{m'}$. This gives us the *admissibility order* $\preceq_{Ad}$:

Definition 10 *Given a framework F, the order $\preceq_{Ad}^F \subseteq \mathcal{M}_{A_F} \times \mathcal{M}_{A_F}$ is defined by $m \preceq_{Ad}^F m'$ iff $\beta_F^m \subseteq \beta_F^{m'}$.*

Note that $\preceq_{Ad}^F$ is a partial pre-order. As usual, given a partial pre-order $\preceq$, we denote by $\prec$ the strict partial order derived from $\preceq$ and define $\approx$ by $m \approx m'$ iff $m \preceq m'$ and $m \succeq m'$. The *non-monotonic admissible inference* relation $\mid\!\sim_{Ad}^F$ is now defined as follows:

Definition 11 *Given a framework F, we define $\mid\!\sim_{Ad}^F$ by $\phi \mid\!\sim_{Ad}^F \psi$ iff $max_{\prec_{Ad}^F}([\alpha_F \wedge \phi]) \subseteq [\psi]$.*

Example 2 *Consider figure 1. We apply non-monotonic admissible inference to the frameworks F_1 and F_2. We will consider the question: what would hold, under the admissible semantics if a were out? For F_1, the most admissible models that satisfy $\mathtt{out}_a$ are $m_1 = (OIO)$ and $m_2 = (OUU)$, with $\beta_{F_1}^{m_1} = \beta_{F_1}^{m_2} = \{b, c\}$. Thus we have $\mathtt{out}_a \mid\!\sim_{Ad}^{F_1} \mathtt{out}_c \vee \mathtt{u}_c$. For F_2, the most admissible models that satisfy $\mathtt{out}_a$ are $m_1 = (OIO), m_2 = (OOI), m_3 =$*

(OUU), *with* $\beta^{m_1}_{F_2} = \{a, b\}$, $\beta^{m_2}_{F_2} = \beta^{m_3}_{F_2} = \{b, c\}$. *Thus we have* $\mathtt{out}_a \not\sim^{F_2}_{Ad} \mathtt{out}_c \vee \mathtt{u}_c$.

Non-monotonic complete inference

Complete models are conflict-free (i.e., satisfying α_F) models that satisfy $\beta_F(x) \wedge \gamma_F(x)$ for all $x \in A_F$. So how do we define 'most complete'? Given two conflict-free models m, m', we can say that m is 'at least as complete' as m' if $\beta^m_F \cap \gamma^m_F \supseteq \beta^{m'}_F \cap \gamma^{m'}_F$. This gives us what we call the $\beta\gamma$ order:

Definition 12 *Given a framework* F*, the order* $\preceq^F_{\beta\gamma} \subseteq \mathcal{M}_{A_F} \times \mathcal{M}_{A_F}$ *is defined by* $m \preceq^F_{\beta\gamma} m'$ *iff* $\beta^m_F \cap \gamma^m_F \subseteq \beta^{m'}_F \cap \gamma^{m'}_F$.

This leaves us with the question of what to do when $\preceq^F_{\beta\gamma}$ is indifferent about two models m, m' (i.e., we have $m \approx^F_{\beta\gamma} m'$). We will then prefer the most admissible one among the two. So, what we actually do, is maximize according to the lexicographic order of first $\preceq^F_{\beta\gamma}$ and then $\preceq^F_{Ad}$. This gives us the *completeness order* $\preceq^F_{Co}$:

Definition 13 *Given a framework* F*, the order* $\preceq^F_{Co} \subseteq \mathcal{M}_{A_F} \times \mathcal{M}_{A_F}$ *is defined by* $m \preceq^F_{Co} m'$ *iff* $m \prec^F_{\beta\gamma} m'$ *or* $(m \approx^F_{\beta\gamma} m'$ *and* $m \preceq^F_{Ad} m')$

Note again that $\preceq^F_{Co}$ is a partial pre-order. We can now define the *non-monotonic complete inference* relation $\mid\!\sim^F_{Co}$:

Definition 14 *Given a framework* F*, we define* $\mid\!\sim^F_{Co}$ *by* $\phi \mid\!\sim^F_{Co} \psi$ *iff* $max_{\prec^F_{Co}}([\alpha_F \wedge \phi]) \subseteq [\psi]$.

Example 3 *We apply non-monotonic complete inference to the frameworks* F_1 *and* F_2 *of figure 1. What would hold, under the complete semantics, if* a *were out? For* F_1*, the most complete model that satisfies* $\mathtt{out}_a$ *is* $m = (OIO)$ *with* $\beta^m_{F_1} \cap \gamma^m_{F_1} = \{b, c\}$*. Thus we have* $\mathtt{out}_a \mid\!\sim^{F_1}_{Co} \mathtt{out}_c$*. For* F_2*, the most complete models that satisfy* $\mathtt{out}_a$ *are* $m_1 = (OIO)$ *and* $m_2 = (OOI)$*, with* $\beta^{m_1}_{F_2} \cap \gamma^{m_1}_{F_2} = \{a, b\}$ *and* $\beta^{m_2}_{F_2} \cap \gamma^{m_2}_{F_2} = \{b, c\}$*. Thus we have* $\mathtt{out}_a \not\sim^{F_1}_{Co} \mathtt{out}_c$.

Intermezzo: Five valued labelings

Before we move on to non-monotonic stable and preferred inference, we will make a remark about a notion that we naturally end up with, namely a *five-valued labeling*. If we look at a model $m \in [\alpha_F]$ that is conflict-free but does not necessarily satisfy the admissibility and completeness restriction, we have, for all $x \in A_F$, five mutually exclusive and collectively exhaustive possibilities:

1. $m \models \mathtt{in}_x$ and $m \models \beta_F(x) \wedge \gamma_F(x)$.
2. $m \models \mathtt{u}_x$ and $m \models \beta_F(x) \wedge \gamma_F(x)$.
3. $m \models \mathtt{out}_x$ and $m \models \beta_F(x) \wedge \gamma_F(x)$.
4. $m \models \mathtt{u}_x$ and $m \models \neg\gamma_F(x) \wedge \beta_F(x)$. This means that all attackers of x are out in m. We call the argument *forced undecided*.
5. $m \models \mathtt{out}_x$ and $m \models \neg\beta_F(x)$. This means that x has no attacker that is in. We call the argument *forced out*.

Note that $\mathtt{in}_x$ and $\mathtt{u}_x$ already imply $\beta_F(x) \wedge \gamma_F(x)$ and $\beta_F(x)$, respectively. So, given a conflict free model m (i.e., $m \in [\alpha_F]$) we can assign one of the five statuses in, undecided, out, forced undecided or forced out to an argument. The concept of a five-valued labeling is a useful one, when reasoning about labelings that are only partially admissible or complete. For example, if we look at the orderings presented previously, we have that:

- The order $\preceq_{Ad}$ minimizes arguments that are forced out,
- The order $\preceq_{\beta\gamma}$ minimizes arguments that are forced either out or undecided.
- The order $\preceq_{Co}$ minimizes first arguments that are forced either out or undecided, and additionally prefers forced undecided over forced out.

Non-monotonic stable inference

Stable models are admissible models with the additional restriction that no arguments are undecided. We will model non-monotonic stable inference not by minimizing undecided arguments but by restricting the models that we consider to those that satisfy the stable restriction. Formally:

Definition 15 *Given a framework* F*, we define* $\mid\!\sim^F_{St}$ *by* $\phi \mid\!\sim^F_{St} \psi$ *iff* $max_{\prec^F_{Ad}}([\alpha_F \wedge \sigma_F \wedge \phi]) \subseteq [\psi]$.

Note that if we had defined $\mid\!\sim^F_{St}$ by minimizing undecided sets, we would not get backwards compatibility—a property we discuss in the following section. Such a definition, however, does seem to correspond to a semantics, namely the *semi-stable* semantics (Caminada, Carnielli, and Dunne 2012). We leave this for future work.

Example 4 *We apply non-monotonic stable inference to the frameworks* F_3 *and* F_4 *of figure 1. What would hold, under the stable semantics, if* a *were out? For* F_3*, the most admissible model that is stable and satisfies* $\mathtt{out}_a$ *is* $m = (OII)$ *with* $\beta^m_{F_1} = \{b, c\}$*. Thus we have, for example,* $\mathtt{out}_a \mid\!\sim^{F_1}_{St} \mathtt{in}_b$*. For* F_4*, the most admissible model that is stable and satisfies* $\mathtt{out}_a$ *is* $m_1 = (OOI)$ *with* $\beta^m_{F_2} = \{b, c\}$*. So here we have* $\mathtt{out}_a \not\sim^{F_1}_{St} \mathtt{in}_b$.

Non-monotonic preferred inference

Preferred models are admissible models that have a maximal (w.r.t. set-inclusion) set of in-labeled arguments. Accordingly, given two conflict-free (i.e., satisfying α_F) models m, m', we take m to be 'at least as preferred' as m' if m is at least as admissible as m' (i.e., if $m' \preceq^F_{Ad} m$). Furthermore, if $\preceq^F_{Ad}$ is indifferent about m and m', we take m to be at least as preferred as m' if $I_m \supseteq I_{m'}$—which sets what we call the Pr order apart from the admissible order. Formally:

Definition 16 *Given a framework* F*, the order* $\preceq^F_{Pr} \subseteq \mathcal{M}_{A_F} \times \mathcal{M}_{A_F}$ *is defined by* $m \preceq^F_{Pr} m'$ *iff* $m \prec^F_{Ad} m'$ *or* $(m \approx^F_{Ad} m'$ *and* $I_m \subseteq I_{m'})$

Note again that $\preceq^F_{Pr}$ is a partial pre-order. We can now define the *non-monotonic preferred inference* relation $\mid\!\sim^F_{Pr}$:

Definition 17 *Given a framework* F*, we define* $\mid\!\sim^F_{Pr}$ *by* $\phi \mid\!\sim^F_{Pr} \psi$ *iff* $max_{\prec^F_{Pr}}([\alpha_F \wedge \phi]) \subseteq [\psi]$.

Example 5 *We apply non-monotonic preferred inference to the framework F_5 in figure 1. What would hold, under the preferred semantics, if a were out? For F_5, the most preferred models that satisfy $\mathtt{out}_a$ are $m_1 = (OIO)$ and $m_2 = (OOI)$ with $\beta^{m_1}_{F_1} = \beta^{m_2}_{F_1} = \{b, c\}$. Thus we have $\mathtt{out}_a \mid\sim^{F_1}_{Co} (\mathtt{in}_b \wedge \mathtt{out}_c) \vee (\mathtt{out}_b \wedge \mathtt{in}_c)$.*

6 Expressivity and properties

We demonstrated non-monotonic inference by a number of examples, most of which showed that it is able to distinguish frameworks that the monotonic inference relations cannot distinguish. Furthermore, the non-monotonic inference relations agree on the consequences of ϕ with their monotonic counterparts, whenever this is consistently possible (i.e. ϕ is credulously accepted). Formally this property, which we call *backwards compatibility*, is expressed as follows.

Proposition 5 *For all F and all $s \in \{Ad, Co, St, Pr\}$), if $\top \not\mid\sim^F_s \neg\phi$ then ($\phi \mid\sim^F_s \psi$ iff $\phi \vdash^F_s \psi$).*

This property implies that $\top \vdash^F_s \phi$ iff $\top \mid\sim^F_s \phi$ whenever F has at least one s-extension, meaning that we can still express skeptical and credulous acceptance.

For admissible non-monotonic inference, we have obtained a precise characterization of its expressivity, stated in terms of a *kernel*. Given a semantics s, a kernel of a framework $F \in \mathcal{F}$ is a framework $F' \in \mathcal{F}$ such that $A_F = A_{F'}$ and $R_{F'} \subseteq R_F$. The idea is that in a kernel all attacks are removed of which the semantics cannot distinguish whether it is present. Whenever a semantics distinguishes exactly those frameworks F, G of which the kernels are different, the kernel is said to characterize the expressivity of the semantics.

The idea of using a kernel to characterize the expressivity of a semantics comes from (Oikarinen and Woltran 2011). Here, kernels are used to study the relation of *strong equivalence*. Two frameworks $F, G \in \mathcal{F}$ are said to be strongly equivalent w.r.t. a semantics s, denoted by $F \equiv_s G$, if and only if for all possible expansions $H \in \mathcal{F}$, we have that $\mathcal{E}_s((A_F \cup A_H, R_F \cup R_H)) = \mathcal{E}_s((A_G \cup A_H, R_G \cup R_H))$. One of the results obtained there is a characterization of the admissible semantics using the *admissible kernel*:

Definition 18 *Given a framework F, the* admissible kernel of F, *denoted by F^{ak} is defined by $F^{ak} = (A_F, R^{ak}_F)$ where $R^{ak}_F = R_F \backslash \{(x, y) \in R_F \mid x \neq y \wedge (x, x) \in R_F \wedge ((y, x) \in R_F \vee (y, y) \in R_F)\}$*

Proposition 6 *For all $F, G \in \mathcal{F}$, $F \equiv_{Ad} G$ iff $F^{ak} = G^{ak}$ (Oikarinen and Woltran 2011).*

We have obtained an analogous result for admissible non-monotonic inference:

Theorem 19 *For all $F, G \in \mathcal{F}$, $\mid\sim^F_{Ad} = \mid\sim^G_{Ad}$ iff $F^{ak} = G^{ak}$.*

The proof of the above is omitted due to space constraints. We leave the characterization of the expressivity of the other inference relations as future work.

7 Conclusions and future work

We presented a new approach to reasoning about the outcome of a framework. The approach is based on a logical labeling language. Different types of inference relations can be defined over this language, each representing a particular type of reasoning that an agent may do on the basis of a framework and a semanantics. We have shown, by example, that this approach allows us to distinguish frameworks in a meaningful way, even if they have the same set of extensions in the traditional approach. Finally, we characterized the expressiveness of admissible non-monotonic inference in terms of a kernel.

Among future work are the characterization of the expressivity of the other inference relations. We also plan to study specialized forms of non-monotonic inference to model, e.g., the difference between interpreting a premise as either an 'intervention' or an 'observation'. We may also look at the possibility of axiomatizing the non-monotonic inference relations, and a generalization that allows premises that violate conflict-freeness. Finally, we will use our inference relations to study properties of (non-)monotonicity w.r.t. adding arguments and attacks.

References

Besnard, P., and Doutre, S. 2004. Checking the acceptability of a set of arguments. In Delgrande, J. P., and Schaub, T., eds., *NMR*, 59–64.

Boella, G.; Hulstijn, J.; and van der Torre, L. W. N. 2005. A logic of abstract argumentation. In Parsons, S.; Maudet, N.; Moraitis, P.; and Rahwan, I., eds., *ArgMAS*, volume 4049 of *Lecture Notes in Computer Science*, 29–41. Springer.

Booth, R.; Kaci, S.; Rienstra, T.; and van der Torre, L. W. N. 2012. Conditional acceptance functions. In *COMMA*, 470–477.

Caminada, M. W. A., and Gabbay, D. M. 2009. A logical account of formal argumentation. *Studia Logica* 93(2-3):109–145.

Caminada, M. W. A.; Carnielli, W. A.; and Dunne, P. E. 2012. Semi-stable semantics. *J. Log. Comput.* 22(5):1207–1254.

Caminada, M. 2006. On the issue of reinstatement in argumentation. In Fisher, M.; van der Hoek, W.; Konev, B.; and Lisitsa, A., eds., *JELIA*, volume 4160 of *Lecture Notes in Computer Science*, 111–123. Springer.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–358.

Egly, U.; Gaggl, S.; and Woltran, S. 2010. Answer-set programming encodings for argumentation frameworks. *Argument and Computation* 1(2):147–177.

Grossi, D. 2010. On the logic of argumentation theory. In van der Hoek, W.; Kaminka, G. A.; Lespérance, Y.; Luck, M.; and Sen, S., eds., *AAMAS*, 409–416. IFAAMAS.

Jakobovits, H., and Vermeir, D. 1999. Robust semantics for argumentation frameworks. *J. Log. Comput.* 9(2):215–261.

Kraus, S.; Lehmann, D. J.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.* 44(1-2):167–207.

Makinson, D. 2005. *Bridges from classical to nonmonotonic logic*, volume 5 of *Texts in Computing*. King's College Publications.

Oikarinen, E., and Woltran, S. 2011. Characterizing strong equivalence for argumentation frameworks. *Artif. Intell.* 175(14-15):1985–2009.

Roos, N. 2010. Preferential model and argumentation semantics. In *Proceedings of the 13th International Workshop on Non-Monotonic Reasoning (NMR-2010)*.