An Algorithm for Computing Semi-Stable Semantics

Martin Caminada

Department of Information and Computing Sciences, Utrecht University

Technical Report UU-CS-2007-010

www.cs.uu.nl

ISSN: 0924-3275

An Algorithm for Computing Semi-Stable Semantics *

Martin Caminada
Utrecht University

April 20, 2007

Abstract

The semi-stable semantics for formal argumentation has been introduced as a way of approximating stable semantics in situations where no stable extensions exist. Semi-stable semantics can be located between stable semantics and preferred semantics in the sense that every stable extension is a semi-stable extension and every semi-stable extension is a preferred extension. Moreover, in situations where at least one stable extension exists, the semi-stable extensions are equal to the stable extensions. In this paper we provide the outline of an algorithm for computing the semi-stable extensions, given an argumentation framework. We show that with a few modifications, the algorithm can also be used for computing stable and preferred semantics.

1 Introduction

Formal argumentation, as a technique for defeasible entailment, has gained popularity since it combines a relatively easy to understand and human-style approach to reasoning with the mathematical rigidness that is required for software implementation [7]. It is also an interesting observation that many formalisms for nonmonotonic reasoning can be expressed as instances of formal argumentation [3].

Formal argumentation, in its simplest and most abstract form, is done using a set of abstract arguments and a defeat relation between these arguments. Since an argument A may be defeated by another argument B which may in its turn be defeated by a third argument C, the status of A (whether it can be accepted or not) partly depends on the status of C. Thus, what is needed is an overall criterion for determining which of the arguments can be considered to be ultimately justified. Several of such criteria have been proposed, the most well-known of these are grounded, preferred and stable semantics [3]. A relatively new proposal is semi-stable semantics [2]. Semi-stable semantics can be placed between stable semantics and preferred semantics, as every stable extension is also a semi-stable extension and every semi-stable extension is also a preferred extension. Moreover, semi-stable semantics can be seen as a way of approximating stable semantics in situations where no stable extensions exist.

This paper is structured as follows. In section 2 we introduce the basic notions of formal argumentation and argument based semantics. In section 3 we show how these notions can be described using the technique of argument labellings. The technique of argument labellings is relevant, since it is the basis of the formal background of the algorithm for computing the semi-stable extensions, which is treated in section 4. In section 5 we then treat an important optimization for the approach proposed until so far. We have decided

*This work was partly supported by the EU ASPIC project

to treat this in a separate section so that the algorithm can be introduced in a step-wise way. Then, in section 6, the actual algorithm is provided, using a high level description in pseudocode.

2 Argumentation and argument-based semantics

In this section, we provide a brief introduction on argument based semantics and the position of semi-stable semantics.

Definition 1. *An* argumentation framework *is a pair* (Ar, def) *where* Ar *is a finite set of arguments and* $def \subseteq Ar \times Ar$.

We say that an argument A *defeats* an argument B iff $(A, B) \in def$.

An argumentation framework can be represented as a directed graph in which the arguments are represented as nodes and the defeat relation is represented as arrows. In several examples throughout this paper, we will use this graph representation.

The shorthand notation A^+ and A^- stands for, respectively, the set of arguments defeated by A and the set of arguments that defeat A. Likewise, if $\mathcal{A}rgs$ is a set of arguments, then we write $\mathcal{A}rgs^+$ for the set of arguments that is defeated by at least one argument in $\mathcal{A}rgs$, and $\mathcal{A}rgs^-$ for the set of arguments that defeat at least one argument in $\mathcal{A}rgs$. In the definition below, $F(\mathcal{A}rgs)$ stands for the set of arguments that are acceptable in the sense of [3].

Definition 2 (defense / conflict-free)**.** *Let* $A \in Ar$ *and* $\mathcal{A}rgs \subseteq Ar$.
We define A^+ *as* $\{B \mid A\ def\ B\}$ *and* $\mathcal{A}rgs^+$ *as* $\{B \mid A\ def\ B\ for\ some\ A \in \mathcal{A}rgs\}$.
We define A^- *as* $\{B \mid B\ def\ A\}$ *and* $\mathcal{A}rgs^-$ *as* $\{B \mid B\ def\ A\ for\ some\ A \in \mathcal{A}rgs\}$.
$\mathcal{A}rgs$ *is* conflict-free *iff* $\mathcal{A}rgs \cap \mathcal{A}rgs^+ = \emptyset$.
$\mathcal{A}rgs$ defends *an argument* A *iff* $A^- \subseteq \mathcal{A}rgs^+$.
We define the function $F : 2^{Ar} \rightarrow 2^{Ar}$ *as*
$F(\mathcal{A}rgs) = \{A \mid A\ is\ defended\ by\ \mathcal{A}rgs\}$.

In the definition below, definitions of grounded, preferred and stable semantics are described in terms of complete semantics, which has the advantage of making the proofs in the remainder of this paper more straightforward. These descriptions are not literally the same as the ones provided by Dung [3], but as was first stated in [1], these are in fact equivalent to Dung's original versions of grounded, preferred and stable semantics.

Definition 3 (acceptability semantics)**.** *Let* $\mathcal{A}rgs$ *be a conflict-free set of arguments.*

- $\mathcal{A}rgs$ *is* admissible *iff* $\mathcal{A}rgs \subseteq F(\mathcal{A}rgs)$.
- $\mathcal{A}rgs$ *is a* complete *extension iff* $\mathcal{A}rgs = F(\mathcal{A}rgs)$.
- $\mathcal{A}rgs$ *is a* grounded *extension iff* $\mathcal{A}rgs$ *is the minimal (w.r.t. set-inclusion) complete extension.*
- $\mathcal{A}rgs$ *is a* preferred *extension iff* $\mathcal{A}rgs$ *is a maximal (w.r.t. set-inclusion) complete extension.*
- $\mathcal{A}rgs$ *is a* stable *extension iff* $\mathcal{A}rgs$ *is a complete extension that defeats every argument in* $Ar \backslash \mathcal{A}rgs$.
- $\mathcal{A}rgs$ *is a* semi-stable *extension iff* $\mathcal{A}rgs$ *is a complete extension where* $\mathcal{A}rgs \cup \mathcal{A}rgs^+$ *is maximal (w.r.t. set-inclusion).*

It is known that for every argumentation framework, there exists at least one admissible set (the empty set), exactly one grounded extension, one or more complete extensions, one or more preferred extensions and zero (!) or more stable extensions. Moreover, when the set of arguments in the argumentation framework is finite, there also exists one or more semi-stable extensions.

In [2] it is proved that every stable extension is also a semi-stable extension, and the every semi-stable extension is also a preferred extension. Moreover, it is observed that if the argumentation framework has at least one stable extension, then the set of semi-stable extensions is equal to the set of stable extensions. That is, when at least one stable extension exists, then stable semantics and semi-stable semantics coincide.

3 A brief introduction to argument labellings

The concepts of admissibility, as well as that of complete, grounded, preferred, stable or semi-stable semantics were originally stated in terms of sets of arguments. It is equally well possible, however, to express these concepts using argument labellings. This approach was originally proposed by Pollock [5] and has recently been extended by Caminada [1], Vreeswijk [7] and Verheij [6]. The idea of a labelling is to associate with each argument exactly one label, which can either be `in`, `out` or `undec`. The label `in` indicates that the argument is explicitly accepted, the label `out` indicates that the argument is explicitly rejected, and the label `undec` indicates that the status of the argument is undecided, meaning that one abstains from an explicit judgement whether the argument is `in` or `out`.

Definition 4. *A* labelling *is a function* $\mathcal{L} : Ar \longrightarrow \{\texttt{in}, \texttt{out}, \texttt{undec}\}$.

We write $\texttt{in}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \texttt{in}\}$, $\texttt{out}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \texttt{out}\}$ and $\texttt{undec}(\mathcal{L})$ for $\{A \mid \mathcal{L}(A) = \texttt{undec}\}$. Sometimes, we write a labelling $\mathcal{L}$ as a triple $(\mathcal{A}rgs_1, \mathcal{A}rgs_2, \mathcal{A}rgs_3)$ where $\mathcal{A}rgs_1 = \texttt{in}(\mathcal{L})$, $\mathcal{A}rgs_2 = \texttt{out}(\mathcal{L})$ and $\mathcal{A}rgs_3 = \texttt{undec}(\mathcal{L})$.

We distinguish three special kinds of labellings. The *all-in labelling* is a labelling that labels every argument `in`. The *all-out labelling* is a labelling that labels every argument `out`. The *all-undec labelling* is a labelling that labels every argument `undec`.

Definition 5. *Let $\mathcal{L}$ be a labelling and A be an argument. We say that:*

1. *A is* illegally `in` *iff A is labelled* `in` *but not all its defeaters are labeled* `out`
2. *A is* illegally `out` *iff A is labelled* `out` *but it does not have a defeater that is labelled* `in`
3. *A is* illegally `undec` *iff A is labelled* `undec` *but either all its defeaters are labelled* `out` *or it has a defeater that is labelled* `in`.

We say that a labelling has no illegal arguments *iff there is no argument that is illegally* `in`*, illegally* `out` *or illegally* `undec`*. We say that an argument is* legally `in` *iff it is labelled* `in` *and is not illegally* `in`*. We say that an argument is* legally `out` *iff it is labelled* `out` *and is not illegally* `out`*. We say that an argument is* legally `undec` *iff it is labelled undec and is not illegally undec.*

Definition 6. *An* admissible labelling *is a labelling without arguments that are illegally* `in` *and without arguments that are illegally* `out`.

Definition 7. *A* complete labelling *is a labelling without arguments that are illegally* `in`*, without arguments that are illegally* `out` *and without arguments that are illegally* `undec`.

Definition 8. *Let $\mathcal{L}$ be a complete labelling.*

- *We say that $\mathcal{L}$ is a* grounded labelling *iff* $\texttt{in}(\mathcal{L})$ *is minimal (w.r.t. set inclusion) among all complete labellings.*
- *We say that $\mathcal{L}$ is a* preferred labelling *iff* $\texttt{in}(\mathcal{L})$ *is maximal (w.r.t. set inclusion) among all complete labellings.*
- *We say that $\mathcal{L}$ is a* stable labelling *iff* $\texttt{undec}(\mathcal{L}) = \emptyset$.
- *We say that $\mathcal{L}$ is a* semi-stable labelling *iff* $\texttt{undec}(\mathcal{L})$ *is minimal (w.r.t. set inclusion) among all complete labellings.*

As an illustration of how the various types of labellings can be applied, consider the examples in Figure 1.

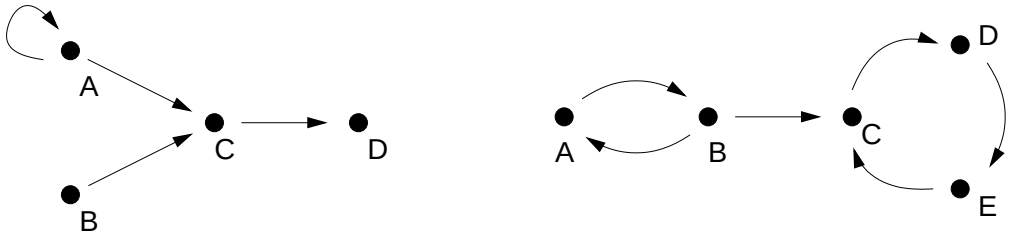

Figure 1: Two argumentation frameworks.

In the example at the left hand side of Figure 1, there exists just one complete labelling: $(\{B, D\}, \{C\}, \{A\})$, which is then automatically also grounded, preferred and semi-stable. The example at the left hand side does not have any stable labellings.

In the example at the right hand side of Figure 1, there exist three complete labellings: $(\emptyset, \emptyset, \{A, B, C, D, E\})$, $(\{A\}, \{B\}, \{C, D, E\})$ and $(\{B, D\}, \{A, C, E\}, \emptyset)$. The first labelling is the grounded labelling. The second and third labellings are both preferred labellings. The third labelling is also a stable and semi-stable labelling.

As for the admissible labellings, it should be mentioned that each complete labelling is also an admissible labelling. However, sometimes there exist admissible labellings that are not complete. Two examples of such labellings for the left hand side of Figure 1 are $(\{B\}, \emptyset, \{A, C, D\})$ and $(\{B\}, \{C\}, \{A, D\})$.

It is interesting to notice that an admissible labelling actually corresponds with the notion of an admissible set.

Theorem 1. *Let (Ar, def) be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$. $\mathcal{A}rgs$ is an admissible set iff there exists an admissible labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$.*

Proof.
"$\Longrightarrow$": Let $\mathcal{A}rgs$ be an admissible set. Now consider a labelling $\mathcal{L}$ with $\mathtt{in}(\mathcal{L}) = \mathcal{A}rgs$, $\mathtt{out}(\mathcal{L}) = \mathcal{A}rgs^+$ and $\mathtt{undec}(\mathcal{L}) = Ar - \mathtt{in}(\mathcal{L}) - \mathtt{out}(\mathcal{L})$. From the fact that $\mathcal{A}rgs$ is conflict-free it follows that $\mathcal{A}rgs \cap \mathcal{A}rgs^+ = \emptyset$, so $\mathcal{L}$ is well-defined.
We now prove that $\mathcal{L}$ has no arguments that are illegally `in`. Let A be an arbitrary argument that is labelled `in` by $\mathcal{L}$. Then $A \in \mathcal{A}rgs$. Let B be an arbitrary defeater of A. The fact that B is an admissible set means that $\mathcal{A}rgs$ contains an argument (say C) that defeats B. Therefore, $B \in \mathcal{A}rgs^+$, so B is labelled `out` by $\mathcal{L}$. As this holds for any arbitrary defeater B of A, it follows that each defeater of A is labelled `out`. Therefore, A is legally `in`.
We now prove that $\mathcal{L}$ has no arguments that are illegally `out`. Let A be an arbitrary argument that is labelled `out` by $\mathcal{L}$. Then $A \in \mathcal{A}rgs^+$. The fact that $A \in \mathcal{A}rgs^+$ means that there is some argument (say B) in $\mathcal{A}rgs$ that defeats A. The fact that B is in $\mathcal{A}rgs$ means that B is labelled `in` by $\mathcal{L}$. Thus, A has a defeater that is labelled `in`. Therefore, A is legally `out`.
"$\Longleftarrow$": Let $\mathcal{L}$ be an admissible labelling and let $\mathcal{A}rgs = \mathtt{in}(\mathcal{L})$.
We first prove that $\mathcal{A}rgs$ is conflict-free. Suppose this is not the case. Then there exist two (possibly the same) arguments $A, B \in \mathcal{A}rgs$ such that A defeats B. The fact that $A, B \in \mathcal{A}rgs$ means that both A and B are labelled `in` by $\mathcal{L}$. But then B would be illegally `in`, which implies that $\mathcal{L}$ is not an admissible labelling. Contradiction.
We now prove that $\mathcal{A}rgs \subseteq F(\mathcal{A}rgs)$. Let $A \in \mathcal{A}rgs$. Then A is labelled `in` by $\mathcal{L}$. Let B be an arbitrary defeater of A. The fact that $\mathcal{L}$ is an admissible labelling means that $\mathcal{L}$ has no arguments that are illegally `in`, so from the fact that A is labelled `in` by $\mathcal{L}$ it follows

that B must be labelled `out`. The fact that $\mathcal{L}$ is an admissible labelling also means that $\mathcal{L}$ has no arguments that are illegally `out`, so from the fact that B is labelled `out` it follows that B must have a defeater (say C) that is labelled `in`. The fact that C is labelled `in` means that $C \in \mathcal{A}rgs$. So, as A is defended by $\mathcal{A}rgs$ against any possible defeater B, we have that $A \in F(\mathcal{A}rgs)$. □

The notion of a complete labelling actually corresponds to the notion of a complete extension.

Theorem 2. *Let (Ar, def) be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$. $\mathcal{A}rgs$ is a complete extension iff there exists a complete labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$.*

Proof.
"$\Longrightarrow$": Let $\mathcal{A}rgs$ be a complete extension. Now consider a labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$, $\texttt{out}(\mathcal{L}) = \mathcal{A}rgs^+$ and $\texttt{undec}(\mathcal{L}) = Ar - \texttt{in}(\mathcal{L}) - \texttt{out}(\mathcal{L})$. Since $\mathcal{A}rgs$ is a complete extension and therefore also an admissible set of arguments, it holds that $\mathcal{L}$ contains no arguments that are illegally `in` or illegally `out`, for reasons explained in the proof of Theorem 1.
We now prove that $\mathcal{L}$ also does not have any arguments that are illegally `undec`. Suppose there is an argument (say A) that is illegally `undec` in $\mathcal{L}$. We distinguish two possibilities:

- A has a defeater (say B) that is labelled `in` by $\mathcal{L}$. Then $B \in \mathcal{A}rgs$, so $A \in \mathcal{A}rgs^+$, so A is labelled `out` by $\mathcal{L}$. Contradiction.
- All defeaters of A are labelled `out` by $\mathcal{L}$. Then all defeaters of A are in $\mathcal{A}rgs^+$. This means that A is defended by $\mathcal{A}rgs$, so $A \in F(\mathcal{A}rgs)$. The fact that $\mathcal{A}rgs$ is a complete extension means that $\mathcal{A}rgs = F(\mathcal{A}rgs)$. Therefore, $A \in \mathcal{A}rgs$. It then follows that A is labelled `in` by $\mathcal{L}$. Contradiction.

The fact that $\mathcal{L}$ is a labelling without arguments that are illegally `in`, illegally `out` or illegally `undec` means that $\mathcal{L}$ is a complete labelling.
"$\Longleftarrow$": Let $\mathcal{L}$ be a complete labelling and $\mathcal{A}rgs = \texttt{in}(\mathcal{L})$. From the fact that $\mathcal{L}$ is a complete labelling, it follows that that $\mathcal{L}$ is also an admissible labelling. For reasons explained in the proof of Theorem 1, it then follows that $\mathcal{A}rgs$ is an admissible set. That is, $\mathcal{A}rgs \subseteq F(\mathcal{A}rgs)$.
We will now prove that $F(\mathcal{A}rgs) \subseteq \mathcal{A}rgs$. Let $A \in F(\mathcal{A}rgs)$. Then, for every argument B that defeats A there exists an argument $C \in \mathcal{A}rgs$ that defeats B. Since $C \in \mathcal{A}rgs$, it holds that C is labelled `in` by $\mathcal{L}$. As $\mathcal{L}$ does not contain any arguments that are illegally `in` or illegally `undec` (because $\mathcal{L}$ is a complete labelling) it then fllows that B has to be labelled `out` by $\mathcal{L}$. It then follows that A has to be labelled `in` by $\mathcal{L}$. Thus, $A \in \mathcal{A}rgs$. □

The notions of a grounded, preferred, stable and semi-stable labelling correspond to he notion of a grounded, preferred, stable and semi-stable extension, respectively.

Theorem 3. *Let (Ar, def) be an argumentation framework and $\mathcal{A}rgs \subseteq Ar$.*

- *$\mathcal{A}rgs$ is a grounded extension iff there exists a grounded labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$*
- *$\mathcal{A}rgs$ is a preferred extension iff there exists a preferred labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$.*
- *$\mathcal{A}rgs$ is a stable extension iff there exists a stable labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$.*
- *$\mathcal{A}rgs$ is a semi-stable extension iff there exists a semi-stable labelling $\mathcal{L}$ with $\texttt{in}(\mathcal{L}) = \mathcal{A}rgs$.*

Proof. Using the results of Theorem 2 this then follows in a straightforward way from Definition 3 and Definition 8. □

Before continuing with the backgrounds of the proposed algorithm, we first state a few useful properties of complete and admissible labellings.

Lemma 1. *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two complete labellings of (Ar, def). It holds that* $\texttt{in}(\mathcal{L}_1) \subseteq \texttt{in}(\mathcal{L}_2)$ *iff* $\texttt{out}(\mathcal{L}_1) \subseteq \texttt{out}(\mathcal{L}_2)$.

Proof.
"$\Longrightarrow$": Suppose $\texttt{in}(\mathcal{L}_1) \subseteq \texttt{in}(\mathcal{L}_2)$. Let $A \in \texttt{out}(\mathcal{L}_1)$. Then A has a defeater (say B) that is labelled `in` by $\mathcal{L}_1$. That is, $B \in \texttt{in}(\mathcal{L}_1)$. From the fact that $\texttt{in}(\mathcal{L}_1) \subseteq \texttt{in}(\mathcal{L}_2)$ it then follows that $B \in \texttt{in}(\mathcal{L}_2)$. Since $\mathcal{L}_2$ does not contain any arguments that are illegally `in` or illegally `undec` it then follows that $A \in \texttt{out}(\mathcal{L}_2)$.
"$\Longleftarrow$": Suppose $\texttt{out}(\mathcal{L}_1) \subseteq \texttt{out}(\mathcal{L}_2)$. Let $A \in \texttt{in}(\mathcal{L}_1)$. Then each defeater of A must be labelled `out` by $\mathcal{L}_1$. From the fact that $\texttt{out}(\mathcal{L}_1) \subseteq \texttt{out}(\mathcal{L}_2)$ it then follows that each defeater of A is also labelled `out` by $\mathcal{L}_2$. Since $\mathcal{L}_2$ does not contain any arguments that are illegally `out` or illegally `undec` it follows that $A \in \texttt{in}(\mathcal{L}_2)$. □

Lemma 2. *Let $\mathcal{L}_1$ be an admissible labelling. There exists a preferred labelling $\mathcal{L}_2$ with* $\texttt{in}(\mathcal{L}_1) \subseteq \texttt{in}(\mathcal{L}_2)$ *and* $\texttt{out}(\mathcal{L}_1) \subseteq \texttt{out}(\mathcal{L}_2)$.

Proof. The fact that $\mathcal{L}_1$ is an admissible labelling implies (Theorem 1) that $\texttt{in}(\mathcal{L}_1)$ is an admissible set. From [3] it then follows that there exists a preferred extension $\mathcal{A}rgs$ that is a superset of the admissible set. Let $\mathcal{L}_2$ be a preferred labelling with $\texttt{in}(\mathcal{L}_2) = \mathcal{A}rgs$. It then holds that $\texttt{in}(\mathcal{L}_1) \subseteq \texttt{in}(\mathcal{L}_2)$. From Lemma 1 it then follows that $\texttt{out}(\mathcal{L}_1) \subseteq \texttt{out}(\mathcal{L}_2)$. □

Lemma 3. *Let $\mathcal{L}$ be a preferred labelling and $\mathcal{L}'$ be an admissible labelling. It holds that:*

1. *if* $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}')$ *then* $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}')$
2. *if* $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}')$ *then* $\texttt{out}(\mathcal{L}) = \texttt{out}(\mathcal{L}')$

Proof. From Lemma 2 it follows that there exists a preferred labelling $\mathcal{L}''$ with $\texttt{in}(\mathcal{L}') \subseteq \texttt{in}(\mathcal{L}'')$ and $\texttt{out}(\mathcal{L}') \subseteq \texttt{out}(\mathcal{L}'')$.

1. Suppose that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}')$. It then follows that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}') \subseteq \texttt{in}(\mathcal{L}'')$. From the fact that both $\mathcal{L}$ and $\mathcal{L}''$ are preferred labellings, it then follows that $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}'')$, so also $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}')$.
2. Suppose that $\texttt{out}(\mathcal{L} \subseteq \texttt{out}(\mathcal{L}')$. It then follows that $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}') \subseteq \texttt{out}(\mathcal{L}'')$. From Lemma 1 it then follows that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}') \subseteq \texttt{in}(\mathcal{L}'')$. From the fact that both $\mathcal{L}$ and $\mathcal{L}''$ are preferred labellings, it then follows that $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}'')$, so also $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}')$. From $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}')$ it follows that $\texttt{in}(\mathcal{L}') \subseteq \texttt{in}(\mathcal{L})$. We can then apply Lemma 1 to obtain $\texttt{out}(\mathcal{L}') \subseteq \texttt{out}(\mathcal{L})$. From this, together with the fact that $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}')$, it follows then that $\texttt{out}(\mathcal{L}) = \texttt{out}(\mathcal{L}')$.

□

4 Formal background of the semi-stable algorithm

Now that the preliminary concepts have been explained, it is time to treat the main question of how to compute, given an argumentation framework, all the semi-stable labellings. The idea is to do this by generating a set $\mathcal{L}abellings$ of admissible labellings that includes at least all preferred labellings. Since every semi-stable labelling is also a preferred labelling [2, 1], this means that $\mathcal{L}abellings$ also contains all semi-stable labellings. We then have to select those labellings in $\mathcal{L}abellings$ with minimal `undec` to obtain the final answer.

How does one generate an admissible labelling? A possible approach is to start with the all-in labelling (the labelling in which every argument is labelled `in`). This labelling trivially satisfies the absence of arguments that are illegally `out`. However, for an admissible labelling also the absence of arguments that are illegally `in` is required, and the all-in labelling may contain many arguments that are illegally `in`. This means we need a way of changing the label of an argument that is illegally `in`, preferrably without creating any arguments that are illegally `out`. This is done using a sequence of *transition steps*. A transition step basically takes an argument that is illegally `in` and relabels it to `out`. It then checks if, as a result of this, one or more arguments have become illegally `out`. If this is the case, then these arguments are relabelled to `undec`. More precisely, a transition step can be described as follows.

Definition 9. *Let $\mathcal{L}$ be a labelling and A an argument that is illegally* `in` *in $\mathcal{L}$. A* transition step *on A in $\mathcal{L}$ consists of the following:*

1. *the label of A is changed from* `in` *to* `out`
2. *for every $B \in \{A\} \cup A^+$, if B is illegally* `out`*, then change the label of B from* `out` *to* `undec`*.*

Theorem 4. *Each transition step preserves the absence of arguments that are illegally* `out`*.*

Proof. Suppose that before the transition step, there are no arguments that are illegally `out`. Let A be an argument that is labelled `out` after the transition step. We distinguish two possibilities:

1. A was also labelled `out` before the transition step. Suppose it no longer has a defeater that is labelled `in`. But then A would have been relabelled as `undec` in the second part of the transition step. Contradiction.
2. A was not labelled `out` before the transition step. Then it must hold that A was labelled `in` before the transition step, and that the transition step was performed on A. This then means that A must have been illegally `in` before the transition step. From the fact that A was no relabelled to `undec` during the transition step, it can be inferred that A has a defeater that is labelled `in`. Therefore, A is legally `in`.

□

A transition sequence starts with an initial labelling $\mathcal{L}_0$, on which a sequence of successive transition steps is applied.

Definition 10. *A* transition sequence *is a list $[\mathcal{L}_0, A_1, \mathcal{L}_1, A_2, \mathcal{L}_2, \ldots, A_n, \mathcal{L}_n]$ ($n \geq 0$) where each A_i ($1 \leq i \leq n$) is an argument that is illegally* `in` *in labelling $\mathcal{L}_{i-1}$ and every $\mathcal{L}_i$ is the result of doing a transition step of A_i on $\mathcal{L}_{i-1}$. A transition sequence is called* terminated *iff $\mathcal{L}_n$ does not contain any argument that is illegally* `in`*.*

As an illustration of how transition steps are applied consider the two examples of Figure 2.

First, consider the example at the left hand side of Figure 2. Assume the initial situation is the all-in labeling $\mathcal{L}_0 = (\{A, B, C\}, \emptyset, \emptyset)$. In this labelling both B and C are illegally `in` since each of them has a defeater that is `in`, so they are both candidates for a transition step. If we select B for a transition step, then the result is a labelling $\mathcal{L}_1 = (\{A, C\}, \{B\}, \emptyset)$. This labelling does not contain any arguments that are illegally `in`, so the transition sequence $[\mathcal{L}_0, B, \mathcal{L}_1]$ is terminated. If, at the other hand, we select C for a transition step then the result is a labelling $\mathcal{L}'_1 = (\{A, B\}, \{C\}, \emptyset)$. This labelling still has an argument that is illegally `in` (B), so we perform another transition step that relabels B from `in` to `out`. However, as a result of doing that, C becomes illegally `out` since it has no longer a defeater that is `in`, so C is relabelled from `out` to `undec`. The

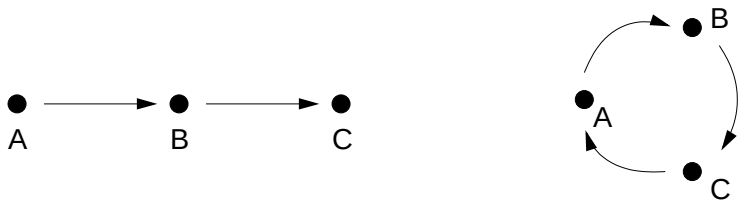

Figure 2: Two argumentation frameworks.

transition step as a whole then yields $\mathcal{L}'_2 = (\{A\}, \{B\}, \{C\})$. This means that there exists a second terminated transition sequence $[\mathcal{L}_0, C, \mathcal{L}'_1, B, \mathcal{L}'_2]$.

Now consider the example at the right hand side of Figure 2. Again, assume that the initial labelling is the all-in labelling, so $\mathcal{L}_0 = (\{A, B, C\}, \emptyset, \emptyset)$. Here, all three arguments are `in`, so each of them can be selected for a transition step. Assume, without loss of generality, that A is selected for a transition step. This then yields a labelling $\mathcal{L}_1 = (\{B, C\}, \{A\}, \emptyset)$. In this labelling, only C is illegally `in` and can be selected for a transition step. During this transition step, after C is relabelled from `in` to `out`, A becomes illegally `out` and is therefore relabelled to `undec`. Thus, the transition step as a whole yields $\mathcal{L}_2 = (\{B\}, \{C\}, \{A\})$. In this labelling B is illegally `in` since it has a defeater (A) that is `undec`. Therefore, a transition step on B is performed during which B is relabelled from `in` to `out`. Directly after doing that, however, not only C is illegally `out` but also B itself is illegally `out`, so both of them are relabelled from `out` to `undec`. Thus, the transition step as a whole yields $\mathcal{L}_3 = (\emptyset, \emptyset, \{A, B, C\})$. This means that there exists a terminated transition sequence $[\mathcal{L}_0, A, \mathcal{L}_1, C, \mathcal{L}_2, B, \mathcal{L}_3]$. It can be verified that in the example at the right hand side of Figure 2 every terminated transition sequence that starts with the all-in labelling finishes with $\mathcal{L}_3$.

Since for any finite argumentation framework, only a finite number of successive transition steps can be performed, this means that (again for finite argumentation frameworks) each terminated transition sequence is finite. Furthermore, for any terminated transition sequence, the final labelling is an admissible labelling, as is stated by the following theorem.

Theorem 5. *Let $[\mathcal{L}_0, A_1, \mathcal{L}_1, A_2, \mathcal{L}_2, \ldots, A_n, \mathcal{L}_n]$ ($n \geq 0$) be a terminated transition sequence where $\mathcal{L}_0$ is the all-in labelling. It holds that $\mathcal{L}_n$ is an admissible labelling.*

Proof. The fact that $\mathcal{L}_0$ does not contain any arguments that are illegally `out`, together with the fact that each transition step preserves the absence of arguments that are illegally `out`, implies that $\mathcal{L}_n$ does not contain any arguments that are illegally `out`. The fact that the transition sequence is terminated implies that $\mathcal{L}_n$ does not contain any arguments that are illegally `in`. Therefore, $\mathcal{L}_n$ is an admissible labelling. □

An interesting observation is that during the course of a transition sequence, the set of `in`-labelled arguments monotonically decreases and the set of `undec`-labelled arguments monotonically increases.

Proposition 1. *Let $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ be a transition sequence. For any $i \leq i \leq n$ it holds that $\texttt{in}(\mathcal{L}_i) \subseteq \texttt{in}(\mathcal{L}_{i-1})$ and $\texttt{undec}(\mathcal{L}_i) \supseteq \texttt{undec}(\mathcal{L}_{i-1})$.*

Proposition 1 is relevant with respect to the algorithm for generating the semi-stable labellings. Suppose that a previously generated terminated transition sequence yielded some admissible labelling $\mathcal{L}$ and we are currently expanding a transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_i, \mathcal{L}_i]$. Now, if it holds that $\texttt{undec}(\mathcal{L}_i) \supsetneq \texttt{undec}(\mathcal{L})$ then we already know that the current transition sequence cannot yield a semi-stable labelling, since expanding it to a terminated

transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ results in a labelling $\mathcal{L}_n$ with $\texttt{undec}(\mathcal{L}_n) \supsetneq \texttt{undec}(\mathcal{L})$ (this follows from Proposition 1 and the fact that $\texttt{undec}(\mathcal{L}_i) \supsetneq \texttt{undec}(\mathcal{L})$). This means we might as well stop expanding the current transition sequence and instead backtrack to another possibility.

We define *Labellings* as the set of all final labellings from terminated transition sequences that start from the all-in labelling.

As we have now obtained that the result of any terminated transition sequence starting from the all-in labelling is an admissible labelling, it directly follows that each element of *Labellings* is an admissible labelling. The next step, then, is to examine whether each semi-stable labelling will be an element of *Labellings*. If this is the case, then we can simply determine the semi-stable labellings as those elements of *Labellings* where undec is minimal. It turns out that this is indeed the case; it can be proved using the following two lemma's.

Lemma 4. *Let $\mathcal{L}$ be a preferred labelling. Let $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$, be a transition sequence where $\mathcal{L}_0$ is the all-in labelling and $\{A_1, \ldots, A_n\} = \texttt{out}(\mathcal{L})$. It holds that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_n)$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_n)$.*

Proof. First, we prove that $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ (where $\mathcal{L}_0$ is the all-in labelling and $\{A_1, \ldots, A_n\} = \texttt{out}(\mathcal{L})$) is indeed a valid transition sequence. For this, it should hold that each A_i is illegally in in $\mathcal{L}_{i-1}$. Suppose there exists an A_i $(1 \leq i \leq n)$ that is legally in in $\mathcal{L}_{i-1}$. Then each defeater B of A_i is out in $\mathcal{L}_{i-1}$. But since every argument that is out in $\mathcal{L}_{i-1}$ is also out in $\mathcal{L}$ this means that A_i would be illegally out in $\mathcal{L}$. This then means that $\mathcal{L}$ would not be a preferred labelling.

The next step is to prove that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_n)$. We prove this by contraposition. Suppose $A \notin \texttt{in}(\mathcal{L}_n)$. Then there is a transition step in the transition sequence that is performed on A. That is, $A = A_i$ for some A_i in $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$. This is because the only way that A can be relabelled from in to something else than in is by performing a transition step on it. But as $\{A_1, \ldots, A_n\} \subseteq \texttt{out}(\mathcal{L})$ it follows that $A \in \texttt{out}(\mathcal{L})$. Therefore, $A \notin \texttt{in}(\mathcal{L})$.

We now prove that $\texttt{undec}(\mathcal{L}) = \emptyset$. Suppose that $\texttt{undec}(\mathcal{L}) \neq \emptyset$. Let $\mathcal{L}_i$ be the the first labelling in the transition sequence (obtained by a transition step on A_i) that labels an argument undec. The fact that the transition step of A_i on $\mathcal{L}_{i-1}$ yields an argument labelled undec means that (Definition 9) after relabelling A_i from in to out, there is some $B \in \{A_i\} \cup A_i^+$ that is illegally out and therefore has to be relabelled undec. This means that (Definition 5) that B does not have a defeater that is labelled in after relabelling A_i from in to out in $\mathcal{L}_{i-1}$. As $\mathcal{L}_{i-1}$ does not contain any arguments labelled undec (recall that $\mathcal{L}_i$ is the *first* labelling that contains an argument labelled undec) this means that all defeaters of B are labelled out. This then implies that a transition step has been done on each defeater of B. Therefore, each defeater of B is an element of $\texttt{out}(\mathcal{L})$. But since B is also labelled out in $\mathcal{L}$ (otherwise it could not have been selected for the transition step that made it out) this would imply that B is illegally out in $\mathcal{L}$ and that therefore $\mathcal{L}$ is not a complete labelling, and therefore also not a preferred labelling. Contradiction.

We now prove that $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_n)$. Suppose that $A \notin \texttt{out}(\mathcal{L}_n)$. From the fact that $\texttt{undec}(\mathcal{L}_n) = \emptyset$ it then follows that $A \in \texttt{in}(\mathcal{L}_n)$. This implies that no transition step has been done on A during the transition sequence. From this, it follows that $A \notin \texttt{out}(\mathcal{L})$. □

Lemma 5. *Let $\mathcal{L}$ be a preferred labelling and let $\mathcal{L}_0$ be a labelling without arguments that are illegally* out *and satisfying $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_0)$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_0)$. There exists a terminated transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ with $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_n)$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_n)$.*

Proof. We first observe that each transition step satisfies the following property: If in the labelling before the transition step $(\mathcal{L}_i)$ it holds that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_i)$ and $\texttt{out}(\mathcal{L}) \subseteq$

$\texttt{out}(\mathcal{L}_i)$ then in the labelling after the transition step ($\mathcal{L}_{i+1}$) it holds that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_{i+1})$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_{i+1})$. Suppose this would not be the case. We distinguish two possibilities:

1. There is an argument A that is labelled `in` in $\mathcal{L}$ but not labelled `in` in $\mathcal{L}_{i+1}$. From the fact that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_i)$ it follows that the transition step was performed on A. This means that A was illegally `in` in $\mathcal{L}_i$, so A has a defeater (say B) that is not `out` in $\mathcal{L}_i$. From the fact that $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_i)$ it follows that B is also not labelled `out` in $\mathcal{L}$. But then A would be illegally `in` in $\mathcal{L}$, so $\mathcal{L}$ would not be a preferred labelling. Contradiction.
2. There is an argument A that is labelled `out` in $\mathcal{L}$ but not labelled `out` in $\mathcal{L}_{i+1}$. The fact that $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_i)$ means that A is labelled `out` in $\mathcal{L}_i$. The fact that A is not labelled `out` in $\mathcal{L}_{i+1}$ then implies that A is labelled `unec` in $\mathcal{L}_{i+1}$. The fact that A is labelled `out` in $\mathcal{L}$ means that A has a defeater (say B) that is labelled `in` in $\mathcal{L}$. The fact that $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_i)$ means that B is also labelled `in` in $\mathcal{L}_i$. This means that the transition step must have been performed on B (otherwise B would still be `in` in $\mathcal{L}_{i+1}$ and A would not have been relabelled from `out` to `undec`). But then B would have been illegally `in` in $\mathcal{L}_i$, so B would also have been illegally `in` in $\mathcal{L}$. Contradiction.

□

Theorem 6. *Let $\mathcal{L}$ be a preferred labelling. There exists a transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ where $\mathcal{L}_0$ is the all-in labelling and $\mathcal{L}_n = \mathcal{L}$.*

Proof. From Lemma 4 it follows that there exists a transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_m, \mathcal{L}_m]$ ($m \geq 0$) where $\mathcal{L}_0$ is the all-in labelling, $\{A_1, \ldots, A_m\} = \texttt{out}(\mathcal{L})$, $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_m)$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_m)$. From Lemma 5 it then follows that there exists a terminated transition sequence $[\mathcal{L}_m, A_{m+1}, \mathcal{L}_{m+1}, \ldots, A_{m+k}, \mathcal{L}_{m+k}]$ with $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_{m+k})$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_{m+k})$. These two transition sequences can be concaternated to a transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_m, \mathcal{L}_m, A_{m+1}, \mathcal{L}_{m+1}, \ldots, A_{m+k}, \mathcal{L}_{m+k}]$. Let $n = m + k$. We then have a transition sequence $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ which satisfies $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}_n)$ and $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}_n)$. The fact that this transition sequence is terminated means that (Theorem 5) $\mathcal{L}_n$ is an admissible labelling. The fact that $\mathcal{L}$ is a preferred labelling implies that (Lemma 3) for any admissible labelling $\mathcal{L}'$ with $\texttt{in}(\mathcal{L}) \subseteq \texttt{in}(\mathcal{L}')$ it holds that $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}')$ and for any admissible labelling with $\texttt{out}(\mathcal{L}) \subseteq \texttt{out}(\mathcal{L}')$ it holds that $\texttt{out}(\mathcal{L}) = \texttt{out}(\mathcal{L}')$. Therefore, we have that $\texttt{in}(\mathcal{L}) = \texttt{in}(\mathcal{L}_n)$ and $\texttt{out}(\mathcal{L}) = \texttt{out}(\mathcal{L}_n)$, so also $\texttt{undec}(\mathcal{L}) = \texttt{undec}(\mathcal{L}_n)$. This means that $\mathcal{L} = \mathcal{L}_n$. □

5 Optimizing the semi-stable algorithm

As was shown in section 4, for the example at the left hand side of Figure 2 there are two terminated transition sequences starting from the all-in labelling: one that yields $(\{A, C\}, \{B\}, \emptyset)$ and one that yields $(\{A\}, \{B\}, \{C\})$. This is because starting from the all-in labelling, we have two choices of arguments to do a transition step on: B or C. This is because both of them are illegally `in` in $\mathcal{L}_0$ (the all-in labelling). If we choose B we will finally end up with a complete labelling, but if we choose C then we will ultimately end up with a labelling that is admissible but not complete (and therefore also not preferred or semi-stable). An interesting question, therefore, is whether there is a way of avoiding such non-complete results by carefully choosing the right arguments to do the transition steps on. While in general this question is difficult to answer, we do propose a simple guideline that is helpful in many cases: choose an argument that is *superillegally* `in` to do a transition step on, if such an argument is available.

Definition 11. *Let $\mathcal{L}$ be a labelling of (Ar, def). An argument A is* superillegally `in` *in $\mathcal{L}$ iff A is labelled* `in` *by $\mathcal{L}$ and is defeated by an argument that is legally* `in` *in $\mathcal{L}$ or* `undec` *in $\mathcal{L}$.*

It directly follows that if an argument is superillegally `in` in $\mathcal{L}$, then it is also illegally `in` in $\mathcal{L}$. The converse, however, may not be the case. As an example, consider again the example at the left hand side of Figure 2. With the all-in labelling, A is legally `in`, B and C are illegally `in`, and only B is superillegally `in`. Thus, it makes sense to select B to do a transition step on.

The reason why arguments that are superillegally `in` are such good candidates to perform a transition step on is that an argument that is superillegally `in` will stay illegally `in` (although it may not necessarily stay *super*illegally `in`) throughout the transition sequence, until a transition step is done on it. Thus, we might as well perform a transition step on the superillegal argument as soon as possible, since this prevents us from doing things we later regret (like performing a transition step on argument C).

Theorem 7. *Let $\mathcal{L}_0$ be a labelling where argument A is superillegally* `in` *and $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_n, \mathcal{L}_n]$ be a transaction sequence where no transaction step is performed on A (that is: $A \notin \{A_1, \ldots, A_n\}$). It holds that A is illegally* `in` *in $\mathcal{L}_n$.*

Proof. We distinguish two cases:

1. A is superillegally `in` because it has a defeater (say B) that is labelled `undec` in $\mathcal{L}_0$. Then B is still labelled `undec` in $\mathcal{L}_n$ so A is still (super)illegally `in` in $\mathcal{L}_n$.
2. A is superillegally `in` because it has a defeater that is legally `in` in $\mathcal{L}_0$. The fact that B is legally `in` in $\mathcal{L}_0$ means that every defeater of B is labelled `out` in $\mathcal{L}_0$. It is possible that during the transition sequence, the label of one of B's defeaters is changed. We distinguish two possibilities:
   (a) None of the transition steps in the transition sequence alters the label of any of B's defeaters. In that case B is still legally `in` in $\mathcal{L}_n$. Therefore, A is still (super)illegally `in` in $\mathcal{L}_n$.
   (b) At least one of B's defeaters (say C) has been relabelled during the course of the transition sequence. Since C was labelled `out` it can only be relabelled as `undec`. As this means that B has become illegally `in` at some place in the transition sequence, it would be a potential candidate to perform a transition step on. If such a transition step was performed, B would have been relabelled from `in` to `undec` (`out` is not an option since every transition step preserves the absence of arguments that are illegally `out`), which would imply that A is still (super)illegally `in` in $\mathcal{L}_n$. If such a transition step was not performed, then B is still `in` (although illegally) in $\mathcal{L}_n$, and A is still illegally (although not *super*illegally) `in` in $\mathcal{L}_n$.

□

From Theorem 7 it follows that it may be a good idea to select an argument that is superillegally `in` to do a transition step on, whenever such an argument is available. An interesting question is how such a strategy would affect the results that were obtained earlier regarding correctness (each transition sequence terminates with an admissible labelling) and completeness (for each preferred labelling, there exists a transition sequence that produces this labelling).

As for correctness, the situation does not change. The result of a terminated transition sequence is always an admissible labelling, regardless of which strategy has been used to select the arguments to do transition steps on.

As for completeness, we need to pay special attention to the possible effects regarding on Lemma 4. When we commit ourselves to a strong preference for superillegal arguments,

can we still choose arguments that are `out` in $\mathcal{L}$? It would help if we could prove that at least one of the superillegal arguments is `out` in $\mathcal{L}$ whenever at least one superillegal argument is present. This would then mean that we have at least one "good" choice available in situations that force us to choose a superillegal argument. It would be even better of we could prove that *every* superillegal argument is `out` in $\mathcal{L}$ whenever at least one superillegal argument is present. This would mean that *every* choice is a good choice, as long as we choose a superillegal argument. It turns out that this is indeed the case.

Lemma 6. *Let $\mathcal{L}$ be a preferred labelling. Let $[\mathcal{L}_0, A_1, \mathcal{L}_1, \ldots, A_i, \mathcal{L}_i]$ be a transition sequence ($i \geq 0$) with $\{A_1, \ldots, A_i\} \subseteq \texttt{out}(\mathcal{L})$. If argument A is superillegally `in` in $\mathcal{L}_i$, then $A \in \texttt{out}(\mathcal{L})$.*

Proof. Suppose that A is superillegally `in` in $\mathcal{L}_i$. Since the transition sequence does not yield arguments that are labelled `undec` (for reasons explained in the proof of Lemma 4) this can only mean that A is defeated by an argument (say B) that is legally `in` in $\mathcal{L}_i$. The fact that B is legally `in` means that every defeater of B is `out` in $\mathcal{L}_i$. This means a transition step has been done on each and every defeater of B during the transition sequence. Thefore, each defeater of B is `out` in $\mathcal{L}$. Since $\mathcal{L}$ is a preferred (and therefore complete) labelling, this then implies that B is labelled `in` in $\mathcal{L}$. Therefore, A is labelled `out` in $\mathcal{L}$. □

Since our new selection strategy ("choose an arbitrary superillegal argument if available") does not affect Lemma 5 and Theorem 6, it can be concluded that one still finds each preferred labelling (and therefore also each stable and semi-stable labelling).

6 The actual algorithm

Since the algoritm starts with the labelling in which every argument is labelled `in`, we assume the presence of the constant `all_in`, which stands for the all-in labelling. There is one global variable (`potential_semi_stables`) which stands for the admissible labellings with minimal `undec` that have been found until now. If, during the search algorithm, one finds that the current labelling is worse (that is: it has a proper superset of `undec` labelled arguments) than an admissible labelling found earlier, then it is time to stop evaluating the current transition sequence, since its final result will not be semi-stable anyway.

If there is no argument that is illegally `in` then we are at the end of a terminated transition sequence and have obtained an admissible labelling. From the previous check, we already know that this admissible labelling is not any worse than what we already have found (it does not have a proper superset of `undec` labelled arguments compared to a previously computed admissible labelling), so we add it to the set of potential semi-stable labellings (`potential_semi_stables`). We then have to check if we found something that is actually *better* than what we found earlier. If so, we need to delete some of the old results (remove it from `potential_semi_stables`).

If we have not reached the end of a terminated transition sequence, then there is at least one argument that is still illegally `in`. We then distinguish two cases. If there is at least one argument that is superillegally `in` then go for the argument that superillegally `in`. There is no need to be selective; any argument that is superillegally `in` will do for a transition step. If, however, there is no argument that is superillegally `in` then we have to try, one by one, each argument that is "normally" illegally `in`. We have to try each of them to see which one will yield a labelling with minimal `undec`.

Thus, the algorithm can be described as follows:

```
01.  potential_semi_stables = ∅;
02.  find_semi-stables(all-in);
03.  print potential_semi_stables;
04.  end;
05.
06.  procedure find_semi-stables(ℒ)
07.     # if we have got something worse than was found earlier,
08.     # then prune the search tree and backtrack
09.     if ∃ℒ′ ∈ potential_semi-stables:  undec(ℒ′) ⊊ undec(ℒ) then return;
10.     # now see if the transition sequence has terminated
11.     if ℒ does not have an argument that is illegally in then
12.       for each ℒ′ ∈ potential_semi-stables
13.         # if the old results are worse than our new labelling:  remove
14.         if undec(ℒ) ⊊ undec(ℒ′) then
15.           potential_semi-stables := potential_semi-stables - ℒ′;
16.         endif;
17.       endfor;
18.       # add our newly found labelling as a candidate
19.       # we already know that it is not worse than what we already have
20.       potential_semi-stables := potential_semi-stables ∪ℒ;
21.       return; # we are done with this one; let's try the next possibility
22.     else
23.       if ℒ has an argument that is superillegally in then
24.         A := some argument that is superillegally in in ℒ;
25.         find_semi-stables(transition_step(A, ℒ));
26.       else
27.         for each argument A that is illegally in in ℒ
28.           find_semi-stables(transition_step(A, ℒ);
29.         endfor;
30.       endif;
31.     endif;
32.  endproc;
```

7 Discussion

It is interesting to observe that the algorithm stated in section 6 can also be used to calculate, respectively, stable semantics and preferred semantics, by applying a few changes.

For stable semantics, the modification is quite straightforward. Basically, the idea (Definition 8) is only to yield labellings without `undec`-labelled arguments. For this, we have to stop expanding a transition sequence as soon as an `undec`-labelled argument is produced. Therefore, we have to replace line 9 by
`if` $\mathtt{undec}(\mathcal{L}) \neq \emptyset$ `then return`.
Furthermore, we do not have to compare the sets of `undec`-labelled arguments of the previous results with the current result, so the lines 12 until 17 can be removed. Then, after renaming the variable `potential_semi-stables` to `stables` and renaming the procedure `find_semi-stables` to `find_stables`, the modifications are finished and the result is an algorithm that calculates all stable extensions of an argumentation framework.

For preferred semantics, the modification is slightly different. The idea is that we still have to check for a condition that allows us to cut off the current transition sequence once we know that it will not yield a useful result. For semi-stable semantics, it can be observed that the set of `undec`-labelled arguments keeps *increasing* as the transition sequence progresses (Proposition 1). For preferred semantics, it can be observed that

the set of `in`-labelled arguments keeps *decreasing* as the transition sequence progresses (Proposition 1). In both cases, there may come a point where the current transition sequence becomes worse than a result found earlier, which means we might as well stop expanding it and instead backtrack to another possibility.

The modification for preferred semantics is done as follows. First, the variable `potential_semi-stables` to renamed as `potential_preferreds` and the procedure `find_semi-stables` is renamed as `find_preferreds`. Line 9 is replaced by:
`if` $\exists\mathcal{L}' \in$ `potential_preferreds such that` $\mathtt{in}(\mathcal{L}') \supsetneq \mathtt{in}(\mathcal{L})$ `then return;`
Line 14 is replaced by:
`if` $\mathtt{in}(\mathcal{L}) \supsetneq \mathtt{in}(\mathcal{L}')$ `then`
The result, then, is an algorithm that calculates, given an argumentation framework, all preferred extensions.

In [1], it was first examined how argument labellings are related to the traditional Dung-style argument semantics.[1] It was found that a complete labelling has a maximal set of `in`-labelled arguments iff it has a maximal set of `out`-labelled arguments. In both cases, the labelling corresponds with a preferred extension. Furthermore, it was found that a complete labelling has a minimal set of `in`-labelled arguments iff it has a minimal set of `out`-labeled arguments iff it has a maximal set of `undec`-labelled arguments. In all three cases, the labelling corresponds with the grounded extension. The only option that is then left to be examined is are the labellings where the set of `undec`-labelled arguments is minimal. It turned out that these did not correspond with any well-known semantics, and this is how semi-stable semantics was discovered [2, 1]. Thus, one can perhaps think of semi-stable semantics as a missing link in the traditional hierarchy of argumentation semantics (see Figure 3).

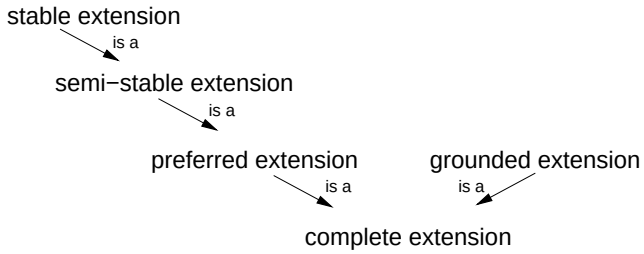

Figure 3: An overview of argumentation semantics.

Nevertheless, semi-stable semantics is more than just a purely theoretical notion. Stable semantics, despite its property of the potential absence of stable extensions, is still being used for the purpose of constraint satisfaction in fields like answer set programming [4]. The idea is that a problem is specified in a declarative way and that the set of potential solutions then corresponds with the stable models of the thus described problem. In cases where no solutions exists, there should therefore also not exist any stable models. Thus, the absence of stable models (or extensions) is not always an undesirable property. This does assume, however, that the original problem was encoded in a way that is perfectly correct. If, for instance, an answer set program contains an error, then the result may well be a total absence of stable models, which is a situation that can be notoriously hard to debug. With semi-stable semantics, however, one obtains one or more models that can serve as a starting point to examine where things went wrong. For instance, consider the example at the left hand side of Figure 1. It has no stable extensions and its (only)

[1] A *reinstatement labelling* in [1] is defined in a slightly different way than a complete labelling in the current paper. It can be shown, however, that both definitions are equivalent.

semi-stable extension is $(\{B, D\}, \{C\}, \{A\})$. The fact that A is labelled `undec` can be seen as an indication of what is "wrong" in this argumentation framework from the perspective of stable semantics. Similarly, if there exists an odd loop that causes the absence of stable models, then this odd loop is flagged `undec` by a semi-stable model. Thus, semi-stable semantics can give a good indication of where to start debugging if no stable model exists. Semi-stable semantics does a better job here than, for instance, preferred semantics. This is because there can be non-stable preferred models (like $(\{A\}, \{B\}, \{C, D, E\})$) even in cases where stable models (like $(\{B, D\}, \{A, C, E\}, \emptyset)$ do exist. With semi-stable semantics, one obtains non-stable semi-stable models only if there is a real problem that prevens the existence of stable models.

One particularly interesting application of semi-stable semantics would be answer set programming and other forms of logic programming that use the stable model semantics. At the time of writing, the author is exploring the possibilities of applying semi-stable semantics to logic programming and answer set programming. We believe this would be a useful approach for analyzing programs for which no stable models exist.

Thanks

We would like to thank Newres Al Haider for helping to develop the initial idea.

References

[1] M.W.A. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence; 10th European Conference, JELIA 2006*, pages 111–123. Springer, 2006. LNAI 4160.

[2] M.W.A. Caminada. Semi-stable semantics. In P.E. Dunne and TJ.M. Bench-Capon, editors, *Computational Models of Argument; Proceedings of COMMA 2006*, pages 121–130. IOS Press, 2006.

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

[4] M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–385, 1991.

[5] J. L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person.* MIT Press, Cambridge, MA, 1995.

[6] Bart Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In Manuela M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India*, pages 623–628, 2007.

[7] G.A.W. Vreeswijk. An algorithm to compute minimally grounded and admissible defence sets in argument systems. In P.E. Dunne and TJ.M. Bench-Capon, editors, *Computational Models of Argument; Proceedings of COMMA 2006*, pages 109–120. IOS, 2006.